Chasing clouds: Differentiable volumetric rasterisation of point clouds as a highly efficient and accurate loss for large-scale deformable 3D registration

Mattias P. Heinrich[1] Alexander Bigalke[1] Christoph Großbröhmer[1] Lasse Hansen[2]

[1]Institute of Medical Informatics, University of Lübeck [2]EchoScout GmbH Germany

mattias.heinrich@uni-luebeck.de lasse@echoscout.ai

Abstract

*Learning-based registration for large-scale 3D point clouds has been shown to improve robustness and accuracy compared to classical methods and can be trained without supervision for locally rigid problems. However, for tasks with highly deformable structures, such as alignment of pulmonary vascular trees for medical diagnostics, previous approaches of self-supervision with regularisation and point distance losses have failed to succeed, leading to the need for complex synthetic augmentation strategies to obtain reliably strong supervision. In this work, we introduce a novel **Di**fferentiable **V**olumetric **R**asterisation **o**f point **C**louds (DiVRoC) that overcomes those limitations and offers a highly efficient and accurate loss for large-scale deformable 3D registration. DiVRoC drastically reduces the computational complexity for measuring point cloud distances for high-resolution data with over 100k 3D points and can also be employed to extrapolate and regularise sparse motion fields, as loss in a self-training setting and as objective function in instance optimisation. DiVRoC can be successfully embedded into geometric registration networks, including PointPWC-Net and other graph CNNs. Our approach yields new state-of-the-art accuracy on the challenging PVT dataset in three different settings without training with manual ground truth: 1) unsupervised metric-based learning 2) self-supervised learning with pseudo labels generated by self-training and 3) optimisation based alignment without learning.*
https://github.com/mattiaspaul/ChasingClouds

1. Introduction

Learning-based point cloud or shape registration has seen increasing interest in computer vision and medical imaging over recent years [56] due to its ability to model 3D motion for sequences of (multi-view) depth images or LiDAR measurements within a dynamic scene as well as for surface and tree-like representations in medical 3D scans. Point clouds offer benefits of representing a 3D scene more sparsely and efficiently than voxelised volumes and they immensely reduce privacy concerns of data sharing since identifying intensity information is removed. However, they come with the difficulty of defining convolutional operators, neighbourhood information and loss functions without a regular grid. Many recent works define graph convolutional operators to solve point cloud classification and semantic segmentation tasks [4, 19, 45, 48, 57].

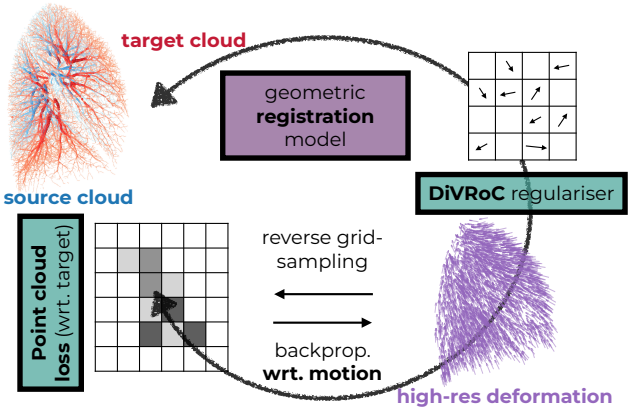

Figure 1. Overview of our key contributions for unsupervised learning of highly deformable 3D point cloud registration.

For scene flow estimation of predominantly rigid objects that move dynamically a number of geometric network approaches for self-supervised and supervised learning of correspondences have been proposed: flow embedding layers (FlowNet3D) [21], bilateral convolutions [13], optimal transport based solutions (FLOT) [34], cost volume based methods with multiple warping steps (PointPWC-Net) [49], iterative approaches (PV-RAFT) [47].

While scene flow estimation usually requires high frame rate throughput on moderately sized point clouds, 3D medical imaging for diagnosis mandates registration accuracies at millimetre-scale on high-resolution clouds but may al-

low for computation times of hundreds of milliseconds to around 1 second even in interventional procedures. For related work in medical registration and lung imaging the use of instance optimisation as a refinement to a robust network prediction has thus become imminent [14, 39, 38] to compensate subtle domain shifts that are introduced by the patient specific anatomy, e.g. the topology of vascular trees is highly varying across a population.

To date, few methods have successfully addressed the registration of highly deformable geometric structures e.g. the lungs, as discussed in the recent NeurIPS paper on robust optimal transport (RobOT) [38]. The authors in particular state that a key lesson from their experiments and ablation studies on the new pulmonary vascular tree (PVT) dataset on respiratory motion estimation was that metric-based unsupervised learning did not yield to a competitive level of accuracy despite exploring a variety of point cloud metrics, including local Laplacian matching [49] as well as Chamfer and Wasserstein distances [10, 28]. They argue that this is due to the complex geometric structure of the lung anatomy and the significant differences in shape found between inspiration and exhale scans that are usually not present in surface meshes or other clean point clouds that unsupervised methods were primarily designed for and evaluated on. RobOT therefore resorts to a sophisticated synthetic simulation of deformable lung motion to provide strong supervision. They propose a multistep approach that comprises affine pre-alignment, PointPWC flow prediction and robust optimal transport with post-processing to set a state-of-the-art accuracy of 2.86mm, which clearly outperforms classical optimisation based methods.

Motivation: We argue that the difficulties of previous attempts to learn point cloud registration of intrinsically deformable structures without stronger supervision stems from three problems of related works. First, using a geometric network that can freely predict sparse displacements with only a soft penalty on the regularity of the deformation makes unsupervised learning with a cost function that exhibits many local minima nearly infeasible. Hence, a stronger link between regularisation and network prediction has to be established. Second, point-based metrics can be prone to sparse differentiability with respect to the hyper-parameter choice, e.g. the nearest neighbours for the Chamfer distance or blur in the sinkhorn metric used in optimal transport [11]. Third, the importance of highly efficient instance optimisation to compensate differences between the learned population model and patient-specific graph topology has so far been partially overlooked.

In this work, we overcome those difficulties by presenting a novel differentiable volumetric rasterisation (DiVRoC) for 3D point cloud losses, which substantially reduces the sparsity of the loss gradients and avoids quantisation errors (e.g. found in occupancy grids [24]) and can achieve new state-of-the-art accuracy in unsupervised registration when embedded in a more explicit regularisation strategy again realised with DiVRoC. The new measure also excels at rapid instance optimisation that can for one be used in self-supervised learning (as pseudo ground truth) and helps to achieve highly accurate alignment of detailed point clouds while remaining competitive in inference run times.

1.1. Contribution

Our newly presented differentiable volumetric rasterisation (DiVRoC) technique is the core strength of our method that constitutes the following key contributions:

1. **Unsupervised learning of highly deformable point cloud registration.** For the first time we demonstrate that geometric registration of vascular trees can be trained in an unsupervised fashion and more effectively than with stronger supervision through synthetically generated pairs.

2. A novel and **highly scalable differentiable volumetric rasterisation technique** for point clouds. This module enables a robust, accurate and fast computation of point cloud distances or losses with respective gradients for the corresponding displacement vectors. It can also be employed to densify and implicitly regularise sparse point cloud motion and helps to substantially stabilise training.

3. **Sub-second large-motion instance optimisation** on point cloud pairs with over 100k 3D points each that can correct residual errors of network predictions and provide pseudo ground truth for self-training of larger models.

4. **New state-of-the-art performance on the challenging PVT dataset.** Ablation studies that demonstrate the benefits of DiVRoC for loss function, implicit regularisation and instance optimisation over previous research.

5. A **new dataset for out-of-domain evaluation** with a total of 1500 manual landmark pair annotations for 30 scans and further large-scale point cloud data of cancer screening scans is created and will be made public.

2. Related work

Deep learning on point clouds. Geometric deep learning [4] is ubiquitously used for effective visual 3D data analysis. The major challenge consists in processing sparse and unordered 3D points without grid structure, prohibiting the use of ordinary convolutions. The pioneering PointNet [35] addressed the problem by extracting point-wise spatial embeddings with a shared MLP, which are aggregated with a permutation-invariant global max-pooling operation. To include local geometric information, PointNet++ [36] proposed hierarchical grouping, and various subsequent works introduced generic convolution operations applicable to irregular domains [19, 45, 48, 57, 51, 22, 44, 18, 52, 16]. Among these, we employ the Dynamic Graph CNN [45]

as backbone in our work, which formulates graph convolutional networks to learn a feature extractor for suitable neighbourhood representations. Specifically, the authors employ a kNN-neighbourhood and define EdgeConvolutions that concatenate feature information from two nodes that are connected by an edge and symmetrically aggregate the information from nonlinear (local) MLPs using a max-pooling over all k neighbours. Basis point sets [33] use prototype point sets to encode each point of an irregular cloud by its vector to the closest keypoint and thus embed the information into a regular structure that can be used as input to a normal MLP.

Medical image analysis: On the contrary, medical image analysis is predominantly addressed using 3D convolutional neural networks [20, 37, 6] as the data is already acquired in a dense voxelised grid. Out of 65 method submissions to the comprehensive 3D medical image registration challenge Learn2Reg [14] only a single approach was based on sparse graph-based analysis. This also highlights an apparent gap in research for applying geometric learning to 3D data in the medical domain.

Point-voxel representations: Notable recent works have proposed the benefits of combining sparse and dense 3D point representations, e.g. by voxelising multi-view depth images into a relatively coarse volume for semantic scene understanding or classification [24, 50]. The Point-voxel CNN (PVCNN) [23] even interleaves sparse and voxel-based point operators for an optimal trade-off between accuracy and computational efficiency (mainly to avoid unpredictable, irregular memory access that e.g. slow down DGCNNs).

Point cloud distances: A number of recent learning-based works have aimed to overcome the shortcomings of the Chamfer distance and Earth Mover's distance (EMD). In Occupancy Networks [25] and DeepSDF [30], the boundary between the inner and outer parts of a model is represented by an indicator function or signed distance function respectively, whereas Deep Point Distances (DPDist) [43] learn an implicit neural functions and a 3D modified Fisher vector representation to estimate the surface locally with reduced complexity. While providing improvements in robustness with respect to different samplings of point clouds, these methods, nevertheless, all rely on a separately learned model of the shapes and cannot be used as plug-in replacement for Chamfer and EMD and might not generalise well to highly deformable objects such as pulmonary vessels.

Computer graphics: Differentiable rendering [29] or rasterisation are frequently used in computer graphics for shape reconstruction, but has so far been mainly focused on surfaces (either represented as vertices or splatting points) [53] and not yet used in the context of deformable 3D registration.

Learning deformable point cloud registration: In this work we focus on self-supervised learning approaches, since manual or synthetic ground truth is extremely difficult to obtain for high-resolution pulmonary vessel trees in large numbers. Following seminal classic optimisation based point cloud registration approaches such as coherent point drift [27], subsequent deep learning based work aimed at extending the successful Gaussian distribution modelling of two point clouds in DeepGMR [54] into a supervised setting. Finding soft correspondences for large point clouds by optimising a transport energy (Earth Mover's or Wasserstein distance [28]) often by means of the sinkhorn algorithm [40, 8, 32] (and its variants) lies at the core of (deep learning based) optimal transport registration [38, 34]. The key objective of FLOT [34] is to learn geometric features from two point clouds that enable a suitable similarity metric for the cost of the optimal transport solver given ground truth correspondences. RobOT [38] extends this concept to synthetic supervision and implicit regularisation using LDDMM or B-splines. FlowNet3D [21] and PointPWC-Net [49] on the other hand, borrow ideas from their 2D optical flow counterparts and learn to (locally) embed the geometric relations of two point clouds to infer their flow using graph convolutions. PV-RAFT [47] leverages the successful iterative strategy of classical methods with trainable step to update the sparse flow. Test-time optimisation (or instance optimisation) was also considered in [12] that iteratively minimises the point distances for pointwise rigid scene motion.

Importantly, we do not claim a new standalone registration architecture but a versatile method that can enhance several recent algorithms. The DiVRoC loss function could be used for rigid motion estimation or shape reconstruction, whereas the implicit regularisation that is based on a differentiable densifying of sparse displacements is useful to stabilise training in many other 3D registration scenarios.

3. Methods

We address the problem of unsupervised sparse point cloud registration. Our input are two point clouds, source $\boldsymbol{S} \in \mathbb{R}^{N_s \times 3}$ and target $\boldsymbol{T} \in \mathbb{R}^{N_t \times 3}$, comprising arbitrary numbers N_s and N_t of 3D coordinates and no additional features. Our goal is to predict a sparse displacement field $\boldsymbol{\psi} \in \mathbb{R}^{N_s \times 3}$ that geometrically aligns the source towards the target point cloud. We assume that one-to-one correspondences do not necessarily exist and therefore pose the registration as regression task. N_s and N_t can be of very high cardinality i.e. point cloud sizes of over 100 thousands are common, which poses computational problems for many graph-based algorithms. To enable learning with the commonly employed registration networks, we also compute coarser but similarly distinctive clouds with equally reduced number of points N_{sub} through local den-

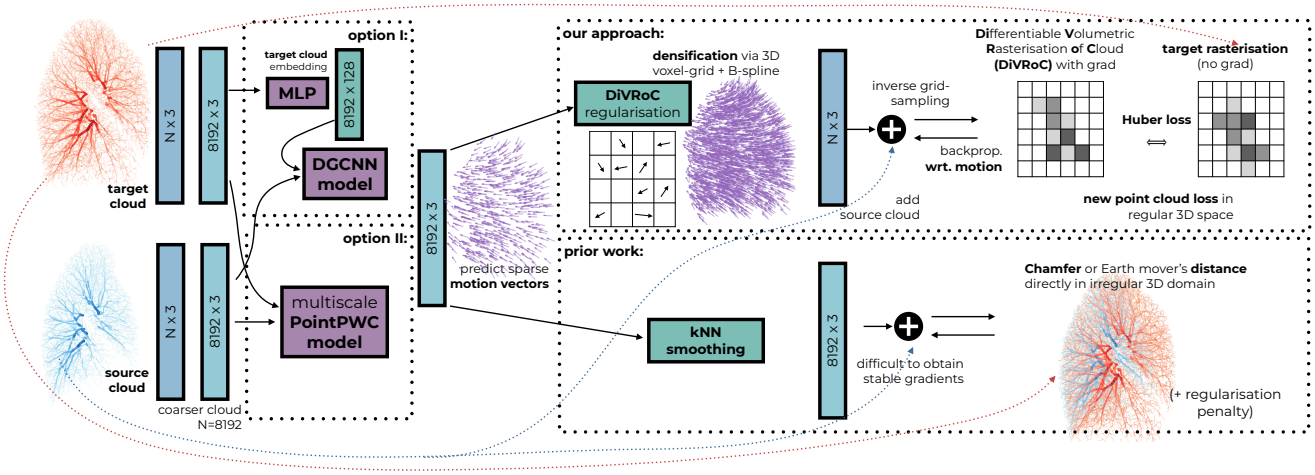

Figure 2. Overview of our proposed concept for learning-based unsupervised large-scale point cloud registration. Given two point clouds a geometric registration model predicts a sparse motion field, which is regularised and densified using differentiable volumetric rasterisation of point clouds (DiVRoc). The new coordinates of the transformed source cloud are then rasterised in a 3D volume and a Huber loss against the fixed target cloud rasterisation is used to train the network (again making use of the stable differentiation of DiVRoc). During inference or for self-supervised learning a subsequent instance optimisation using Adam and the two novel DiVRoC modules is employed. The alternative solution of prior work that employs k-NN smoothing and a Chamfer loss in the irregular domain is shown as comparison.

sity estimation and non-maximum suppression (NMS). The reduced point clouds (and displacements) are denoted with a superscript *. We retain the high-resolution clouds for loss computation and fine-tuning with instance optimisation. Nevertheless our method could also be employed directly to larger point clouds.

Our approach comprises three main parts: 1) a geometric neural network that receives the two point clouds as input and predicts an estimated field $\boldsymbol{\psi}$, 2) an implicit regularisation that smoothes the raw network predictions and enables the interpolation from the coarse to the high-resolution cloud, achieved by an intermediate densification onto a regular grid and 3) our novel unsupervised DiVRoC loss function that measures the similarity of the transformed source and target clouds. To obtain optimal results these steps are prepended to a final fine-tuning using instance (test-time) optimisation on the full-resolution clouds without employing the network but only steps 2) and 3). In the following, the DiVRoC module and each individual step of the registration framework is explained in more detail.

3.1. Differentiable volumetric rasterisation of point clouds (DiVRoC)

Bi- or trilinear interpolation is ubiquitously used in image processing and defines the process of sampling a value in between regular grid points, which e.g. is necessary when warping an image with a displacement field of sub-pixel accuracy. Differentiable image sampling was introduced in the concept of deep learning for spatial transformer networks [15] and has since been used by most learning-based registration approaches (e.g. [26]). Following the notation of [9] trilinear interpolation can be formulated as

$$\boldsymbol{y}(\boldsymbol{p}) = \sum_{\boldsymbol{q}} G(\boldsymbol{q}, \boldsymbol{p}) \cdot \boldsymbol{x}(\boldsymbol{q}), \tag{1}$$

where $\boldsymbol{p} = \boldsymbol{p_0} + \boldsymbol{\Delta p}$ represent sampling coordinates $\boldsymbol{p_0} \in \mathbb{N}^3$ displaced with a sub-voxel offset $\boldsymbol{\Delta p} \in \mathbb{R}^3$, $\boldsymbol{q} \in \mathbb{N}^3$ is iterated over all integer coordinates (the whole grid) of the image $\boldsymbol{x} \in \mathbb{R}^{n_x \times n_y \times n_z}$ and G defines the interpolation kernel. For trilinear interpolation the kernel is three-dimensional and can be decomposed into:

$$G(\boldsymbol{q}, \boldsymbol{p}) = g(q_x, p_x) \cdot g(q_y, p_y) \cdot g(q_z, p_z), \tag{2}$$

where g is defined as $g(q_x, p_x) = \max(0, 1 - |q_x - p_x|)$. This is usually only non-zero for exactly eight grid points. In pytorch [31] this operation is called `grid_sample` and efficiently implemented with respective Jacobians with respect to the gradients of $\boldsymbol{\Delta p}$. `grid_sample` enables us to return a tensor of interpolated values given a gridded image/volume and a tensor of sampling coordinates.

Now, let us think about the opposite situation where we have a set of interpolated values and corresponding sampling coordinates and want to **rasterise** an image/volume on a grid. That means the information of each point with non-integer (off-grid) coordinates is now distributed across all eight immediate (integer) neighbours according to the interpolation weight from before. We can thus simply reverse the order of $\boldsymbol{p}$ and $\boldsymbol{q}$ in Eq. 2:

$$\boldsymbol{x}(\boldsymbol{q}) = \sum_{\boldsymbol{p}} G(\boldsymbol{p}, \boldsymbol{q}) \cdot \boldsymbol{y}(\boldsymbol{p}), \tag{3}$$

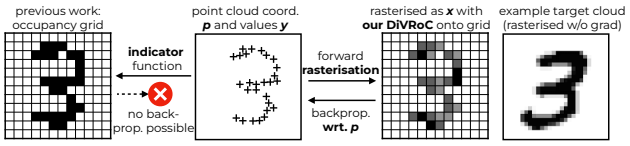

Figure 3. Detailed concept of proposed differentiable volumetric rasterisation of point cloud that serves as core layer in our method. In contrast to occupancy grids DiVRoC provides a well-defined gradient for the relative displacement of point cloud coordinates.

and iterate over all elements of our value vector $\boldsymbol{y}$ and again the summation is very sparse. The total sum (mass) of values in $\boldsymbol{x}$ and $\boldsymbol{y}$ will be equal.

Implementation of forward and derivative operators: Since efficient (GPU) implementations should avoid custom loops and directly opt for the best available computational backend (e.g. CuDNN) we use the following mathematical reformulation to employ the built-in `jacobian` from pytorch's `autograd` tools. Differentiating $-\sum \frac{1}{2}(\texttt{grid\_sample}(\boldsymbol{x},\boldsymbol{p})-\boldsymbol{y})^2$ with respect to $\boldsymbol{x}$ yields the desired result. This implementation enables us to rapidly rasterise 3D volumes from very large point clouds. In practice the value vector $\boldsymbol{y}$ is composed of only ones in this case. Furthermore, the information can be smoothed using a Gaussian kernel (that can efficiently be applied for each dimension separately). Note that while the idea is in principle similar to occupancy grids that are e.g. used in VoxNet [24], it is crucially different. Occupancy grids use nearest neighbour interpolation, which leads to less accurate rasterisations (especially for smaller 3D grids) and they are not differentiable with respect to the point positions. Yet, this is necessary to use the rasterised volume as loss for a deformable point cloud registration network.

To obtain the custom backward path, i.e. for $\boldsymbol{z} = \frac{\partial \mathcal{L}}{\partial \boldsymbol{x}}$ through our DiVRoC layer we need two computations: the gradient with respect to the $\boldsymbol{y}$ is simply the forward mode $\texttt{grid\_sample}(\boldsymbol{z},\boldsymbol{p})$. Gradients with respect to $\boldsymbol{p}$ are obtained using the reverse (ie. the Jacobian) of $\texttt{grid\_sample}(\boldsymbol{z},\boldsymbol{p})$ multiplied by the negated input $-\boldsymbol{y}\cdot N$. Here N is the number of elements in $\boldsymbol{y}$. When dealing with multichannel point clouds and mini-batches some specific care needs to be taken to use correctly shaped tensors as detailed in our open source code.

3.2. Geometric registration networks

Our approach is in general agnostic to the geometric registration architecture. We embed our method into two graph-based learning frameworks for point registration.

First, we select PointPWC-Net [49], which was successfully employed within the RobOT approach that set state-of-the-art performance for the PVT dataset. Through its key components, a multi-scale architecture with intermediate point cloud warping steps, and a cost volume layer adapted to sparse displacement candidates it enables good coverage for large deformations. It comprises ≈100 layers and 8 million trainable parameters. So far it relied on stronger supervision for deformable point registration and struggled with weak unsupervised losses[1]

Second, we also propose an adaptation of the general purpose DGCNN (dynamic graph convolutional network). DGCNNs are widely used for global classification or point-wise semantic segmentation for a single point cloud and features that reside on the same coordinates [45]. For point registration, a method to encode the target point cloud is required. Inspired by the basis point set [33] and [1], we propose to capture this information by assigning the geometric 3D vector that maps a point from the source to its closest spatial neighbour in the target cloud. For each of these 3D vectors a 128D embedding is learned using local MLPs with ShuffleNet [55] 1×1 convolutions. Together this model comprises 1.7 million trainable parameters and is also used with input cloud sizes of 8192.

3.3. Implicit regularisation and densification

As pointed out before, unsupervised learning of highly deformable point cloud registration may be prone to difficulties in defining a stable gradient that avoids the inherent local optima. E.g. two small local sub-parts of the pulmonary vascular tree can look very similar leading to a high risk of mixed-up correspondences. We argue that a strong regularisation is necessary to empower the model to overcome such problematic locations by filling in more reliable information from neighbouring parts of the point cloud.

One strategy that was proposed for the PointPWC-Net [49] is to use k-nearest neighbour smoothing (implicit regularisation) and an additional explicit regularisation penalty to avoid differing motion vectors for spatially close points. We found empirically that this strategy is insufficient to tackle the specific challenges of the PVT dataset.

Consequently, we propose a stronger implicit regularisation by mapping the sparse displacement field into a dense grid (*densification*) using DiVRoC and applying a spatial B-spline smoothing in this voxelised 3D space. Our regularisation implements a cardinal quadratic B-spline smoothing by two iterative box filter steps of kernel size $5 \times 5 \times 5$ with padding. The differentiable rasterisation function can be used as before with the extension to a four-channel input (3D coordinates plus unit vector). Once those homogenous coordinates of the displacements ψ^* (estimated by our network) are extrapolated to a regular voxel grid (densified), we can also resample the resolution to the original point cloud size. Hence, we obtain ψ and use our point distance in the highest resolution.

[1] see Sec. "Unsupervised loss functions" in the Supplement of https://dblp.org/rec/conf/nips/ShenFLCEEN21 for details.

3.4. Novel point distance based on DiVRoC

The most important innovation that can be realised with DiVRoC is a density based point distance that provides a very reliable gradient for deformable 3D point registration and is scalable to high-resolution clouds. Recall, the Chamfer distance that is e.g. minimised in iterative closest point (ICP) [2] requires computing a nearest neighbour for every source point in the target cloud and/or vice-versa. Because the nearest neighbour (`argmin`) operation is not differentiable itself the gradient will be obtained from its directional offset. This has the disadvantage that the spatial proximity to a higher density of points that are subsequent (second, third, etc.) neighbours would be ignored.

Our approach works as follows: First, we rasterise the high-resolution target point cloud $\boldsymbol{T}$ with its points $\boldsymbol{p_T}$ using DiVRoC (cf. Eq. 3). This yields a voxelised representation $\boldsymbol{x}_T$ and does not track gradients. Second, we obtain the network prediction ψ (densified and regularised, see above) that represents the relative displacements $\boldsymbol{\Delta p}$ for source points that align $\boldsymbol{S}$ towards $\boldsymbol{T}$. The forward path rasterises the spatially warped cloud $\boldsymbol{S}_\psi = \boldsymbol{S} + \psi$ as $\boldsymbol{x}_{S_\psi}$ into a comparable voxel grid. Here, we track gradients and require differentiable operators. We use the Huber norm $\mathcal{L}_H(\boldsymbol{a}, \boldsymbol{b}) = \sum_i \ell_H(a_i, b_i)$ (or SmoothL1Loss) with

$$\ell_H(a_i, b_i) = \begin{cases} \frac{1}{2\beta}(a_i - b_i)^2, & \text{if } |a_i - b_i| < \beta \\ |a_i - b_i| - \frac{1}{2\beta}, & \text{otherwise} \end{cases}$$

as distance loss. As detailed in Sec. 3.1 it is possible to obtain a derivative of $\mathcal{L}_H(\boldsymbol{S}_\psi, \boldsymbol{T})$ with respect to ψ using our DiVRoC implementation.

3.5. Adam Instance Optimisation and Self-Training

The two previous steps (implicit regularisation and DiVRoC point distance) together can already be used for optimisation-based point registration. When defining the intermediate B-spline grid that reparameterises ψ as a trainable parameter, we can directly employ the well-known Adam optimiser [17] to obtain competitive registration results. Here, the displacement vectors of ψ are sampled at the points $\boldsymbol{p_S}$ and subsequently optimised directly based on $\mathcal{L}_H$. This coarse regular grid is initialised with zeros. Due to the intrinsic regularity of the B-spline transformation a reasonable optimisation landscape is obtained. Nevertheless, using a single-scale is not sufficient to model the highly complex lung motion and a robust coarse scale alignment, either using the presented networks or another optimisation based initialisation is necessary.

Self-Training: Due to the fact that the instance optimisation can directly optimise the alignment of high-resolution point clouds it can also provide a reasonable pseudo ground truth for a subsequent network refinement. We propose a self-training setting that enables us to further improve the PPWC network. We use pre-computed predictions from an earlier model (DGCNN) with Adam fine-tuning on the training set as pseudo ground truth for supervision (see ablation study 6 in Sec. 4.2). The loss now becomes directly the registration error. Note, that more sophisticated techniques such as the mean teacher paradigm could also be used [3].

4. Experiments and results

We apply our proposed method and variants thereof on the pulmonary vascular tree (PVT) dataset published under CC BY-NC-SA 3.0 licence in 2021 by [38] [2]. The PVT dataset comprises 1000 pairs of high-resolution point clouds that were obtained from processing inspiration and exhale CT scans of patient with chronic obstructive pulmonary disease (COPD) as part of the IRB-approved COPDgene study (NCT00608764). The motion is particularly challenging because subjects were asked to perform forced inspiration and exhale with volume changes of a factor of 2 or more. These point clouds depict the vascular tree of both lungs in high detail, making them ideal to study local deformations that correlate with lung ventilation for clinical assessment of COPD severity. But they comprise no manual ground truth making it a hard deep learning problem. For evaluation another ten vascular tree of the DIRLAB-COPDgene [5] are provided for which anatomical correspondences (300 per scan) of expert raters are available as ground truth.

We only train our models on the unsupervised 1000 pairs and evaluate on the 10 DIRLAB cases using the target registration error (TRE) in millimetres. Our validation comprises a detailed ablation study that evaluates each of our contributions and the important design choices that make unsupervised learning possible on this complex data. All our experiments contain both the network prediction alone and the refined field with Adam instance optimisation.

4.1. Implementation details and run times

The point cloud registration networks were used as provided by public implementation, with the following modifications. The normalisation in PointPWC was replaced from BatchNorm to GroupNorm to solve inconsistencies across training and inference of the original implementation. The embedding of the target point cloud $\boldsymbol{T}$ in the coordinate space of $\boldsymbol{S}$ as 128D input features for the DGCNN was implemented as 6 Conv + GroupNorm + ReLU blocks - kernel size=1, group-size=4 and shuffle permutation [55] - and a channel size of 512 except the first and last layer. Each network received clouds of size 8192 as input. The implicit regularisation employs DiVRoC with a fixed sized grid of size $19 \times 19 \times 19$ and a Gaussian kernel with $\sigma = 3.5$,

[2] https://github.com/uncbiag/shapmagn

Table 1. Mean target registration error and std-dev of variants of the proposed DiVRoC within our ablation experiments on the PVT1010 dataset, evaluated each with and without Adam fine-tuning (instance optimisation). TRE before registration was 23.62mm on average.

DiVRoC regulariser	DiVRoC distance	DGCNN	self-train PPWC	without Adam	with Adam
kNN-smooth	✓	✓	✗	$22.2_{\pm 13.0}$	$11.4_{\pm 10}$
✓	Chamfer	✓	✗	$7.56_{\pm 5.8}$	$4.49_{\pm 5.6}$
✓	rasterise+STN	✓	✗	$9.18_{\pm 5.8}$	$3.15_{\pm 4.9}$
✓	✓	PPWC	✗	$8.96_{\pm 4.8}$	$3.21_{\pm 4.9}$
✓	✓	✓	✗	$7.35_{\pm 4.6}$	$2.71_{\pm 4.3}$
✓	✓	✓	✓	$\mathbf{5.96}_{\pm 4.0}$	$\mathbf{2.39}_{\pm 3.4}$

followed by a quadratic B-spline smoothing kernel with a size of 5. We experimented with higher resolution grids but found it could destabilise training and result in no/little performance gains. The DiVRoC metric is implemented with a Gaussian kernel with $\sigma = 0.7$ and a grid size of $76\times76\times76$ for the first stage (network training and prediction). The grid size is doubled for the fine-tuning stage with Adam. All networks were trained with Adam using an initial learning rate of $\eta = 0.001$ and step learning rate scheduling with $\gamma = 0.5$ after every 8 out of 64 epochs with a mini-batch size of 6. We employ a (sigmoidal) ramp-up phase of 150 iterations in which a Euclidean norm is added to the predicted displacements and hence a zero-displacement is encouraged to avoid divergence at the early stage of training. Each training was completed in approximately 2.5 hours on a RTX A40 and our models require only 2-3 GByte of VRAM. All feed-forward network run in less than 100 milliseconds (PointPWC 90ms, DGCNN 50ms).

The Adam instance optimisation uses 50 iterations with a learning rate of 0.01 and uses a displacement grid of size $38\times38\times38$ and employs the highest resolution clouds. It requires about 250ms, whereas a Chamfer based instance optimisation would require 1.7s, we thus achieve a more than 6-fold speed-up. Source code and a more detailed analysis of the computational complexity for different point cloud sizes, rasterisation grid spacings and metrics is given in the supplementary material.

4.2. Ablation studies

Table 1 quantitatively compares alternatives that represent the current state-of-the-art solutions for unsupervised point-cloud registration and can be seen as variants of our method. Fig. 2 shows our concepts in contrast to those alternate design and loss choices.

1) We use 3-fold **kNN-smoothing** as alternative to our proposed DiVRoC regularisation - that performs B-spline smoothing on a coarse regular grid and densifies the correspondence field. We use k=150 and an MSE-Loss between this smoothed network output and a B-spline regularised version of it. Its accuracy is very poor (TRE=22.2/11.4mm) and shows the importance of DiVRoC's differentiable regularisation that stabilises the unsupervised learning.

2) Next, we employ the **Chamfer distance** with a loss weighting of 0.01 (hyper-parameter empirically chosen on a single test scan pair) and applied it to the reduced point cloud sizes $N_{sub} = 8192$ (denoted with superscript * in Sec. 3), because otherwise the gradient would become less stable and the optimisation very memory and resource intensive. It yields worse performance in particular for instance optimisation (Adam) with TRE=7.56/4.49mm. We also experimented with different versions of the Sinkhorn loss [8] and Earth Mover's distance [28] but could not find any successful setting. This mirrors the analysis of [38] that found the lung geometry to complex for those losses. In addition, we tried using the local curvature as proposed in [49] but again found the loss to be unsuitable for the PVT dataset.

3) Furthermore, we evaluate the benefit of our proposed direct differentiability of our point cloud loss with respect to motion vectors in contrast to using a spatial transformer loss [15], denoted as *rasterise+STN* (see also Fig. 3), which yields 15-20% higher TRE than our method. For this variant both initial clouds are rasterised statically (without gradient), but the target cloud volume is now spatially warped for improved alignment and thus the derivative can be obtained from a forward `grid_sample` operation as done e.g. in [26] for 3D images.

4) To evaluate the influence of architecture choice, we compare our registration adaptation of the DGCNN with the widely used PointPWC-Net (abbreviated as PPWC). We decided to keep the additional implicit DiVRoC regulariser after the output of PPWC to enforce more smoothness and found it was necessary to avoid diverging training runs. Also, all other hyperparameters for the training and inference procedure are kept the same. Our baseline model is listed as row 5) in Table 1 and performs 15-20% better.

6) When using self-training to enable a further fine-tuning of the network, the DiVRoC distance is no longer used and replaced by a Euclidean error of predictions versus pseudo ground truth. The pseudo ground truth was generated using our baseline unsupervised DGCNN model with Adam instance optimisation. To make the most out of this now stronger supervision, we train the PointPWC that has more capacity and add augmentation to the target point clouds for which we draw a new random subset of size 8192 from the high-resolution cloud in each iteration. To deal with a potential shift in the geometry of the randomly sub-sampled clouds compared to the local density and NMS strategy used before, we run the PointPWC $8\times$ at inference

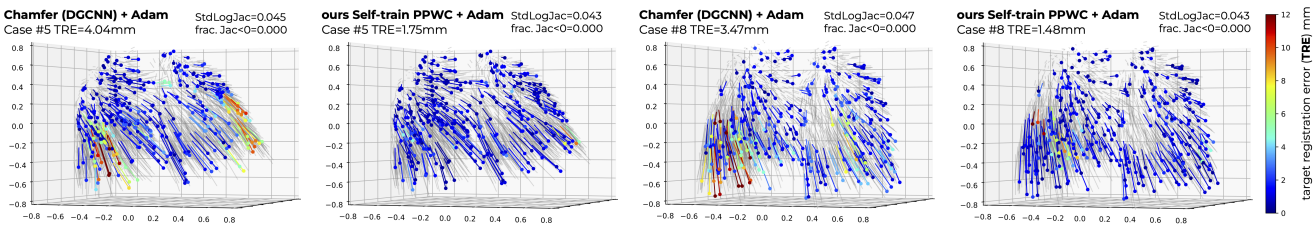

Figure 4. Visual comparison of results from our representative implementation of prior work (Chamfer loss) compared to our DiVRoC loss with self-training (more detailed renderings and more cases in supplementary material) demonstrates the robustness of our approach to handle even cases of challenging motion. Gray lines show predicted displacements on the points of the moving cloud, and coloured dots and lines show the moving landmarks and the flow interpolated to them, with the colour encoding the error magnitude (TRE). Blue arrows indicate high agreement with manual ground truth.

with new target cloud samples (test time augmentation) and average the results. With this we obtain substantial further improvements: 3.21→2.39 mm and overall best results.

4.3. Comparison to state-of-the-art

Visual results with high-resolution point cloud renderings before and after alignment are shown in Fig. 5 and our supplementary material. We quantitatively compare our results to the best published methods from the literature in Table 2 based on manual landmarks and demonstrate clear advantages over the recent D-RobOT (LDDMM) [38] method outperforming it by as much as 16.4% (2.39 vs 2.86 mm) without the need for training with synthetic deformations. In addition, we also restrict ourselves to the raw point clouds without further geometric features ([38] use also the vessel radius) and are substantially faster. To highlight the challenges of the dataset one can see that CPD (non-rigid), still a competitive method for many other benchmarks, reduces the TRE only by 60% versus initial whereas we reach an 88% error reduction. To explore the capability of our DiVRoC method to be used directly for optimisation of 3D registration without any trainable network, we add another variant that uses Adam instance optimisation twice. First, as initialisation for which we use a coarser scale and halve both displacement and rasterisation grid. And then second as done for the network variants. We can show that this method itself almost reaches state-of-the-art performance owing to the robust novel point distance that it optimises.

4.4. Longitudinal Cancer Screening Dataset

To evaluate the performance of our approach for out-of-domain (OOD) examples (PVT1010 comprises only point cloud pairs with identical scanning and breathhold protocol), we have created a new point cloud dataset for longitudinal cancer screening by searching for public low-dose CT acquisitions in TCIA of cancer patients with follow-up scans after one year [7, 41, 42]. We automatically extract lung vessel segmentations using the TotalSegmentator [46] and convert those into sparse clouds with approximately

Table 2. Mean target registration error, percentiles, and inference times of our method for point cloud registration on the PVT1010 dataset compared to the state of the art. Note that all our methods use Adam instance optimisation with DiVRoC.

Method	TRE [mm]	25% [mm]	75% [mm]	Time [s]
initial	23.30	13.18	31.65	-
ICP (affine) [2]	15.05	9.60	20.01	0.52
CPD (non-rigid) [27]	9.30	5.95	11.83	332.60
RobOT (raw) [38]	9.41	4.89	13.04	**0.15**
S-RobOT (LDDMM) [38]	5.48	2.86	7.14	42.30
D-RobOT (LDDMM) [38]	2.86	1.25	3.11	1.92
2x Adam (not learned)	2.93	**0.97**	2.24	0.53
ours DGCNN	2.71	0.99	2.31	0.30
ours PPWC (self-train)	**2.39**	**0.97**	**2.19**	1.02

90k points each. Note, that these steps are different to [38] leading to a substantial domain gap. To avoid simplistic point cloud pairs we measure the lung volume overlap after rigid pre-alignment and select only cases with large motion. A medical expert annotated 15 CT scans (baseline + follow-up) with 100 manual landmarks pairs each, that are only used for test evaluation. More details are in the supplementary material and a larger dataset with 500 point cloud pairs will be made publicly available to the community.

Our results demonstrate how well our method generalises without any further training or hyperparameter optimisation on the OOD-data. Given an TRE of 10.16±2.29 mm after rigid alignment, we achieve 5.91±1.83 mm (PPWC self-train) and 2.45±0.49 mm (PPWC self-train + Adam) respectively, which is substantially better than the Chamfer distance based baseline with 9.03±2.84 mm and 5.11±1.09 mm (w/o or with Adam).

Limitations: Our unsupervised 3D registration model relies more on instance optimisation than other methods that were trained with stronger supervision. This is owed to the fact that the complexity of the deformation can make the

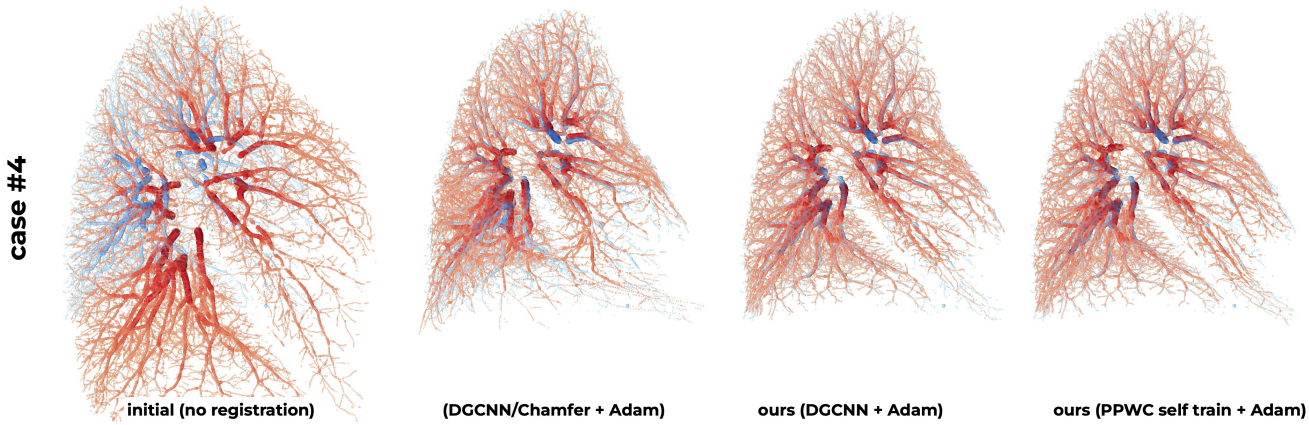

Figure 5. High-resolution point cloud renderings before and after alignment one exemplary Case #4 demonstrating the large extent of motion that has to be compensated. The advantages of our approach compared to using the Chamfer distance for training the DGCNN and instance optimisation are visible in the lower part of the lungs. For this case our best setting (PPWC with self-training) and the simpler DGCNN with DiVRoC perform on par.

training of an unconstrained highly parameterised network unstable for the challenging lung registration task as described in [38]. Furthermore, our networks use a restricted point cloud size of 8192, whereas the instance optimisation efficiently processes clouds with over 100k points. In future work, a coarse-to-fine approach during training can further reduce the reliance on instance optimisation. In addition, our experiments do not include (multi-)rigid motion benchmarks since this is considered to be an easier task and out of the direct scope of the presented solution for deformable motion estimation. However, we created a new lung point cloud dataset as additional out-of-domain validation and demonstrated the generalisation of our models.

5. Conclusion

We have presented a novel **Di**fferentiable **V**olumetric **R**asterisation **o**f point **C**louds (DiVRoC) technique that provides an accurate loss and regularisation for large-scale point cloud alignment and overcomes the challenges for unsupervised learning for highly deformable structures. Embedded into PointPWC or DGCNN, DiVRoC yields new state-of-the-art accuracy on the challenging PVT dataset without training with manual ground truth for 1) unsupervised metric-based learning 2) self-supervised learning with pseudo labels and 3) optimisation based alignment.

References

[1] Zachary MC Baum, Yipeng Hu, and Dean C Barratt. Real-time multimodal image registration with partial intraoperative point-set data. *Medical image analysis*, 74:102231, 2021. 5

[2] PJ Besl and Neil D McKay. A method for registration of 3-d shapes. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 14(02):239–256, 1992. 6, 8

[3] Alexander Bigalke, Lasse Hansen, and Mattias P Heinrich. Adapting the mean teacher for keypoint-based lung registration under geometric domain shifts. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2022: 25th International Conference, Singapore, September 18–22, 2022, Proceedings, Part VI*, pages 280–290. Springer, 2022. 6

[4] Michael M Bronstein, Joan Bruna, Yann LeCun, Arthur Szlam, and Pierre Vandergheynst. Geometric deep learning: going beyond euclidean data. *IEEE Signal Processing Magazine*, 34(4):18–42, 2017. 1, 2

[5] Richard Castillo, Edward Castillo, David Fuentes, Moiz Ahmad, Abbie M Wood, Michelle S Ludwig, and Thomas Guerrero. A reference dataset for deformable image registration spatial accuracy evaluation using the copdgene study archive. *Physics in Medicine & Biology*, 58(9):2861, 2013. 6

[6] Heang-Ping Chan, Ravi K Samala, Lubomir M Hadjiiski, and Chuan Zhou. Deep learning in medical image analysis. *Deep Learning in Medical Image Analysis: Challenges and Applications*, pages 3–21, 2020. 3

[7] Kenneth Clark, Bruce Vendt, Kirk Smith, John Freymann, Justin Kirby, Paul Koppel, Stephen Moore, Stanley Phillips, David Maffitt, Michael Pringle, Lawrence Tarbox, and Fred Prior. The Cancer Imaging Archive (TCIA): Maintaining and Operating a Public Information Repository. *J. Digit. Imaging*, 26(6):1045–1057, Dec. 2013. 8

[8] Marco Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. *Advances in neural information processing systems*, 26, 2013. 3, 7

[9] Jifeng Dai, Haozhi Qi, Yuwen Xiong, Yi Li, Guodong Zhang, Han Hu, and Yichen Wei. Deformable convolutional networks. In *Proceedings of the IEEE international conference on computer vision*, pages 764–773, 2017. 4

[10] Jean Feydy. *Geometric data analysis, beyond convolutions*. PhD thesis, Université Paris-Saclay Gif-sur-Yvette, France, 2020. 2

[11] Jean Feydy, Benjamin Charlier, François-Xavier Vialard, and Gabriel Peyré. Optimal transport for diffeomorphic registration. In *Medical Image Computing and Computer Assisted Intervention- MICCAI 2017: 20th International Conference, Quebec City, QC, Canada, September 11-13, 2017, Proceedings, Part I 20*, pages 291–299. Springer, 2017. 2

[12] Zan Gojcic, Or Litany, Andreas Wieser, Leonidas J Guibas, and Tolga Birdal. Weakly supervised learning of rigid 3d scene flow. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5692–5703, 2021. 3

[13] Xiuye Gu, Yijie Wang, Chongruo Wu, Yong Jae Lee, and Panqu Wang. Hplflownet: Hierarchical permutohedral lattice flownet for scene flow estimation on large-scale point clouds. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 3254–3263, 2019. 1

[14] Alessa Hering, Lasse Hansen, Tony CW Mok, Albert CS Chung, Hanna Siebert, Stephanie Häger, Annkristin Lange, Sven Kuckertz, Stefan Heldmann, Wei Shao, et al. Learn2reg: comprehensive multi-task medical image registration challenge, dataset and evaluation in the era of deep learning. *IEEE Transactions on Medical Imaging*, 2022. 2, 3

[15] Max Jaderberg, Karen Simonyan, Andrew Zisserman, et al. Spatial transformer networks. *Advances in neural information processing systems*, 28, 2015. 4, 7

[16] Li Jiang, Hengshuang Zhao, Shu Liu, Xiaoyong Shen, Chi-Wing Fu, and Jiaya Jia. Hierarchical point-edge interaction network for point cloud semantic segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 10433–10441, 2019. 2

[17] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In Yoshua Bengio and Yann LeCun, editors, *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, 2015. 6

[18] Jiaxin Li, Ben M Chen, and Gim Hee Lee. So-net: Self-organizing network for point cloud analysis. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 9397–9406, 2018. 2

[19] Yangyan Li, Rui Bu, Mingchao Sun, Wei Wu, Xinhan Di, and Baoquan Chen. Pointcnn: Convolution on x-transformed points. *Advances in neural information processing systems*, 31, 2018. 1, 2

[20] Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen Awm Van Der Laak, Bram Van Ginneken, and Clara I Sánchez. A survey on deep learning in medical image analysis. *Medical image analysis*, 42:60–88, 2017. 3

[21] Xingyu Liu, Charles R Qi, and Leonidas J Guibas. Flownet3d: Learning scene flow in 3d point clouds. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 529–537, 2019. 1, 3

[22] Yongcheng Liu, Bin Fan, Shiming Xiang, and Chunhong Pan. Relation-shape convolutional neural network for point cloud analysis. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8895–8904, 2019. 2

[23] Zhijian Liu, Haotian Tang, Yujun Lin, and Song Han. Point-voxel cnn for efficient 3d deep learning. *Advances in Neural Information Processing Systems*, 32, 2019. 3

[24] Daniel Maturana and Sebastian Scherer. Voxnet: A 3d convolutional neural network for real-time object recognition. In *2015 IEEE/RSJ international conference on intelligent robots and systems (IROS)*, pages 922–928. IEEE, 2015. 2, 3, 5

[25] Lars Mescheder, Michael Oechsle, Michael Niemeyer, Sebastian Nowozin, and Andreas Geiger. Occupancy networks: Learning 3d reconstruction in function space. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4460–4470, 2019. 3

[26] Tony CW Mok and Albert CS Chung. Large deformation diffeomorphic image registration with laplacian pyramid networks. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part III 23*, pages 211–221. Springer, 2020. 4, 7

[27] Andriy Myronenko and Xubo Song. Point set registration: Coherent point drift. *IEEE transactions on pattern analysis and machine intelligence*, 32(12):2262–2275, 2010. 3, 8

[28] Trung Nguyen, Quang-Hieu Pham, Tam Le, Tung Pham, Nhat Ho, and Binh-Son Hua. Point-set distances for learning representations of 3d point clouds. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 10478–10487, 2021. 2, 3, 7

[29] Michael Niemeyer, Lars Mescheder, Michael Oechsle, and Andreas Geiger. Differentiable volumetric rendering: Learning implicit 3d representations without 3d supervision. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3504–3515, 2020. 3

[30] Jeong Joon Park, Peter Florence, Julian Straub, Richard Newcombe, and Steven Lovegrove. Deepsdf: Learning continuous signed distance functions for shape representation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 165–174, 2019. 3

[31] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32, 2019. 4

[32] Gabriel Peyré, Marco Cuturi, et al. Computational optimal transport. *Center for Research in Economics and Statistics Working Papers*, (2017-86), 2017. 3

[33] Sergey Prokudin, Christoph Lassner, and Javier Romero. Efficient learning on point clouds with basis point sets. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 4332–4341, 2019. 3, 5

[34] Gilles Puy, Alexandre Boulch, and Renaud Marlet. Flot: Scene flow on point clouds guided by optimal transport.

In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXVIII*, pages 527–544. Springer, 2020. 1, 3

[35] Charles R Qi, Hao Su, Kaichun Mo, and Leonidas J Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 652–660, 2017. 2

[36] Charles Ruizhongtai Qi, Li Yi, Hao Su, and Leonidas J Guibas. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. *Advances in neural information processing systems*, 30, 2017. 2

[37] Dinggang Shen, Guorong Wu, and Heung-Il Suk. Deep learning in medical image analysis. *Annual review of biomedical engineering*, 19:221–248, 2017. 3

[38] Zhengyang Shen, Jean Feydy, Peirong Liu, Ariel H Curiale, Ruben San Jose Estepar, Raul San Jose Estepar, and Marc Niethammer. Accurate point cloud registration with robust optimal transport. *Advances in Neural Information Processing Systems*, 34:5373–5389, 2021. 2, 3, 6, 7, 8, 9

[39] Hanna Siebert, Lasse Hansen, and Mattias P Heinrich. Fast 3d registration with accurate optimisation and little learning for learn2reg 2021. In *Biomedical Image Registration, Domain Generalisation and Out-of-Distribution Analysis: MICCAI 2021 Challenges: MIDOG 2021, MOOD 2021, and Learn2Reg 2021, Held in Conjunction with MICCAI 2021, Strasbourg, France, September 27–October 1, 2021, Proceedings*, pages 174–179. Springer, 2022. 2

[40] Richard Sinkhorn and Paul Knopp. Concerning nonnegative matrices and doubly stochastic matrices. *Pacific Journal of Mathematics*, 21(2):343–348, 1967. 3

[41] The National Lung Screening Trial Research Team. Data from the National Lung Screening Trial (NLST). https://doi.org/10.7937/TCIA.HMQ8-J677. 8

[42] The National Lung Screening Trial Research Team, DR Aberle, AM Adams, CD Berg, WC Black, JD Clapp, RM Fagerstrom, IF Gareen, C Gatsonis, PM Marcus, and JD Sicks. Reduced Lung-Cancer Mortality with Low-Dose Computed Tomographic Screening. *N. Engl. J. Med.*, 365(5):395–409, June 2011. 8

[43] Dahlia Urbach, Yizhak Ben-Shabat, and Michael Lindenbaum. Dpdist: Comparing point clouds using deep point cloud distance. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XI 16*, pages 545–560. Springer, 2020. 3

[44] Lei Wang, Yuchun Huang, Yaolin Hou, Shenman Zhang, and Jie Shan. Graph attention convolution for point cloud semantic segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10296–10305, 2019. 2

[45] Yue Wang, Yongbin Sun, Ziwei Liu, Sanjay E Sarma, Michael M Bronstein, and Justin M Solomon. Dynamic graph cnn for learning on point clouds. *Acm Transactions On Graphics (tog)*, 38(5):1–12, 2019. 1, 2, 5

[46] Jakob Wasserthal, Hanns-Christian Breit, Manfred T. Meyer, Maurice Pradella, Daniel Hinck, Alexander W. Sauter, Tobias Heye, Daniel T. Boll, Joshy Cyriac, Shan Yang, Michael Bach, and Martin Segeroth. TotalSegmentator: Robust segmentation of 104 anatomic structures in CT images. *Radiology: Artificial Intelligence*, 5(5):e230024, 2023. 8

[47] Yi Wei, Ziyi Wang, Yongming Rao, Jiwen Lu, and Jie Zhou. Pv-raft: Point-voxel correlation fields for scene flow estimation of point clouds. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 6954–6963, 2021. 1, 3

[48] Wenxuan Wu, Zhongang Qi, and Li Fuxin. Pointconv: Deep convolutional networks on 3d point clouds. In *Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition*, pages 9621–9630, 2019. 1, 2

[49] Wenxuan Wu, Zhi Yuan Wang, Zhuwen Li, Wei Liu, and Li Fuxin. Pointpwc-net: Cost volume on point clouds for (self-) supervised scene flow estimation. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part V 16*, pages 88–107. Springer, 2020. 1, 2, 3, 5, 7

[50] Zhirong Wu, Shuran Song, Aditya Khosla, Fisher Yu, Linguang Zhang, Xiaoou Tang, and Jianxiong Xiao. 3d shapenets: A deep representation for volumetric shapes. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1912–1920, 2015. 3

[51] Mutian Xu, Runyu Ding, Hengshuang Zhao, and Xiaojuan Qi. Paconv: Position adaptive convolution with dynamic kernel assembling on point clouds. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3173–3182, 2021. 2

[52] Yifan Xu, Tianqi Fan, Mingye Xu, Long Zeng, and Yu Qiao. Spidercnn: Deep learning on point sets with parameterized convolutional filters. In *Proceedings of the European conference on computer vision (ECCV)*, pages 87–102, 2018. 2

[53] Wang Yifan, Felice Serena, Shihao Wu, Cengiz Öztireli, and Olga Sorkine-Hornung. Differentiable surface splatting for point-based geometry processing. *ACM Transactions on Graphics (TOG)*, 38(6):1–14, 2019. 3

[54] Wentao Yuan, Benjamin Eckart, Kihwan Kim, Varun Jampani, Dieter Fox, and Jan Kautz. Deepgmr: Learning latent gaussian mixture models for registration. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part V 16*, pages 733–750. Springer, 2020. 3

[55] Xiangyu Zhang, Xinyu Zhou, Mengxiao Lin, and Jian Sun. Shufflenet: An extremely efficient convolutional neural network for mobile devices. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 6848–6856, 2018. 5, 6

[56] Zhiyuan Zhang, Yuchao Dai, and Jiadai Sun. Deep learning based point cloud registration: an overview. *Virtual Reality & Intelligent Hardware*, 2(3):222–246, 2020. 1

[57] Haoran Zhou, Yidan Feng, Mingsheng Fang, Mingqiang Wei, Jing Qin, and Tong Lu. Adaptive graph convolution for point cloud analysis. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 4965–4974, 2021. 1, 2