Eurographics Workshop on Visual Computing for Biology and Medicine (2023)
C. Hansen, J. Procter, and R. G. Raidou (Editors)

Neural Deformable Cone Beam CT

L. Birklein[1], E. Schömer[1], R. Brylka[2], U. Schwanecke[2] and R. Schulze[3]

[1] Johannes Gutenberg University, Mainz, Germany
[2] RheinMain University of Applied Sciences, Wiesbaden, Germany
[3] University of Bern, Switzerland

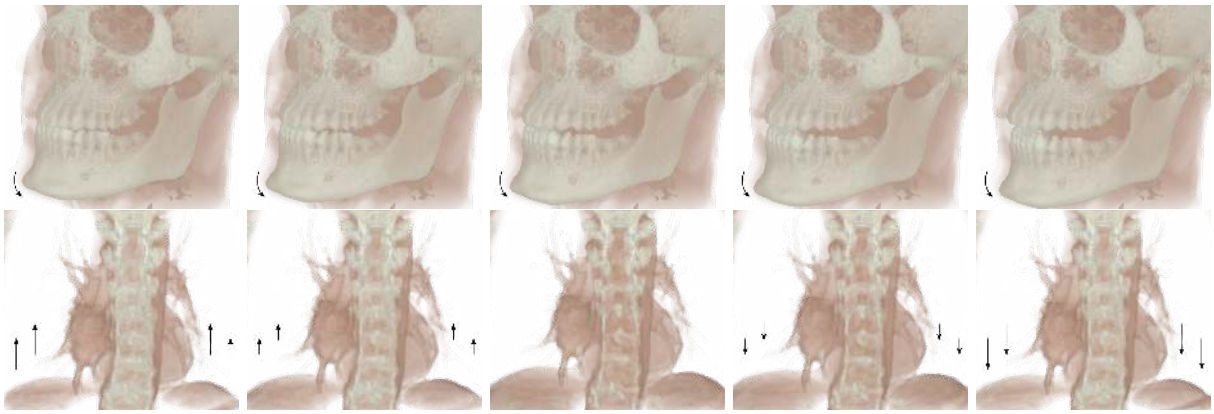


Figure 1: *We propose a motion-aware cone beam CT reconstruction method based on neural inverse rendering, which is applicable to a variety of different scenarios. In the first row we show the reconstruction of a dental CBCT scan of a patient opening the mouth, the bottom row shows the inhale and exhale phase of a 4D-CBCT of the thorax.*

Abstract

In oral and maxillofacial cone beam computed tomography (CBCT), patient motion is frequently observed and, if not accounted for, can severely affect the usability of the acquired images. We propose a highly flexible, data driven motion correction and reconstruction method which combines neural inverse rendering in a CBCT setting with a neural deformation field. We jointly optimize a lightweight coordinate based representation of the 3D volume together with a deformation network. This allows our method to generate high quality results while accurately representing occurring patient movements, such as head movements, separate jaw movements or swallowing. We evaluate our method in synthetic and clinical scenarios and are able to produce artefact-free reconstructions even in the presence of severe motion. While our approach is primarily developed for maxillofacial applications, we do not restrict the deformation field to certain kinds of motion. We demonstrate its flexibility by applying it to other scenarios, such as 4D lung scans or industrial tomography settings, achieving state-of-the art results within minutes with only minimal adjustments.

CCS Concepts

• ***Computing methodologies*** → ***Reconstruction; Volumetric models; Motion processing;*** *Neural networks;*

1. Introduction

Significant technical and algorithmic advances in recent decades have made X-ray computed tomography one of the most important medical imaging modalities. A specific type of computed tomography, invented in the 1990s for oral and maxillofacial radiology, is cone beam computed tomography (CBCT), in which the individual X-rays form a cone, and a flat panel detector is used to capture unique X-ray images. Today, CBCT is not only used in dentistry but also in other medical fields, such as interventional radiology or image-guided radiotherapy, or even in industrial processes, such as quality control of components.

A CBCT scanner rotates around the object under investigation and generates hundreds of images during the imaging process. Thereby, CBCT acquisition times can range from 10 to 40

seconds for typical dentomaxillofacial applications up to several minutes, e.g. for 4D-CBCT thorax scans of the lung. Traditional CBCT reconstruction algorithms, most notably the FDK algorithm [FDK84], but also iterative reconstruction methods [GBH70, AK84], assume stationary, non-deforming objects. The examined object's movement during acquisition can lead to image distortions and other artifacts like blurring, streaks, and double contours within the reconstructed volume [SHG*11]. In some cases, such as thorax scans of the lung, the patient's respiratory motions are unavoidable. Still, involuntary patient movements are also a consistently reported phenomenon in dentomaxillofacial CBCT scans [SNMS*17, HDO*13]. These often severely limit the clinical value of the reconstruction and frequently lead to a repeat of the imaging process, which in turn increases the patient's exposure to radiation [SNMS*15, TJJ*15, SNW16, SNCS*18].

In 2020, a new approach to scene reconstruction called neural radiance fields (NeRFs) [MST*21] has been introduced. NeRFs allow the synthesis of a continuous volumetric representation of an object using only a sparse set of two-dimensional images. Given this synthetic representation, new, previously unseen views can be generated. Later, neural radiance fields were extended to capture non-rigidly deforming scenes by optimizing an additional continuous volumetric deformation field that warps each observed point into a canonical 5D NeRF [PSB*21, PCPMMN20, LCM*22]. Similar to this technique, in this paper we present a reconstruction method for non-stationary deformable objects solely based on the acquired X-ray images without the need for any prior knowledge. The quality of our reconstructions is equal to or better than specifically tailored state-of-the-art reconstruction methods.

In summary our main contributions are 1. a lightweight, memory efficient system for fast neural reconstruction in X-ray cone beam computed tomography, 2. a prior-free deformable reconstruction model for various medical scenarios containing patient motion and 3. the integration of a sparsity inducing regularization term into the deformation field.

2. Related work

Motion correction In general, motion correction in CBCT has attracted a lot of interest in recent years, since it has applications in many different fields, such as CBCTs of the beating heart, 4D-CBCT thorax scans, or (mostly rigid) motion correction in maxillofacial and oral scenarios. Typically, approaches applied to deforming regions such as the heart or lung make use of the periodicity of the motion and divide the projection images into different phases, or make use of some sort of prior [FBSK14, MJR16, BMA*16, RCWH18]. Deformable CBCT also has applications outside the medical field. Zang et al. developed two algorithms for continuously deforming objects in CBCT [ZIT*18, ZIT*19]. However, those rely heavily on a specific acquisition strategy, which allows multiple initial reconstructions, that can later be iteratively improved by estimating the deformation between them. In a medical context, especially in maxillofacial CBCT, this possibility is not given.

Many successful prior-free motion correction methods can only handle the case of single rigid patient motions [OJS*17, SJP*21, NST*19], or multiple rigid motions within one field of view (FOV) [BNS*23, BMA*16]. They rely on a variety of different metrics, such as consistency conditions [BXA*17, PMM*19], autofocus [HBR*17, RBSF13, SSY*17] or reprojection error [NST*19, OJS*17, BMA*16]. Our method differs from these in that it is not restricted to rigid motions and at the same time does not rely on any external information, such as priors or periodicity.

Neural rendering and reconstruction Starting with NeRFs (Mildenhall et al. [MST*21]) in 2020, neural (inverse) rendering has become a very active field of research. Its original main goal is to synthesize novel views of complex 3D scenes which are captured by a number of photographs from different positions and directions. From here, research has gone in many different directions, e.g. improving reconstruction speed by utilizing a parametric input encoding, such as a grid [CLI*20, JSM*20, LGL*20], a tree [TLY*21] or hashing [MESK22], or by supporting deformable and moving scenes [PSB*21, PCPMMN20, LCM*22]. Neural rendering and reconstruction based techniques have also been successfully applied to computed tomography, e.g. by Rückert et al. [RWL*22], Sun et al. [SLX*21] or Shen et al [SPX23]. Recently, this technique has been used for 4D-CBCT scans of the lung [ZSPM23], this work however still needs an available prior 4D CT scan and was only able to produce results of dimension 64^3. Another recent work [RKA*21] also models the reconstruction as an implicit neural representation and jointly optimizes a deformation field. Their deformation field is modeled as a tensor of polynomial coefficients, such that each voxel is warped using a polynomial of degree k. This approach is quite memory intensive and the resulting reconstructions are supported only up to a resolution of 256^3. Inspired by those approaches we combine neural reconstruction and neural deformation fields into a novel method, capable of producing high-quality, high-resolution results.

3. Method

The key idea of our approach is to combine neural cone beam CT reconstruction with a neural deformation field, i.e. a deformation field realized by a neural network. This allows us to represent arbitrary motions and deformations, e.g., unwanted patient movements such as head rotations, swallowing, etc. Contrary to classic, voxel-based representations of the reconstructed volume, we employ a coordinate-based representation, utilizing a multi layer perceptron (MLP) together with a parametric input encoding. This volume representation is continuous, i.e., allowing samples at arbitrary points $\mathbf{x} \in \mathbb{R}^3$ and is also fully differentiable due to the nature of MLPs. To account for deformations and patient motion, every input $\mathbf{x}$ together with a timestamp $t \in [0,1]$ is passed through a neural deformation field, which models the current deformation state of the object. Its output $\hat{\mathbf{x}} \in \mathbb{R}^3$ is fed into the continuous volume representation, called density network, which then computes the density $\sigma(\hat{\mathbf{x}})$ at the given position. For a brief overview see Fig. 2. As the deformation field is also implemented via an MLP with input encoding, the whole pipeline is fully differentiable and can be optimized using a (stochastic) gradient descent algorithm.

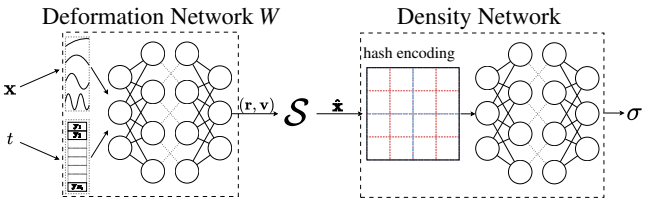


Figure 2: *Layout of the network architecture.*

3.1. Neural Computed Tomography

In general, computed tomography aims to find the solution of the inverse problem

$$I = I_0 \exp\left(-\int_{t_n}^{t_f} \sigma(\gamma(t))\,\mathrm{d}t \right) \tag{1}$$

which can be derived using Beer-Lamberts law. Here, I_0 is the initial intensity at the source and I is the measured intensity at the detector along the ray γ. The value $\sigma(\mathbf{x})$ is the attenuation of the scanned object (effectively the density) at the spatial position $\mathbf{x} \in \mathbb{R}^3$. In practice, Eq. (1) is usually transformed into the logarithmic domain

$$\log(I_0) - \log(I) = \int_{t_n}^{t_f} \sigma(\gamma(t))\,\mathrm{d}t \tag{2}$$

as its discretization then results in a sum of all values along each ray γ, i.e.

$$b := \log(I_0) - \log(I) \approx \sum_{i=1}^{N_\gamma} \sigma(\gamma(t_i))\delta_i =: \tilde{b} \tag{3}$$

where N_γ is the number of samples along the ray γ, and δ_i are the (possibly) varying step sizes, depending on the ray sampling strategy.

Choosing a differentiable loss function $\mathcal{L}$ makes it possible to optimize Eq.(3) for σ in a gradient descent based fashion. Traditionally, $\mathcal{L}$ is chosen to be the $L2$ norm of the residual, as this results in a simple least-squares problem that can be solved by algorithms with high convergence rates, such as the conjugate gradient method for least squares (CGLS).

In our approach, the classic voxel-based representation of σ is substituted by a neural network, which is adjusted in each training iteration by feeding the re-projection error $\mathcal{L}(\tilde{b}, b)$ back into the network at each input sampling point $\mathbf{x} = \gamma(t)$. As this is optimized using a gradient descent algorithm that does not depend on higher order derivatives, we have more freedom in choosing a suitable loss function $\mathcal{L}$ and in the implementation, we decided for Huber loss [Hub64], as it turns out to be more stable against outliers and less susceptible to noise in real-world data.

3.1.1. Input Encoding

In practice, it is beneficial to encode the three dimensional spatial input coordinates into a domain of higher dimension [TSM*20] because neural networks tend to produce overly smooth results [RBA*18] and thus need a very high number of training steps until high-frequency information is learned (if ever).

Earlier works in this field encoded the input into a domain of higher dimension and frequency, e.g., by

$$\mathrm{Enc}(\mathbf{x}) = (\sin(2^0\pi\mathbf{x}), \cos(2^0\pi\mathbf{x}), \ldots, \sin(2^k\pi\mathbf{x}), \cos(2^k\pi\mathbf{x})) \tag{4}$$

for a given maximum k, aka frequency encoding, as used in the original NeRF paper [MST*21]. This encoding enables the network to learn higher frequency features earlier in the training process.

More recently, parameterized input encodings have emerged, where the encoding itself is also optimized in the training process. In such network architectures, most of the information is actually embedded in the encoding, and the network can be of very small size (in some cases only one hidden layer). Müller et al. [MESK22] presented the multiresolution hash encoding, a parameterized input encoding that allows for short training times on a single GPU. Here, each spatial input vector $\mathbf{x}$ is located within a multiresolution grid, with resolutions increasing from $N_{\min}$ to $N_{\max}$ across L levels, of which each grid point at each level is associated with an F-dimensional vector of parameters (feature vector) via a spatial hash function with a fixed hash table size T. The final input of the MLP is computed by using linear interpolation of the feature vectors at each level (proportional to the distance of $\mathbf{x}$ to its neighboring grid points) and then concatenating the resulting feature vectors of each level.

3.2. Neural Deformation Fields

We model the motion and deformation of a subject undergoing capture with a deformation field Γ composed of a neural network W (more precisely, an MLP with input encoding) acting on input coordinates $\mathbf{x} \in \mathbb{R}^3$ and timestamp $t \in [0,1]$, and an embedding $\mathcal{S}$ (see Sec. 3.2.1). Γ represents the deformation of the subject at each given time t, which in our case is the normalized index of the current acquisition image, so it can be seen as a function

$$\Gamma : \mathbb{R}^4 \to \mathbb{R}^3,\ (\mathbf{x}, t) \mapsto \hat{\mathbf{x}}.$$

The coordinate $\hat{\mathbf{x}}$ within the canonical, or default, state is then entered into the reconstruction network representing σ. Essentially, Γ bends each ray γ before it is traced through the density network. Some previous work enforces the canonical state to be e.g., at time $t = 0$; however, in our experiments, we find that anchoring this state to a particular frame is not advantageous.

The positional input $\mathbf{x}$ of the deformation network W is encoded using frequency encoding, as in Eq. (4). Using frequency encoding for the time parameter can also lead to a "wobble effect" in the deformation field. Therefore we encode the time parameter t using a 1D grid of N_t feature vectors $\mathbf{y} \in \mathbb{R}^{F_t}$ with linear interpolation, which are also optimized during the learning process. Encoding t in this way allows the deformation field to remain stationary for frames in which motion is absent. We chose $F_t = 16$ in all cases and N_t to be the number of projection images, if not stated otherwise, thus assigning each image one feature vector. In all experiments $\mathbf{x}$ is encoded using four frequencies, i.e. setting $k = 3$ in Eq.(4). We employ the coarse-to-fine regularization scheme to the positions $\mathbf{x}$, as described by Park et al. [PSB*21], activating higher frequencies consecutively over the course of 16'000 iterations.

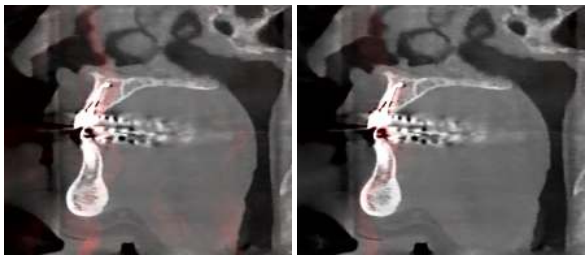


Figure 3: *Elastic energy (red) in a motion-impaired clinical application within the same slice when using the regularization term $\|\log\Sigma\|_F^2$ (left image) compared to $\|\Sigma - I\|_1$ (right image). Our sparsity inducing regularization term is more effective in removing non-rigid deformations, visible as fewer red streaks in the right image.*

3.2.1. Embedding

The easiest and probably most canonical implementation of the deformation field would be to learn an offset vector $\Delta\mathbf{x}$ for each input $(\mathbf{x},t)$ and simply add it to the position, s.t.

$$\hat{\mathbf{x}} = \mathbf{x} + \Delta\mathbf{x}.$$

This formulation was implemented by Pumarola et al. [PCPMMN20] and is sufficient to represent any possible deformation. As Park et al. [PSB*21] already stated, it has the disadvantage that rotations are dependent on the input position and the offset vector will vary depending on the distance to the rotation axis. We adopt the method of Park et al. where the network learns an $\mathrm{SE}(3)$ field, i.e. the output of W is a vector $(\mathbf{r},\mathbf{v}) \in \mathbb{R}^6$ describing a rotation with angle $\theta = \|\mathbf{r}\|$ and axis $\hat{\mathbf{r}} = \mathbf{r}/\theta$ and a translation $\mathbf{v}$, which are input into the embedding $\mathcal{S}$. The canonical coordinate $\hat{\mathbf{x}}$ can be computed by $\hat{\mathbf{x}} = e^{[\mathbf{r}]_\times}\mathbf{x} + G\mathbf{v}$, where $e^{[\mathbf{r}]_\times}$ is the matrix exponential which can be found using Rodrigues' formula

$$e^{[\mathbf{r}]_\times} = I + \frac{\sin(\theta)}{\theta}[\mathbf{r}]_\times + \frac{1-\cos(\theta)}{\theta^2}[\mathbf{r}]_\times^2 \tag{5}$$

and

$$G = I + \frac{1-\cos(\theta)}{\theta^2}[\mathbf{r}]_\times + \frac{\theta-\sin(\theta)}{\theta^3}[\mathbf{r}]_\times^2. \tag{6}$$

This means Γ can be composed as $\Gamma = \mathcal{S}\circ W$. Since every single step of the pipeline is differentiable, the network can be trained by feeding the gradient of $\mathcal{L}$ w.r.t. the input of σ into the backward pass of the deformation network (or the embedding to be precise).

3.3. Regularization

Regularization of the deformation field plays a crucial part in the training process. As the loss function is defined via the reprojection error in Eq. (3), which simply sums up attenuation values along each ray γ, expansions and contractions in the deformation field along the given ray have no influence on the actual error. To combat this, Park et al. [PSB*21] use an elastic regularization term, which penalizes singular values $\neq 1$ of the Jacobian J_Γ of Γ for any fixed timestamp t. For $J_\Gamma = U\Sigma V^T$, their formulated regularization term is $\lambda_{\text{elastic}}\|\log\Sigma\|_F^2$. The idea is to minimize the amount of compression and expansion in the deformation field by encouraging the linear part of Γ (given by J_Γ) to be a pure rotation. They use the logarithmic singular values to achieve equal weight of contractions and expansions of the same factor.

This term can also be applied to our scenario. However, since the $L2$ norm is used, small deformations become insignificant in the regularization term. This poses a problem as even small deformations might accumulate and can introduce unwanted effects, e.g. shearing or global deformations. We only want to allow few non-rigid areas in the deformation field and therefore replace the $L2$ with the $L1$ norm. Also, we do not use logarithmic singular values because strong contractions, such as those around the teeth when a patient opens or closes the mouth, would be heavily weighted. So we penalize the absolute deformation, leading to

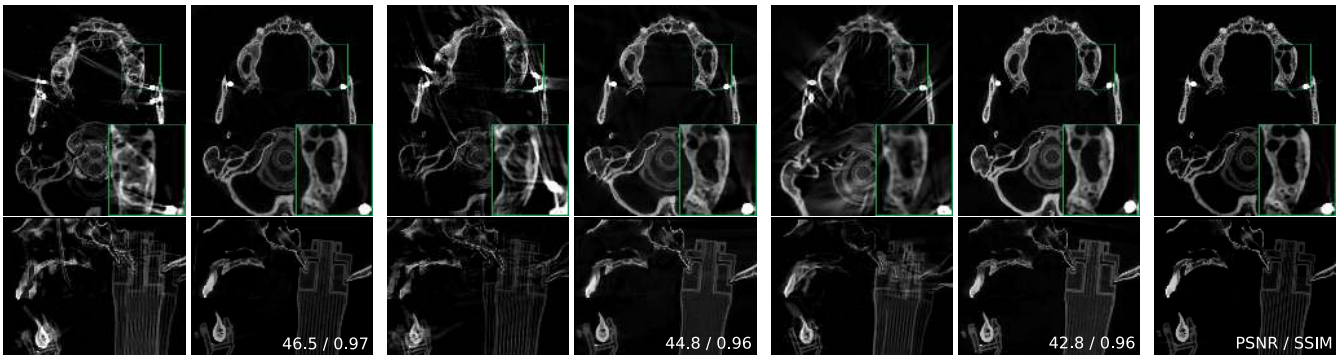
$$L_{\text{elastic-s}}(\mathbf{x}) = \lambda_{\text{elastic}}\|\Sigma - I\|_1,$$

and set $\lambda_{\text{elastic}} = 10^{-3}$. This encourages the network to learn mostly rigid motions overall, as we expect the patient movements in our context to be mostly of rigid nature. In contrast to Park et al. [PSB*21] we do not apply the Geman-McClure robust error function [GM85]. We find that this can leave the network stuck in certain unfavourable states, as (wrongfully) placed strong deformations in early iterations can persist in the final reconstruction. Also, very large gradients of $L_{\text{elastic-s}}$ are inherently prevented by omitting the logarithm and using the $L1$ norm. We weight the regularization term with the density $\sigma(\hat{\mathbf{x}})$ at the current location to penalize deformations in high density matter, such as bones, stronger than in soft tissue or empty space. In addition the output of both σ and W is $L2$ regularized with $\lambda_{L2} = 10^{-6}$ to prevent the networks from overfitting to certain local minima. This value could be tuned for each application, but we have instead opted for a fixed value to minimize per-scenario adjustments. We employ no regularization in the time domain. This can implicitly be achieved by choosing a smaller grid size N_t. Contrary to many other solvers for inverse problems, and some closely related work [RWL*22], we do not explicitly apply a total variation (TV) regularization term, since the network tends towards smooth results anyways.

3.4. Implementation

Our implementation is based on the *instant-ngp* project by Müller et al. [MESK22], which we altered and extended to our needs. This implementation makes use of the *tiny-cuda-nn* framework [Mü21], which provides data structures for fast inference and training times, however requires manual computation of gradients. For the hash grid input encoding of the density network, we adopted some key parameters of the grid, namely $L = 16$ layers, each with feature vectors of size $F = 2$. We chose a hash table size of $T = 2^{19}$, as smaller sizes will lead to reconstructions of diminished spatial resolution and much larger sizes severely slower the training process. These values provide a trade off between reconstruction quality, convergence speed and training efficiency. The original authors provide a thorough investigation of the mentioned meta parameters. As they also steer the total number of training parameters, they directly influence the size of the reconstruction on the disk. With the chosen values the full reconstruction takes up less than 23 MB. The coarsest resolution of the grid $N_{\min}$ was set to 16 and $N_{\max}$ to 1024.

Similar to the original implementation we cull empty space by



(a) *FB+FS* **(b)** *FB+HS* **(c)** *HB+FS* **(d)** *Reference*

Figure 4: *Visual comparison of the moved skull with different scan geometries. (a) Motion-corrected full scan (360°), (b) short scan (186°), (c) half-beam full scan and (d) high-fidelity unmoved reconstruction by vendor. The left column in each scenario shows the provided vendor's reconstruction, the right one ours. We report the PSNR and SSIM compared to the high-fidelity scan.*

using an occupancy grid of size 64^3, which indicates whether the current cell contains any information. We set a threshold of $\sigma(\mathbf{x}) \leq 10^{-1}$. Additionally, empty cells are randomly activated with a chance of 10%. We use a uniform step size to sample points on each ray throughout the whole volume, with random perturbation of the starting point. It is set such that the maximum number of samples per ray is capped at $N_\gamma = 512$. Choosing a bigger number without simultaneously increasing the batch size lowers the convergence rate drastically, setting N_γ too small results in diminished spatial resolution. We use the Adam [KB17] optimizer with a learning rate of 3×10^{-3} for the density and 10^{-2} for the deformation network, which are exponentially decayed during the training process. Additionally, the weights of the density network's MLP are decayed with an L2 regularization of factor 10^{-6}, the deformation network's weights as well as the feature vectors $\mathbf{y}$ with a factor of 10^{-5}. The deformation network is trained with $\beta_1 = 0.9, \beta_2 = 0.99$ and $\varepsilon = 10^{-13}$. For the density network we set $\beta_1 = 0.9, \beta_2 = 0.95$ and decay ε from 10^{-5} to 10^{-15} during the training process. We find that initially using a larger epsilon reduces the resolution in early iterations, as small gradients in higher levels of the hash encoding lose significance, which helps the deformation field in cases of strong movements. Since typical NeRF scenarios take place in the visible light spectrum, mostly only the surface of an object is of interest, enabling very effective early ray termination. In an X-ray setting this is not possible, which inherently increases the average number of samples per ray. For this reason we increased the learning batch size when compared to the original implementation of *instant-ngp* to 2^{19} to achieve a reasonable number of rays per batch.

Network architecture Both our networks are very slim, fully connected MLPs with 64 neurons per layer. We use only three hidden layers in both networks and the ReLU activation function. The deformation network features a linear output layer and the density network's output is passed through a Squareplus [Bar21] output activation to ensure meaningful positive density values.

	Whole head		Accuitomo 170		Local Tomo	
Motion	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
None	39.1	0.99	39.2	0.98	31.5	0.96
Cont.	31.2	0.94	29.7	0.92	28.2	0.89
Sudden	35.5	0.97	35.5	0.97	31.7	0.94

Table 1: *Comparison between our results and the ground truth using the synthetically generated data sets. The case without motion provides a baseline for the reconstruction quality.*

4. Results

4.1. Maxillofacial CBCT

In this section we evaluate our algorithm on synthetic and actual scan data in maxillofacial and oral applications. All results presented in this section were obtained after training for 35'000 iterations.

4.1.1. Synthetic data

We applied our algorithm to synthetically generated CBCT projection data of a parameterized head model, which were generated by Birklein et. al [BNS*23]. During the simulated scan, the model undergoes rigid motion of the cranium while simultaneously opening its mouth. Forwards projections were performed using two different motion profiles, one where the patient performs a continuous, high frequency motion of the cranium (15° rotation and 6 mm translation) while slowly opening the mouth (5 mm) and one of a sudden motion of the cranium (3° rotation, 2 mm translation) followed by a sudden mouth opening (also 5 mm). The exact motion pattern for the first scenario can be seen in [BNS*23], Fig. 5. The second scenario resembles the motion pattern observed in our clinical scenarios more closely. Tab. 1 compares the motion corrected reconstruction to the ground truth. The first row of each scenario is an unmoved reconstruction and provides a baseline to the reconstruction quality. We evaluate three different device configurations, first, one where the whole head is in the FOV in every frame, the second one resembles the configuration with the widest FOV of the Accuitomo

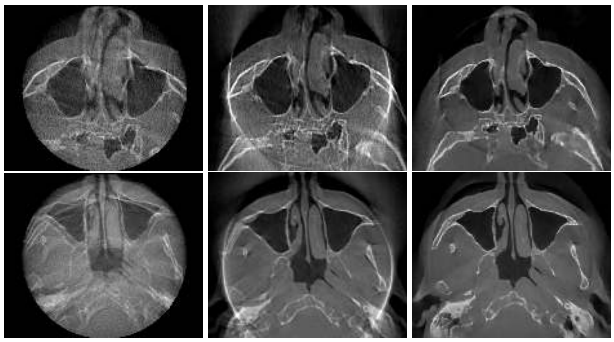
170 device and the third one produces local tomography scans with truncation effects, since only a small part of the patient can be seen in each image. In all test cases the result is of high visual quality and free of motion artifacts. A single sudden movement, such as a patient nodding, poses no problem for our algorithm and both metrics used in Tab. 1 come very close to the baseline reconstruction. It is even possible to compensate strong motions in narrow local tomography scenarios. The motion patterns of the continuous motion test cases were created by using a random walk for each of the six degrees of freedom, inducing very rapid changes in motions. Analogous to Park et al. [PSB*21] we find our method struggles with such rapid changes of motions, which explains the worse SSIM and PSNR values in the continuous motion test profiles.

4.1.2. Scan data

To validate the proposed method on actual acquisitions, which always contain some amount of noise and scatter in practice, we apply our algorithm to scans of a controllable moving skull in various scenarios as well as to motion-impaired patient data from a clinical application. In all acquisitions from real CBCT machines we also optimize for the emitted energy, i.e. I_0 from Eq. (1) to account for slight calibration errors.

Movable skull We placed a skull mounted on a Stewart platform inside two different CBCT machines. The first one is a 3D Accuitomo 170 device and the second one a KaVo OP 3D Vision. Additionally to rigid cranial motions, this skull also supports separate motions of the lower jaw. Fig. 4d shows the vendor reconstruction of an unmoved, high fidelity scan (900 images) which we use as reference. For the results shown in Fig. 4 (a)-(c) we simulated a single strong nodding motion of the patient (about 1.2 cm in total) combined with a sudden motion of the lower jaw (about 7 mm) before employing our algorithm. The data for Fig. 4a was acquired using the Accuitomo 170 standard protocol (full beam, 512 images). In our reconstruction no motion artifacts are visible at all. Also notice how the bright metal spring in the highlighted area shows less artefacts than in the reference reconstruction. Fig. 4b was created using the short scan protocol (full beam, 265 images) of the same device. Here the machine only performs half a rotation around the patient. Our reconstruction still shows some blurriness, which is for example visible in the highlighted box, but does not show motion artefacts like streaks or double contours, as the vendor's does. Lastly, Fig. 4c was acquired using a device which pursues a lateral-offset detector strategy, that is employing a full rotation around the patient using a half-beam geometry, such that each image contains (about) half the skull. The effect of the skull's motion can very clearly be seen in the top row of Fig. 4c, as the reconstruction seems to be cut in half. Our reconstruction does not show this artefact anymore. In a previous work [BNS*23] correcting motions in setups (b) and (c) was not yet possible.

Clinical application Fig. 5 shows reconstructions in actual clinical applications. In the left column one can see reconstruction of the device's vendor (Fig. 5a), the middle one shows the motion-corrected result of Birklein et al. [BNS*23], and in the right images our result (Fig. 5c) can be seen. Our algorithm was capable of removing visible motion artifacts in both setups, which appear in the left reconstructions very noticeably as double contours and streaks, especially in the bottom image. To some extent they still appear in the middle column, e.g. the zygomatic bone in the second example still shows double contours. The first patient performed a slight motion towards the end of the scan, while the strong artifacts in the second scenario are induced by a sudden forwards-motion of the patient of about 7 mm halfway during the scan. Also notice the low level of noise of the proposed reconstruction method.



(a) *Vendor* **(b)** *Birklein et al.* **(c)** *Ours*

Figure 5: *Slices of two different motion-impaired clinical CBCT acquisitions. Top row shows the reconstruction of a short scan (186°), bottom row a full scan.*

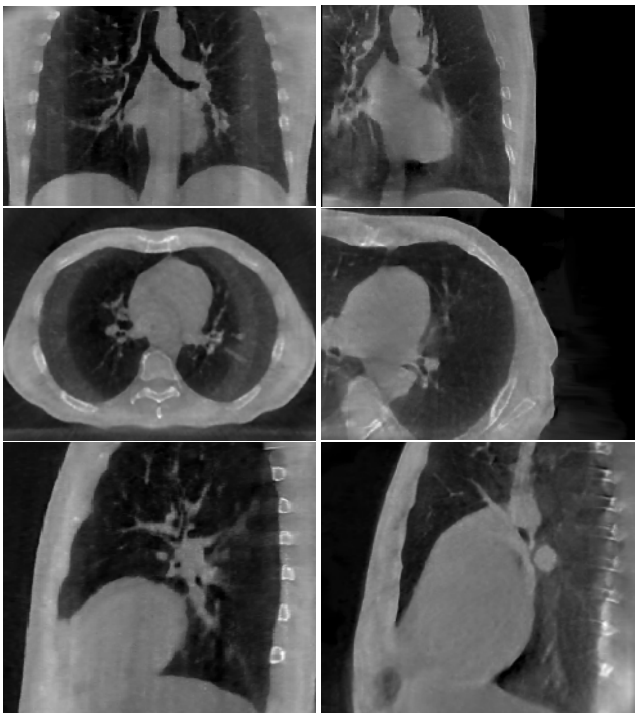
Runtimes The runtime of our algorithm is independent from the data set since we fix the training batch size and number of iterations. It is mostly dominated by the elastic regularization term, as its computation requires additional backwards passes of the deformation network. We obtained the presented results after 25 mins on an Nvidia RTX 3090 GPU.

4.2. Applications beyond maxillofacial CBCT

In this section we explore further applications of the proposed method, as motion correction and deformable CBCT is not only of interest in maxillofacial and oral scenarios. Since we do not expect much rigid motion but rather non-rigid deformations, e.g. breathing motion in 4D lung scans, we do not apply the elastic regularization term in any of the following results. This also shortens the runtime considerably to less than 10 minutes.

4.2.1. 4D-CBCT

In lung cancer treatment, 4D-CBCT is a technique typically employed to provide accurate tumor localization in the space-time domain by binning the 2D projection images into different breathing phases (typically 10) of the patient [SZRvH05] and performing reconstructions of each phase. Compared to traditional 3D-CBCT this can mitigate artifacts introduced by the breathing motion and additionally makes it possible to track moving tissue over time. If applied naively however it leads to very strong streaking artifacts due to the limited number of projections per phase. For this reason, 4D-CBCT scan times are typically much longer than in 3D-CBCT [BAMR*19]. In the recent past a lot of effort went into the



(a) *Clinical Varian* **(b)** *Clinical Elekta*

Figure 6: *Start of the inhale phase of real one minute 4D-CBCT lung scans. The data sets were part of the SPARE Challenge [SGL*19]. 4D animations are provided in the supplemental material of this paper.*

development of algorithms capable of producing high quality results with short scan times. Some methods make use of a priori motion models computed from previously acquired 4D planning CTs [RWvHS09, MJR16], while others match the reconstructions of the different phases to a common fixed frame [RCWH18, WG13], which is more similar to our deformation approach.

With our method, binning the projections into different phases is not strictly necessary, as our deformation field is contiguous in the time domain and is being optimized simultaneously with the reconstruction. It is possible to make use of the respiratory phase signal by sorting the acquisition images according to this signal. We find that this measure can improve the quality of the results, as it ensures a minimal deformation between two consecutive images. Many devices employ a so called bowtie filter, effectively reducing radiation dose in peripheries of the patient [YRLL19]. As our method depends on the projection error of the reconstruction, it is necessary to have an approximate knowledge of the attenuation characteristics of the filter used. In all our test cases we could estimate those from the projection data. Fig. 6 shows two examples of clinical 4D-CBCTs of the lung. The images depict slices of the start of the inhale cycle. Fig. 6a stems from a one minute scan with a full rotation half-beam geometry (680 projections), Fig. 6b from a half scan with full-beam geometry (340 projections).

We also performed a quantitative evaluation of our approach using the Monte-Carlo simulated acquisitions from the SPARE challenge [SGL*19]. We find that the reconstruction quality of our approach is, on average, better than all other algorithms tested, both in terms of RMSE as well as SSIM, in three out of four categories. In the very narrow planning target volume (PTV) category, our method struggles a bit but creates results of comparable quality nevertheless. As Fig. 7 shows, the spatio-temporal accuracy of our method is state-of-the-art. Please note that we do not make use of any priors (as some other methods do) but are purely data driven.

	Body	Lung	PTV	Bone
RMSE	1.662	1.569	1.621	1.576
$\times 10^{-3}$	0.367	0.282	0.301	0.320
SSIM	0.766	0.716	0.774	0.691
	0.053	0.068	0.054	0.085

Table 2: *Mean (first value) and standard deviation (second value) of RMSE and SSIM values of the SPARE challenge benchmark.*

Adjustments To ensure small motions between consecutive frames we use the provided phase signal and sort the projection images accordingly. In the input encoding for the deformation field we decrease the number of cells N_t used for the time coordinate t to 10. This number is inspired by the common binning process in 4D-CBCT and leads to a smoother deformation field (along the time axis), therefore resulting in a smoother breathing motion. At the same time the learning rate of the deformation network is decreased to 10^{-3} and the learning rate of the feature vectors $\mathbf{y}$ is further decreased by a factor of 5×10^{-2} compared to the MLP's. This increases the stability, especially in cases with high rates of scatter or in low dose simulations.

4.2.2. Deformable CBCT in non-medical applications

4D-CBCT also has use cases outside of the medical field. Zang et al. [ZIT*18, ZIT*19] explore applications of deformable CT and develop two successive algorithms to solve this problem using their altered acquisition strategy. We applied our algorithm to the rose dataset, which was kindly made public. Note that we had to use a subsampled version (factor ½ for width and height, respectively, and skipping every second frame), since the full resolution would not fit in our GPU VRAM.

Fig. 8 shows the same slice of the reconstruction at three different timestamps from left to right: start of the acquisition, halfway through the acquisition and at the end. Our simple formulation of only a deformation field and a reference volume is sufficient to create a result of high quality in this case. In this scenario the elastic regularization is not applied either. Please note that it is not possible to accurately represent topological changes in the object, such as leaves that are fused at the end of the scan but not at the beginning.

5. Limitations

We model the 4D scene as a reference 3D volume and a 4D deformation field. As both are continuous and differentiable in each coordinate, our formulation is not able to represent topological

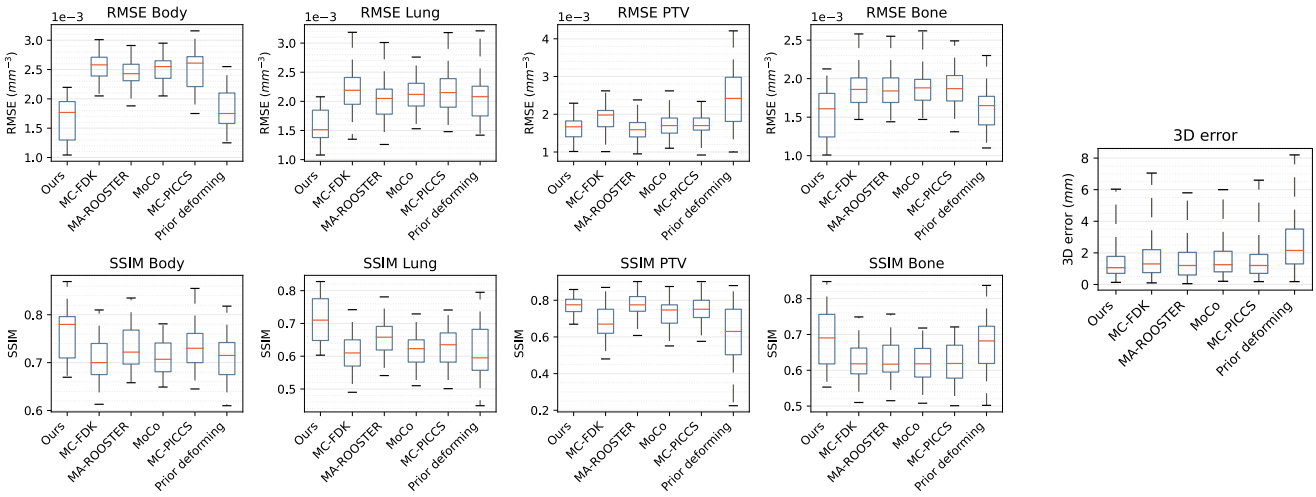


Figure 7: *Quantitative evaluation of 4D-CBCT scans of Monte-Carlo simulated data sets compared to other algorithms (we adapted Fig. 5 and 7(a) from Shieh et al. [SGL*19]). Left hand side: RMSE and SSIM values compared to the ground truth. Right hand side: magnitude of the translation error of the PTV area when aligning to the ground truth. Results of our method are leftmost.*

changes which might happen during a scan, as for example in Sec. 4.2.2. Park et al. [PSH*21] combat this by introducing the time component into the density network through a "slicing network". First experiments showed that this might also be possible with our approach, however we could not create consistent results, as the timing of when to introduce these coordinates into the network seems to be crucial and may vary from one setting to another. We hope to overcome this limitation in future research. Fig. 9 shows the reconstruction of a highly viscous fluid at the beginning and end of the scan, respectively. Although the reconstructions appear surprisingly convincing at first glance, in cases of such extreme deformation using only one reference volume does not seem to be enough, as they still contain many artifacts (notice the "swirls" in the bottom part of the left image).

Although we discourage non-rigid deformations in maxillofacial applications by applying the elastic regularization term, the deformation network may find deformations which are not present in the data (see Fig. 3 for example). This can happen for different reasons, including truncation effects due to local tomography or physical effects like noise, beam hardening, or scatter, which are explicitly ignored in the forward projection. Also, miscalibrations can play a role, e.g., if I_0 from Eq. (1) is set wrong and our optimization cannot recover it. While our method generally applies to local tomography, these deformations seem more pronounced in scenarios with smaller ROIs. Suppose present motion is expected, or even known, to be solely rigid. In that case, it can easily be incorporated into the process by setting the number of frequencies for the input encoding of Γ to 0.

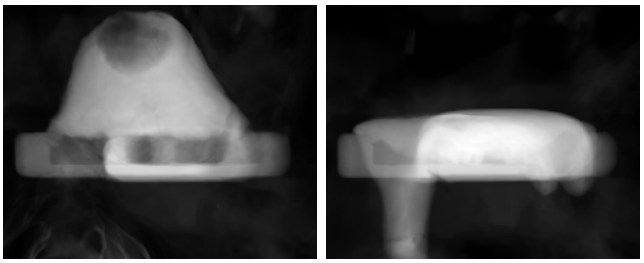


Figure 9: *Reconstruction of a highly viscous fluid undergoing a strong deformation during the acquisition. This data set was also created by Zhang et al. [ZIT*18]*

Figure 8: *Deforming 4D-CBCT scan of a wilting rose scanned over the course of nine hours. For the sake of visualization, the colors have been inverted. The data set is provided by Zhang et al. [ZIT*18].*

6. Conclusion

We presented a novel method for deformable cone beam CT based on neural inverse rendering, applicable to a number of different scenarios. Our approach requires no prior information and produces 4D reconstructions of equal or better quality than task specific reconstruction methods in a matter of minutes. We could demonstrate its effectiveness in dental, 4D lung as well as industrial CBCT settings, both on synthetic and real-world data, showcasing a great flexibility.

References

[AK84] ANDERSEN A., KAK A.: Simultaneous algebraic reconstruction technique (sart): A superior implementation of the art algorithm. *Ultrasonic Imaging 6*, 1 (1984), 81–94. URL: https://www.sciencedirect.com/science/article/pii/0161734684900087, doi:https://doi.org/10.1016/0161-7346(84)90008-7. 2

[BAMR*19] BRYCE-ATKINSON A., MARCHANT T., RODGERS J., BUDGELL G., MCWILLIAM A., FAIVRE-FINN C., WHITFIELD G., VAN HERK M.: Quantitative evaluation of 4d cone beam ct scans with reduced scan time in lung cancer patients. *Radiotherapy and Oncology 136* (2019), 64–70. URL: https://www.sciencedirect.com/science/article/pii/S0167814019301495, doi:https://doi.org/10.1016/j.radonc.2019.03.027. 6

[Bar21] BARRON J. T.: Squareplus: A softplus-like algebraic rectifier, 2021. arXiv:2112.11687. 5

[BMA*16] BERGER M., MÜLLER K., AICHERT A., UNBERATH M., THIES J., CHOI J.-H., FAHRIG R., MAIER A.: Marker-free motion correction in weight-bearing cone-beam ct of the knee joint. *Medical Physics 43* (03 2016), 1235–1248. doi:10.1118/1.4941012. 2

[BNS*23] BIRKLEIN L., NIEBLER S., SCHÖMER E., BRYLKA R., SCHWANECKE U., SCHULZE R.: Motion correction for separate mandibular and cranial movements in cone beam ct reconstructions. *Medical Physics 50*, 6 (2023), 3511–3525. URL: https://aapm.onlinelibrary.wiley.com/doi/abs/10.1002/mp.16347, arXiv:https://aapm.onlinelibrary.wiley.com/doi/pdf/10.1002/mp.16347, doi:https://doi.org/10.1002/mp.16347. 2, 5, 6

[BXA*17] BERGER M., XIA Y., AICHINGER W., MENTL K., UNBERATH M., AICHERT A., RIESS C., HORNEGGER J., FAHRIG R., MAIER A.: Motion compensation for cone-beam ct using fourier consistency conditions. *Physics in Medicine & Biology 62*, 17 (aug 2017), 7181. URL: https://dx.doi.org/10.1088/1361-6560/aa8129, doi:10.1088/1361-6560/aa8129. 2

[CLI*20] CHABRA R., LENSSEN J. E., ILG E., SCHMIDT T., STRAUB J., LOVEGROVE S., NEWCOMBE R. A.: Deep local shapes: Learning local SDF priors for detailed 3d reconstruction. *CoRR abs/2003.10983* (2020). URL: https://arxiv.org/abs/2003.10983, arXiv:2003.10983. 2

[FBSK14] FLACH B., BREHM M., SAWALL S., KACHELRIESS M.: Deformable 3d–2d registration for CT and its application to low dose tomographic fluoroscopy. *Physics in Medicine and Biology 59*, 24 (nov 2014), 7865–7887. URL: https://doi.org/10.1088/0031-9155/59/24/7865, doi:10.1088/0031-9155/59/24/7865. 2

[FDK84] FELDKAMP L. A., DAVIS L. C., KRESS J. W.: Practical cone-beam algorithm. *J. Opt. Soc. Am. A 1*, 6 (Jun 1984), 612–619. URL: https://opg.optica.org/josaa/abstract.cfm?URI=josaa-1-6-612, doi:10.1364/JOSAA.1.000612. 2

[GBH70] GORDON R., BENDER R., HERMAN G. T.: Algebraic reconstruction techniques (art) for three-dimensional electron microscopy and x-ray photography. *Journal of Theoretical Biology 29*, 3 (1970), 471–481. URL: https://www.sciencedirect.com/science/article/pii/0022519370901098, doi:https://doi.org/10.1016/0022-5193(70)90109-8. 2

[GM85] GANAN S., MCCLURE D.: Bayesian image analysis: An application to single photon emission tomography. *Amer. Statist. Assoc* (1985), 12–18. 4

[HBR*17] HAHN J., BRUDER H., ROHKOHL C., ALLMENDINGER T., STIERSTORFER K., FLOHR T., KACHELRIESS M.: Motion compensation in the region of the coronary arteries based on partial angle reconstructions from short scan ct data. *Medical Physics 44* (08 2017). doi:10.1002/mp.12514. 2

[HDO*13] HANZELKA T., DUSEK J., OCASEK F., KUCERA J., SEDY J., BENES J., PAVLIKOVA G., FOLTAN R.: Movement of the patient and the cone beam computed tomography scanner: objectives and possible solutions. *Oral Surgery, Oral Medicine, Oral Pathology and Oral Radiology 116*, 6 (2013), 769–773. doi:https://doi.org/10.1016/j.oooo.2013.08.010. 2

[Hub64] HUBER P. J.: Robust Estimation of a Location Parameter. *The Annals of Mathematical Statistics 35*, 1 (1964), 73 – 101. URL: https://doi.org/10.1214/aoms/1177703732, doi:10.1214/aoms/1177703732. 3

[JSM*20] JIANG C. M., SUD A., MAKADIA A., HUANG J., NIESSNER M., FUNKHOUSER T. A.: Local implicit grid representations for 3d scenes. *CoRR abs/2003.08981* (2020). URL: https://arxiv.org/abs/2003.08981, arXiv:2003.08981. 2

[KB17] KINGMA D. P., BA J.: Adam: A method for stochastic optimization, 2017. arXiv:1412.6980. 5

[LCM*22] LIU J.-W., CAO Y.-P., MAO W., ZHANG W., ZHANG D. J., KEPPO J., SHAN Y., QIE X., SHOU M. Z.: Devrf: Fast deformable voxel radiance fields for dynamic scenes. *arXiv preprint arXiv:2205.15723* (2022). 2

[LGL*20] LIU L., GU J., LIN K. Z., CHUA T.-S., THEOBALT C.: Neural sparse voxel fields. *NeurIPS* (2020). 2

[MESK22] MÜLLER T., EVANS A., SCHIED C., KELLER A.: Instant neural graphics primitives with a multiresolution hash encoding. *ACM Trans. Graph. 41*, 4 (July 2022), 102:1–102:15. URL: https://doi.org/10.1145/3528223.3530127, doi:10.1145/3528223.3530127. 2, 3, 4

[MJR16] MORY C., JANSSENS G., RIT S.: Motion-aware temporal regularization for improved 4d cone-beam computed tomography. *Physics in Medicine & Biology 61*, 18 (sep 2016), 6856. URL: https://dx.doi.org/10.1088/0031-9155/61/18/6856, doi:10.1088/0031-9155/61/18/6856. 2, 7

[MST*21] 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 (dec 2021), 99–106. URL: https://doi.org/10.1145/3503250, doi:10.1145/3503250. 2, 3

[Mü21] MÜLLER T.: tiny-cuda-nn, 4 2021. URL: https://github.com/NVlabs/tiny-cuda-nn. 4

[NST*19] NIEBLER S., SCHÖMER E., TJADEN H., SCHWANECKE U., SCHULZE R.: Projection-based improvement of 3d reconstructions from motion-impaired dental cone beam ct data. *Medical Physics 46* (07 2019). doi:10.1002/mp.13731. 2

[OJS*17] OUADAH S., JACOBSON M., STAYMAN J. W., EHTIATI T., WEISS C., SIEWERDSEN J. H.: Correction of patient motion in cone-beam CT using 3d–2d registration. *Physics in Medicine & Biology 62*, 23 (nov 2017), 8813–8831. URL: https://doi.org/10.1088/1361-6560/aa9254, doi:10.1088/1361-6560/aa9254. 2

[PCPMMN20] PUMAROLA A., CORONA E., PONS-MOLL G., MORENO-NOGUER F.: D-NeRF: Neural Radiance Fields for Dynamic Scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2020). 2, 4

[PMM*19] PREUHS A., MAIER A., MANHART M., KOWARSCHIK M., HOPPE-PREUHS E., FOTOUHI J., NAVAB N., UNBERATH M.: Symmetry prior for epipolar consistency. *International Journal of Computer Assisted Radiology and Surgery 14* (07 2019). doi:10.1007/s11548-019-02027-8. 2

[PSB*21] PARK K., SINHA U., BARRON J. T., BOUAZIZ S., GOLDMAN D. B., SEITZ S. M., MARTIN-BRUALLA R.: Nerfies: Deformable neural radiance fields. *ICCV* (2021). 2, 3, 4, 6

[PSH*21] PARK K., SINHA U., HEDMAN P., BARRON J. T., BOUAZIZ S., GOLDMAN D. B., MARTIN-BRUALLA R., SEITZ S. M.: Hypernerf: A higher-dimensional representation for topologically varying neural radiance fields. *ACM Trans. Graph. 40*, 6 (dec 2021). 8

[RBA*18] RAHAMAN N., BARATIN A., ARPIT D., DRÄXLER F., LIN

M., HAMPRECHT F. A., BENGIO Y., COURVILLE A. C.: On the spectral bias of neural networks. In *International Conference on Machine Learning* (2018). 3

[RBSF13] ROHKOHL C., BRUDER H., STIERSTORFER K., FLOHR T.: Improving best-phase image quality in cardiac ct by motion correction with mam optimization. *Medical Physics 40*, 3 (2013), 031901. URL: https://aapm.onlinelibrary.wiley.com/doi/abs/10.1118/1.4789486, arXiv:https://aapm.onlinelibrary.wiley.com/doi/pdf/10.1118/1.4789486, doi:https://doi.org/10.1118/1.4789486. 2

[RCWH18] RIBLETT M. J., CHRISTENSEN G. E., WEISS E., HUGO G. D.: Data-driven respiratory motion compensation for four-dimensional cone-beam computed tomography (4d-cbct) using groupwise deformable registration. *Medical Physics 45*, 10 (2018), 4471–4482. URL: https://aapm.onlinelibrary.wiley.com/doi/abs/10.1002/mp.13133, arXiv:https://aapm.onlinelibrary.wiley.com/doi/pdf/10.1002/mp.13133, doi:https://doi.org/10.1002/mp.13133. 2, 7

[RKA*21] REED A. W., KIM H., ANIRUDH R., MOHAN K., CHAMPLEY K., KANG J., JAYASURIYA S.: Dynamic ct reconstruction from limited views with implicit neural representations and parametric motion fields. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)* (Los Alamitos, CA, USA, oct 2021), IEEE Computer Society, pp. 2238–2248. URL: https://doi.ieeecomputersociety.org/10.1109/ICCV48922.2021.00226, doi:10.1109/ICCV48922.2021.00226. 2

[RWL*22] RÜCKERT D., WANG Y., LI R., IDOUGHI R., HEIDRICH W.: Neat: Neural adaptive tomography. *ACM Trans. Graph. 41*, 4 (jul 2022). URL: https://doi.org/10.1145/3528223.3530121, doi:10.1145/3528223.3530121. 2, 4

[RWvHS09] RIT S., WOLTHAUS J. W. H., VAN HERK M., SONKE J.-J.: On-the-fly motion-compensated cone-beam ct using an a priori model of the respiratory motion. *Medical Physics 36*, 6Part1 (2009), 2283–2296. URL: https://aapm.onlinelibrary.wiley.com/doi/abs/10.1118/1.3115691, arXiv:https://aapm.onlinelibrary.wiley.com/doi/pdf/10.1118/1.3115691, doi:https://doi.org/10.1118/1.3115691. 7

[SGL*19] SHIEH C.-C., GONZALEZ Y., LI B., JIA X., RIT S., MORY C., RIBLETT M., HUGO G., ZHANG Y., JIANG Z., LIU X., REN L., KEALL P.: Spare: Sparse-view reconstruction challenge for 4d cone-beam ct from a 1-min scan. *Medical Physics 46*, 9 (2019), 3799–3811. URL: https://aapm.onlinelibrary.wiley.com/doi/abs/10.1002/mp.13687, arXiv:https://aapm.onlinelibrary.wiley.com/doi/pdf/10.1002/mp.13687, doi:https://doi.org/10.1002/mp.13687. 7, 8

[SHG*11] SCHULZE R., HEIL U., GROSS D., BRUELLMANN D., DRANISCHNIKOW E., SCHWANECKE U., SCHOEMER E.: Artefacts in cbct: a review. *Dentomaxillofac Radiol 40 5* (2011), 265–273. 2

[SJP*21] SUN T., JACOBS R., PAUWELS R. J., TIJSKENS E., FULTON R. R., NUYTS J.: A motion correction approach for oral and maxillofacial cone-beam CT imaging. *Physics in Medicine & Biology* (apr 2021). URL: https://doi.org/10.1088/1361-6560/abfa38, doi:10.1088/1361-6560/abfa38. 2

[SLX*21] SUN Y., LIU J., XIE M., WOHLBERG B., KAMILOV U. S.: Coil: Coordinate-based internal learning for imaging inverse problems, 2021. arXiv:2102.05181. 2

[SNCS*18] SPIN-NETO R., COSTA C., SALGADO D., ZAMBRANA N., GOTFREDSEN E., WENZEL A.: Head motion during cone-beam computed tomography: Analysis of frequency and influence on image quality. *Dentomaxillofac Radiol*, 47 (2018), 20170216. 2

[SNMS*15] SPIN-NETO R., MATZEN L., SCHROPP L., GOTFREDSEN E., WENZEL A.: Factors affecting patient movement and re-exposure in cbct examination. *Oral Surgery, Oral Medicine, Oral Pathology and Oral Radiology 119* (02 2015). doi:10.1016/j.oooo.2015.01.011. 2

[SNMS*17] SPIN-NETO R., MATZEN L. H., SCHROPP L., GOTFREDSEN E., WENZEL A.: Detection of patient movement during cbct examination using video observation compared with an accelerometer-gyroscope tracking system. *Dentomaxillofacial Radiology 46*, 2 (2017), 20160289. doi:10.1259/dmfr.20160289. 2

[SNW16] SPIN-NETO R., WENZEL A.: Patient movement and motion artefacts in cone beam computed tomography of the dentomaxillofacial region: a systematic literature review. *Oral Surg Oral Med Oral Pathol Oral Radiol*, 121 (2016), 425–433. 2

[SPX23] SHEN L., PAULY J., XING L.: Nerp: Implicit neural representation learning with prior embedding for sparsely sampled image reconstruction, 2023. arXiv:2108.10991. 2

[SSY*17] SISNIEGA A., STAYMAN J., YORKSTON J., SIEWERDSEN J., ZBIJEWSKI W.: Motion compensation in extremity cone-beam ct using a penalized image sharpness criterion. *Physics in medicine and biology 62 9* (2017), 3712–3734. 2

[SZRvH05] SONKE J.-J., ZIJP L., REMEIJER P., VAN HERK M.: Respiratory correlated cone beam ct. *Medical Physics 32*, 4 (2005), 1176–1186. URL: https://aapm.onlinelibrary.wiley.com/doi/abs/10.1118/1.1869074, arXiv:https://aapm.onlinelibrary.wiley.com/doi/pdf/10.1118/1.1869074, doi:https://doi.org/10.1118/1.1869074. 6

[TJJ*15] TADINADA A., JALALI E., JADHAV A., SCHINCAGLIA G. P., YADAV S.: Artifacts in cone beam computed tomography image volumes: An illustrative depiction. *Journal of the Massachusetts Dental Society 64* (07 2015), 12–5. 2

[TLY*21] TAKIKAWA T., LITALIEN J., YIN K., KREIS K., LOOP C. T., NOWROUZEZAHRAI D., JACOBSON A., MCGUIRE M., FIDLER S.: Neural geometric level of detail: Real-time rendering with implicit 3d shapes. *CoRR abs/2101.10994* (2021). URL: https://arxiv.org/abs/2101.10994, arXiv:2101.10994. 2

[TSM*20] TANCIK M., SRINIVASAN P. P., MILDENHALL B., FRIDOVICH-KEIL S., RAGHAVAN N., SINGHAL U., RAMAMOORTHI R., BARRON J. T., NG R.: Fourier features let networks learn high frequency functions in low dimensional domains. *NeurIPS* (2020). 3

[WG13] WANG J., GU X.: Simultaneous motion estimation and image reconstruction (smeir) for 4d cone-beam ct. *Medical Physics 40*, 10 (2013), 101912. URL: https://aapm.onlinelibrary.wiley.com/doi/abs/10.1118/1.4821099, arXiv:https://aapm.onlinelibrary.wiley.com/doi/pdf/10.1118/1.4821099, doi:https://doi.org/10.1118/1.4821099. 7

[YRLL19] YANG K., RUAN C., LI X., LIU B.: Data of ct bow tie filter profiles from three modern ct scanners. *Data in Brief 25* (2019), 104261. URL: https://www.sciencedirect.com/science/article/pii/S2352340919306158, doi:https://doi.org/10.1016/j.dib.2019.104261. 7

[ZIT*18] ZANG G., IDOUGHI R., TAO R., LUBINEAU G., WONKA P., HEIDRICH W.: Space-time tomography for continuously deforming objects. *ACM Trans. Graph. 37*, 4 (jul 2018). URL: https://doi.org/10.1145/3197517.3201298, doi:10.1145/3197517.3201298. 2, 7, 8

[ZIT*19] ZANG G., IDOUGHI R., TAO R., LUBINEAU G., WONKA P., HEIDRICH W.: Warp-and-project tomography for rapidly deforming objects. *ACM Trans. Graph. 38*, 4 (jul 2019). URL: https://doi.org/10.1145/3306346.3322965, doi:10.1145/3306346.3322965. 2, 7

[ZSPM23] ZHANG Y., SHAO H.-C., PAN T., MENGKE T.: Dynamic cone-beam ct reconstruction using spatial and temporal implicit neural representation learning (stinr). *Physics in Medicine & Biology 68*, 4 (feb 2023), 045005. URL: https://dx.doi.org/10.1088/1361-6560/acb30d, doi:10.1088/1361-6560/acb30d. 2