Deep Learning Based Geometric Registration for Medical Images: How Accurate Can We Get Without Visual Features?

Lasse Hansen(✉) and Mattias P. Heinrich

Institute of Medical Informatics, Universität zu Lübeck, Lübeck, Germany
{hansen,heinrich}@imi.uni-luebeck.de

Abstract. As in other areas of medical image analysis, e.g. semantic segmentation, deep learning is currently driving the development of new approaches for image registration. Multi-scale encoder-decoder network architectures achieve state-of-the-art accuracy on tasks such as intra-patient alignment of abdominal CT or brain MRI registration, especially when additional supervision, such as anatomical labels, is available. The success of these methods relies to a large extent on the outstanding ability of deep CNNs to extract descriptive visual features from the input images. In contrast to conventional methods, the explicit inclusion of geometric information plays only a minor role, if at all. In this work we take a look at an exactly opposite approach by investigating a deep learning framework for registration based solely on geometric features and optimisation. We combine graph convolutions with loopy belief message passing to enable highly accurate 3D point cloud registration. Our experimental validation is conducted on complex key-point graphs of inner lung structures, strongly outperforming dense encoder-decoder networks and other point set registration methods. Our code is publicly available at https://github.com/multimodallearning/deep-geo-reg.

Keywords: Deformable registration · Geometric learning · Belief propagation

1 Introduction

Current learning approaches for medical image analysis predominantly consider the processing of volumetric scans as a dense voxel-based task. However, the underlying anatomy could in many cases be modelled more efficiently using only a sparse subset of relevant geometric keypoints. When sufficient amounts of labelled training data are available and the region of interest can be robustly initialised, sparse surface segmentation models have been largely outperformed by dense fully-convolutional networks in the past few years [12]. However, dense learning based image registration has not yet reached the accuracy of conventional methods for the estimation of large deformations where geometry matters

A. Feragen et al. (Eds.): IPMI 2021, LNCS 12729, pp. 18–30, 2021.
https://doi.org/10.1007/978-3-030-78191-0_2

- e.g. for inspiration-expiration lung CT alignment. The combination of iconic (image-based) and geometric registration approaches have excelled in deformable lung registration but they are often time-consuming and rely on multiple steps of pre-alignment, mask-registration, graph-based optimisation and multi-level continuous refinement with different image-based metrics [22]. In this work, we aim to address 3D lung registration as a purely geometric alignment of two point clouds (a few thousand 3D points for inhale and exhale lungs each). While this certainly reduces the complexity of the dense deformable 3D registration task, it may also reduce the accuracy since intensity- and edge-based clues are no longer present. Yet, we demonstrate in our experimental validation that even this limited search range for potential displacements leads to huge and significant gains compared to dense learning based registration frameworks - mainly stemming from the robustness of our framework to implicitly learn the geometric alignment of vessel and airway trees.

1.1 Related Work

Point Cloud Learning: Conventional point cloud registration (iterative closest point, coherent point drift) [18] often focused on the direct alignment of unstructured 3D points based on their coordinates. Newer work on graph convolutional learning has demonstrated that relevant geometric features can be extracted from point clouds with neighbourhood relations defined on kNN graphs and enable semantic labeling or global classification of shapes, object parts and human poses and gestures [3,21]. Graph Convolutional Networks (GCN) [13] define localised filter and use a polynomial series of the graph Laplacian (Tschebyscheff polynomials) further simplified to the immediate neighbourhood of each node. The graph attention networks introduced in [26] are a promising extension based on attention mechanism. Similarly, dynamic edge convolutions [27] achieve information propagation by learning a function that predicts pairwise edge weights based on previous features of both considered nodes.

Learning Based Image Registration: In image registration, learning based methods have surpassed their untrained optimisation-based counterparts in terms of accuracy and speed for 2D optical flow estimation, where millions of realistic ground truth displacement fields can be generated [25]. Advantages have also been found for certain 3D medical registration tasks, for which thousands of scans with pixel-level expert annotations are available and the complexity of deformations is well represented in the training dataset [2,16,28]. As evident from a recent medical registration challenge [11], deep learning has not yet reached the accuracy and robustness for inspiration to expiration CT lung registration, where detailed anatomical labels are scarce (learning lobe alignment might not directly translate into low registration errors [10]) and the motion is large and complex. Even for the simpler case of shallow breathing in 4D CT, few learning-based works have come close to the best conventional methods (e.g. [22]) despite increasingly complex network pipelines [6].

Learning Graphical Registration: More recent research in computer vision has also explored geometric learning for 3D scene flow [14] that aims to register two 3D point clouds by finding soft correspondences. The challenge stems from the difficulty of jointly embedding two irregular point cloud (sub-)sets to enable end-to-end learning of geometric features and correspondence scores. Other recent approaches in point set registration/matching combine deep feature learning with GCNs and classical optimisation techniques, to solve the optimal transport [20] or reformulate traditional matching algorithms into deep network modules [23]. In the medical domain, combining sparse MRF-based registration [24] and multi-level continuous refinement [22] yielded the highest accuracy for two 3D lung benchmarks comprising inspiration and expiration [4,17].

We strongly believe that geometry can be a key element in advancing learning based registration and that the focus on visual features and fully-convolutional networks has for certain applications diverted research from mathematically proven graphical concepts that can excel within geometric networks.

1.2 Contribution

We propose a novel geometric learning method for large motion estimation across lung respiration that combines graph convolutional networks on keypoint clouds with sparse message passing. Our method considers geometric registration as soft correspondence search between two keypoint clouds with a restricted set of candidates from the moving point cloud for each fixed keypoint. **1)** We are the first to combine edge convolutions as end-to-end geometric feature learning from sparse keypoints with differentiable loopy belief propagation (discrete optimisation) for regularisation of displacements on a kNN graph adapted to irregular sets of candidates for each node. **2)** Our compact yet elegant networks, demonstrate surprisingly large gains in accuracy and outperform deep learning approaches that make use of additional visual clues by more than 50% reduced target registration errors for lung scans of COPD patients. **3)** We present a further novel variant of our approach that discretises the sparse correspondence probabilities using differentiable extrapolation for a further six fold gain in computational efficiency and with similar accuracy.

2 Methods

2.1 Loopy Belief Propagation for Regularised Registration of Keypoint Graphs

We aim to align two point clouds, a fixed point cloud P_f ($|P_f| = N_f$) and a moving point cloud P_m ($|P_m| = N_m$). They consist of distinctive keypoints $\mathbf{p}_{f_i} \in P_f$ and $\mathbf{p}_{m_i} \in P_m$. We further define a symmetric k-nearest neighbour (kNN) graph on P_f with edges $(ij) \in E$ that connect keypoints $\mathbf{p}_{f_i}$ and $\mathbf{p}_{f_j}$. A displacement vector $\mathbf{v}_i \in V$ for each fixed keypoint $\mathbf{p}_{f_i}$ is derived from soft correspondences from a restricted set of possible candidates $\mathbf{c}_i^p \in C_i$ (determined by l-nearest

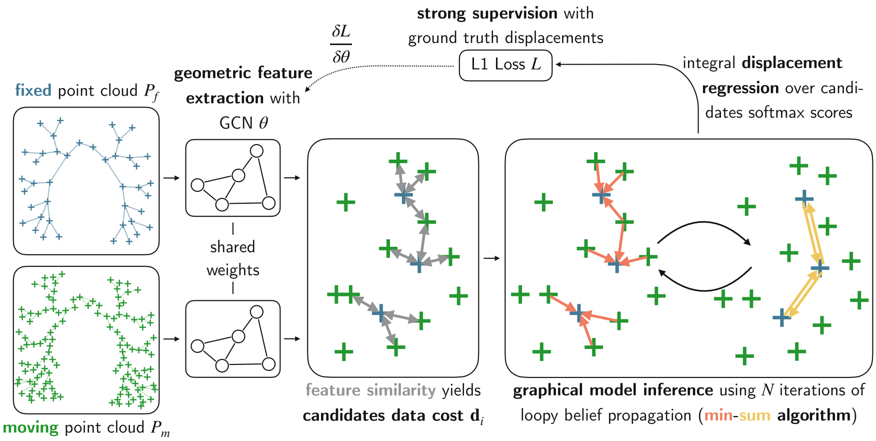
strong supervision with ground truth displacements
$\frac{\delta L}{\delta \theta}$
L1 Loss L
integral displacement regression over candidates softmax scores
geometric feature extraction with GCN θ
fixed point cloud P_f
shared weights
moving point cloud P_m
feature similarity yields candidates data cost $\mathbf{d}_i$
graphical model inference using N iterations of loopy belief propagation (min-sum algorithm)

Fig. 1. Overview of our proposed method for accurate point cloud alignment using geometric features combined with loopy belief propagation in an end-to-end trainable deep learning registration framework.

neighbour search ($|C_i| = l$) in the moving point cloud P_m). The regularised motion vector field V is inferred using loopy belief propagation enforcing spatial coherence of motion vectors. The data cost d_i^p ($\mathbf{d}_i = (d_i^1, \dots, d_i^p, \dots, d_i^l)$) for a fixed point $\mathbf{p}_{f_i}$ and a single candidate $\mathbf{c}_i^p$ is modeled as

$$d_i^p = \|\theta(\mathbf{p}_{f_i}) - \theta(\mathbf{c}_i^p)\|_2^2, \tag{1}$$

where $\theta(.)$ denotes a general feature transformation of the input point (e.g. deep learning based geometric features, cf. Sect. 2.2). Especially in this case of sparse to sparse inference, missing or noisy correspondences can lead to severe registration errors. Therefore, a robust regularisation between neighbouring fixed keypoints (defined by edges $(ij) \in E$) is enforced by penalizing the deviation of relative displacements. The regularisation cost r_{ij}^{pq} ($\mathbf{r}_{ij}^q = (r_{ij}^{1q}, \dots, r_{ij}^{pq}, \dots, r_{ij}^{lq})$) for two fixed keypoints $\mathbf{p}_{f_i}, \mathbf{p}_{f_j}$ and candidates $\mathbf{c}_i^p, \mathbf{c}_j^q$ can then be described as

$$r_{ij}^{pq} = \left\|(\mathbf{c}_i^p - \mathbf{p}_{f_i}) - (\mathbf{c}_j^q - \mathbf{p}_{f_j})\right\|_2^2. \tag{2}$$

To compute the marginal distributions of soft correspondences over the fixed kNN graph we employ N iterations of loopy belief propagation (min-sum algorithm) with outgoing messages $\mathbf{m}_{i\to j}^t$ from $\mathbf{p}_{f_i}$ to $\mathbf{p}_{f_j}$ at iteration t defined as

$$\mathbf{m}_{i\to j}^t = \min_{1,\dots,q,\dots l} \left(\mathbf{d}_i + \alpha \mathbf{r}_{ij}^q - \mathbf{m}_{j\to i}^{t-1} + \sum_{(h,i)\in E} \mathbf{m}_{h\to i}^{t-1}\right). \tag{3}$$

The hyperparameter α weights the displacement deviation penalty and thus controls the smoothness of the motion vector field V. Initial messages $\mathbf{m}_{i\to j}^0$ are set to 0. A graphical description of the presented message passing scheme is also

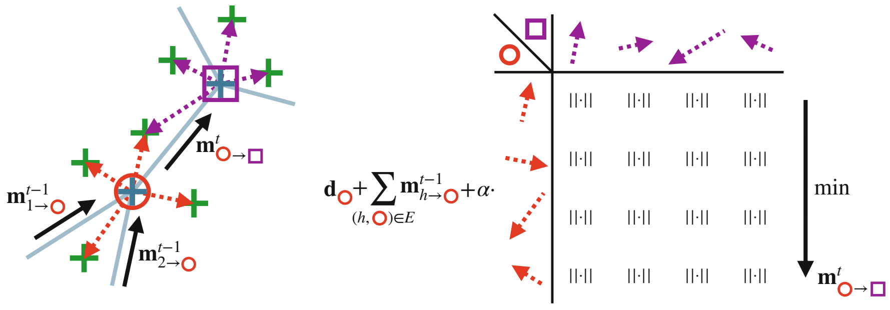
$\mathbf{m}^{t-1}_{1\to\bigcirc}$
$\mathbf{m}^{t}_{\bigcirc\to\square}$
$\mathbf{m}^{t-1}_{2\to\bigcirc}$
$\mathbf{d}_{\bigcirc}+\sum_{(h,\bigcirc)\in E}\mathbf{m}^{t-1}_{h\to\bigcirc}+\alpha\cdot$
$\|\cdot\|$
min
$\mathbf{m}^{t}_{\bigcirc\to\square}$

Fig. 2. Illustration of proposed message passing scheme for keypoint registration. The current outgoing message for the considered keypoint is composed of the candidates data cost and incoming messages from the previous iteration. In addition, the squared deviation (weighted by α) of candidate displacements is minimised for a coherent motion across the kNN graph. Reverse messages are not shown for visual clarity.

shown in Fig. 2 and for further in-depth details on efficient belief propagation the reader is referred to [5].

Fast Approximation Using a Discretised Candidates Space: While the proposed message passing approach is easily parallelisable, it still lacks some efficiency as the number of messages to compute for each keypoint is dependent on the number of neighbours k. We propose to reduce the number of message computations per node to 1 by discretising the sparse candidates cost $\mathbf{d}_i$ in a dense cost volume D_i with fixed grid resolution r. Voxelisation of sparse input has been used in point cloud learning to speed up computation [15]. D_i can be efficiently populated using nearest neighbour interpolation at (normalised) relative displacement locations $\mathbf{o}_\mathbf{i}^\mathbf{P} = (o^p_{i_x}, o^p_{i_y}, o^p_{i_z}) = \mathbf{c}^p_i - \mathbf{p}_{f_i}$, evaluating

$$D_i(u,v,w) = \frac{1}{N_{u,v,w}} \sum_{p=1}^{l} \mathbb{I}\big[\lfloor o^p_{i_x} r\rfloor = u,\, \lfloor o^p_{i_y} r\rfloor = v,\, \lfloor o^p_{i_z} r\rfloor = w\big] d^p_i, \tag{4}$$

where (following notations in [15]) $\mathbb{I}[\cdot]$ denotes a binary indicator that specifies whether the location $\mathbf{o}_\mathbf{i}^\mathbf{P}$ belongs to the voxel grid (u,v,w) and $N_{u,v,w}$ is a normalisation factor (in case multiple displacements end up in the same voxel grid). By operating on the dense displacement space D_i, we can employ an efficient quadratic diffusion regularisation using min convolutions [5] that are separable in dimensions and also avoid the costly computation of k different messages per node. Approximation errors stem solely from the discretisation step.

2.2 Geometric Feature Extraction with Graph Convolutional Neural Networks

Distinctive keypoint graphs that describe plausible shapes contain inherent geometric information. These include local features such as curvature but also global semantics of the graph (e.g. surface or structure connectivity). Recent work on data-driven graph convolutional learning has shown that descriptive geometric features can be extracted from point clouds with neighbourhood relations defined based on kNN graphs. Edge convolutions [27] can be interpreted as irregular equivalents to dense convolutional kernels. Following notations in [27] we define edge features $\mathbf{e}_{ij} = h_\theta(\mathbf{f}_i, \mathbf{f}_j - \mathbf{f}_i)$, where $\mathbf{f}_i$ denote F-dimensionsal features on points $\mathbf{p}_i \in P$ (first feature layer given as $\mathbf{f}_i = \mathbf{p}_i$). The edge function h_θ computes the Euclidean inner product of the learnable parameters $\theta = (\theta_1, \ldots, \theta'_F)$ with $\mathbf{f}_i$ (keypoint information) and $\mathbf{f}_j - \mathbf{f}_i$ (local neigbourhood information). The F'-dimensional feature output $\mathbf{f'}_i$ of an edge convolution is then given by

$$\mathbf{f}_i^{'} = \max_{(i,j)\in E} \mathbf{e}_{ij}, \tag{5}$$

where the max operation is to be understood as a dimension-wise aggregation function. Employing multiple layers of edge convolutions in a graph neural network and applying it to the fixed and moving point clouds (P_f, P_m) yields descriptive geometric features, which can be directly used to compute candidate data costs (see Eq. 1).

2.3 Deep Learning Based End-to-End Geometric Registration Framework

Having described the methodological details, we now summarise the full end-to-end registration framework (see Fig. 1 for an overview). Input to the registration framework are the fixed P_f and moving P_m point cloud. In a first step, descriptive geometric features are extracted from P_f and P_m with a graph convolutional network θ (shared weights). The network consists of three edge convolutional layers, whereby edge functions are implemented as three layers of 1×1 convolutions, instance normalisation and leaky ReLUs. Feature channels are increased from 3 to 64. Two 1×1 convolutions output the final 64-dimensional point feature embeddings. Thus, the total number of free trainable parameters of the network is 26880. In general, the moving cloud will contain more points than the fixed cloud (to enable an accurate correspondence search). To account for this higher density of P_m, the GCN θ acts on the kNN graph for P_f and on the $3k$NN graph for P_m. As described in Sect. 2.1 the geometric features $\theta(P_f)$ and $\theta(P_m)$ are used to compute the candidates cost and final marginal distributions are obtained from N iterations of (sparse or discretised) loopy belief propagation. As all operations in our optimisation step are differentiable the network parameters can be trained end-to-end. The training is supervised with ground truth motion vectors $\hat{\mathbf{v}}_i \in \hat{V}$ (based on 300 available manual annotated and corresponding landmark pairs) using an L1 loss (details on integral regression of the predicted motion vectors V from the marginals in Sect. 2.4).

2.4 Implementation Details: Keypoints, Visual Features and Integral Loss

While our method is generally applicable to a variety of point cloud tasks, we adapted parts of our implementation to keypoint registration of lung CT.

Keypoints: We extract Förstner keypoints with non-maximum suppression as described in [8]. A corner score (distinctiveness volume) is computed using $D(x) = 1/\operatorname{trace}\left((G_\sigma * (\nabla F \nabla F^T))^{-1}\right)$, where G_σ describes a Gaussian kernel and ∇F spatial gradients of the fixed/moving scans computed with a seven-point stencil. Additionally, we modify the extraction to allow for a higher spatial density of keypoints in the moving scan by means of trilinear upsampling of the volume before non-maximum suppression. Only points within the available lung masks are considered.

Visual Features: To enable a fair comparison to state-of-the-art methods that are based on image intensities, we also evaluate variants of all geometric registration approaches with local MIND-SSC features [9]. These use a 12-channel representation of local self-similarity and are extracted as small patches of size $3 \times 3 \times 3$ with stride $= 2$. The dimensionality is then further reduced from 324 to 64 using a PCA (computed on each scan pair independently).

Integral Loss: As motivated before, we aim to find soft correspondences that enable the estimation of relative displacements, without directly matching a moving keypoint location, but rather a probability for each candidate. A softmax operator over all candidates is applied to the negated costs after loopy belief propagation (multiplied by a heuristic scalar factor). These normalised predictions are integrated over the corresponding relative displacements. When considering a discretised search space (the dLBP variant), final displacements are obtained via integration over the fully quantised 3D displacement space.

To obtain a dense displacement field for evaluation (landmarks do not necessarily coincide with keypoints), all displacement vectors of the sparse keypoints are accumulated in a displacement field tensor using trilinear extrapolation and spatial smoothing. This differentiable dense extrapolation enables the use of an L1 loss on (arbitrary) ground truth correspondences.

3 Experiments and Results

To demonstrate the effectiveness of our novel learning-based geometric 3D registration method, we perform extensive experimental validation on the DIR-Lab COPDgene data [4] that consists of 10 lung CT scan pairs at full inspiration (fixed) and full expiration (moving), annotated with 300 expert landmarks each. Our focus lies in evaluating point cloud registration without visual clues and we extract a limited number of keypoints (point clouds) in fixed (≈2000 each) and moving scans (≈6000 each) within the lungs. Since, learning benefits from a variability of data, we add 25 additional 3D scan pairs showing inhale-exhale CT from the EMPIRE10 [17] challenge, for which no landmarks are publicly

Table 1. Results of methods based on geometric features and optimisation on the COPDgene dataset [4]. We report the target registration error (TRE) in millimeters for individual cases as well as the average distance and standard deviation over all landmarks. The average GPU runtime in seconds is listed in the last row.

	Init	CPD	CPD+GF	sLBP	sLBP+GF (ours)	dLBP+GF (ours)
# 01	26.33	3.02	2.75	2.55	**1.88**	2.14
# 02	21.79	10.83	**5.96**	8.69	6.22	6.69
# 03	12.64	1.94	1.88	1.56	**1.53**	1.68
# 04	29.58	2.89	2.84	3.57	**2.63**	3.01
# 05	30.08	3.01	2.70	3.01	**2.02**	2.42
# 06	28.46	3.22	3.65	2.85	**2.21**	2.69
# 07	21.60	2.52	2.44	1.87	**1.64**	1.83
# 08	26.46	3.85	3.58	2.08	**1.93**	2.14
# 09	14.86	2.83	2.58	1.53	**1.55**	1.82
# 10	21.81	3.57	5.57	3.15	**2.79**	3.72
Avg.	23.36	3.77	3.40	3.08	**2.44**	2.81
Std.	11.86	2.54	1.35	2.09	1.40	1.50
Time		7.63	7.66	2.91	3.05	**0.49**

available and we only include automatic correspondences generated using [8] for supervision. We performed leave-one-out cross validation on the 10 COPD scans with sparse-to-dense extrapolation for landmark evaluation. Training was performed with a batch size of 4 and an initial learning rate of 0.01 for 150 epochs. All additional hyperparamters for baselines and our proposed methods (regularisation cost weighting α, scalar factor for integral loss, etc.) were tuned on case #04 of the COPDgene dataset and left unaltered for the remaining folds.

Overall, we compare five different algorithms that work purely on geometric information, five further methods that use visual input features and one deep-learning baseline for dense intensity registration (the winner of the Learn2Reg 2020 challenge LapIRN [16]). Firstly, we compare our proposed sparse-LBP regularisation with geometric feature learning (sLBP+GF) to a version without geometric learning (sLBP) and coherent point drift [18] without (CPD) and with geometric feature learning (CPD+GF). The non-learning based methods directly use the keypoint coordinates (x, y, z) as input features. In addition, we evaluate the novel discretisation of sparse candidates that is again integrated into an end-to-end geometric learning with differentiable LBP regularisation (dLBP+GF) and leads to substantial efficiency gains. The results clearly demonstrate the great potential of keypoint based registration for the complex task of large deformable lung registration. Numerical and qualitative results are shown in Table 1 and Fig. 3, respectively. Even the baseline methods using no features at all, CPD and sLBP, where inference is based only on optimisation on the extracted keypoint graphs, achieve convincing target registration errors

Table 2. Results of methods based on visual features on the COPDgene dataset [4]. We report the average TRE and standard deviation in millimeters over all landmarks. The average GPU runtime in seconds is listed in the last column. For easier comparison we also add the results of our "geometry only" approaches.

	Avg	Std	Time
init	23.36	11.86	
FLOT+MIND	5.87	1.30	1.63
LapIRN	4.99	1.98	1.08
FE+MIND	3.83	1.21	16.71
sPDD+MIND	3.16	0.69	2.17
CPD+MIND	2.40	0.81	13.12
sLBP+MIND (ours)	**1.74**	0.38	4.65
sLBP+GF (ours)	2.44	1.40	3.05
dLBP+GF (ours)	2.81	1.50	**0.49**

of 3.77 mm and 3.08 mm. Adding learned geometric features within our proposed geometric registration framework leads to relative improvements of 10% (CPD+GF) and 20% (sLBP+GF), respectively. For the efficient approximation of our proposed appraoch (dLBP+GF) the TRE increases by approximate 0.35 mm but at the same time the average runtime is improved six fold to just below 0.5 s (which is competitive with dense visual deep learning methods such as LapIRN (cf. Table 2)). A statistical test (Wilcoxon signed-rank test calculated over all landmark pairs) with respect to our proposed method (sLBP+GF) shows that improvements on all other comparison methods are highly significant ($p < 0.001$).

We made great efforts to use state-of-the-art learning-based 3D scene flow registration methods and obtained only meaningful results when incorporating the visual MIND features for FLOT [20] and heavily adapting the FlowNet3d embedding strategy [14] (denoted as FE+MIND). FlowNet3d aims to learn a flow embeddings (FE) using a concatenation of two candidate sets (from connected graph nodes), which does not lead to satisfactory results due to the permutation invariant nature of these sparse candidates. Hence, we designed a layer that captures all pairwise combinations and leads to a higher dimensional intermediate tensor that is fed into 1×1 convolutions and is projected (with max-pooling) to a meaningful message vector. For FLOT, we replaced the feature extraction with the handcrafted MIND-PCA embeddings and also removed the refinement convolutions after the optimal transport block (we observed severe overfitting in our training setting when employing the refinement). The sPDD method is based on the probabilistic dense displacement (PDD) network and was modified to operate on the sparse fixed keypoints (instead of a regular grid as in the original published work [7]). Results for the state-of-the-art learning based 3D scene flow registration methods and further comparison experiments using visual input features can be found in Table 2. Our proposed sparse

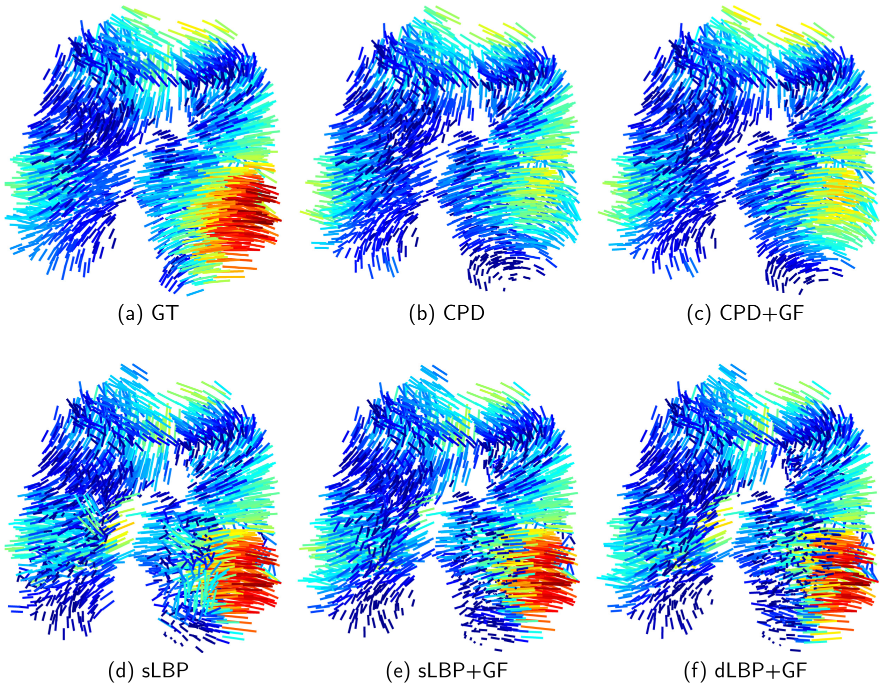
(a) GT
(b) CPD
(c) CPD+GF
(d) sLBP
(e) sLBP+GF
(f) dLBP+GF

Fig. 3. Qualitative results of different geometric methods ((b)–(f)) on case # 01 of the COPDgene dataset [4]. The ground truth motion vector field is shown in (a). Different colors encode small (blue) and large motion (red). (Color figure online)

registration approach using visual MIND features (sLBP+MIND) achieves a TRE well below 2 mm and thus, improves on the geometry based equivalent (sLBP+GF) by 0.7 mm. However, the extraction of visual features slows down the inference time by 1.6 and 4.1 (dLBP+GF) seconds, respectively. Notably, all proposed geometric registration methods achieve results on par with or significantly better (e.g. more than 50% gain in target registration error w.r.t the dense multi-scale network LapIRN) than the deep learning based comparison methods with additional visual features. Conventional registration methods achieve TREs around 1 to 1.5 mm with runtimes of 3 to 30 min [1,8,22].

4 Discussion and Conclusion

We believe our concept clearly demonstrates the advantages of decoupling feature extraction and optimisation by combining parallelisable differentiable message passing for sparse correspondence finding with graph convolutions for geometric feature learning. Our method enables effective graph-based pairwise regularisation and compact networks for robustly capturing geometric context for large deformation estimation. It is much more capable for 3D medical image registration as adaptations of scene flow approaches, which indicates that these methods may be primarily suited for aligning objects with repetitive semantic object/shape parts that are well represented in large training databases.

We demonstrated that even without using visual features, the proposed geometric registration substantially outperforms very recent deep convolutional registration networks that excelled in other medical tasks. The reason for this large performance gap can firstly lie in the complexity of aligning locally ambiguous structures (vessels, airways) that undergo large deformations and that focusing on relatively few relevant 3D keypoints is a decisive factor in learning meaningful geometric transformations. Our new idea to discretise the sparse candidate displacements into a dense embedding using differentiable extrapolation yields immensive computational gains by reducing the number of message computations (from $k = 9$ to 1 per node) and thereby also enabling future use within alternative regularisation algorithms.

While our experimental analysis was so far restricted to lung anatomies, we strongly believe that graph-based regularisation models combined with geometric learning will play an important role for tackling other large motion estimation tasks, the alignment of anatomies across subjects for studying shape variations and tracking in image-guided interventions. Being able to work independently of visual features opens new possibilities for multimodal registration, where our method only requires comparable keypoints to be found, e.g. using probabilistic edge maps [19]. In addition, the avoidance of highly parameterised CNNs can establish new concepts to gain a better interpretability of deep learning models.

References

1. Avants, B.B., Epstein, C.L., Grossman, M., Gee, J.C.: Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. Med. Image Anal. (MedIA) **12**(1), 26–41 (2008)
2. Balakrishnan, G., Zhao, A., Sabuncu, M.R., Guttag, J., Dalca, A.V.: VoxelMorph: a learning framework for deformable medical image registration. IEEE Trans. Med. Imaging (TMI) **38**(8), 1788–1800 (2019)
3. Bronstein, M.M., Bruna, J., LeCun, Y., Szlam, A., Vandergheynst, P.: Geometric deep learning: going beyond Euclidean data. IEEE Signal Process. Mag. (SPM) **34**(4), 18–42 (2017)
4. Castillo, R., et al.: A reference dataset for deformable image registration spatial accuracy evaluation using the COPDgene study archive. Phys. Med. Biol. **58**(9), 2861 (2013)
5. Felzenszwalb, P.F., Huttenlocher, D.P.: Efficient belief propagation for early vision. Int. J. Comput. Vis. (IJCV) **70**(1), 41–54 (2006). https://doi.org/10.1007/s11263-006-7899-4
6. Fu, Y., et al.: LungRegNet: an unsupervised deformable image registration method for 4d-CT lung. Med. Phys. **47**(4), 1763–1774 (2020)
7. Heinrich, M.P.: Closing the gap between deep and conventional image registration using probabilistic dense displacement networks. In: Shen, D., et al. (eds.) MICCAI 2019, Part VI. LNCS, vol. 11769, pp. 50–58. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32226-7_6

8. Heinrich, M.P., Handels, H., Simpson, I.J.A.: Estimating large lung motion in COPD patients by symmetric regularised correspondence fields. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015, Part II. LNCS, vol. 9350, pp. 338–345. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24571-3_41
9. Heinrich, M.P., Jenkinson, M., Papież, B.W., Brady, S.M., Schnabel, J.A.: Towards realtime multimodal fusion for image-guided interventions using self-similarities. In: Mori, K., Sakuma, I., Sato, Y., Barillot, C., Navab, N. (eds.) MICCAI 2013, Part I. LNCS, vol. 8149, pp. 187–194. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-40811-3_24
10. Hering, A., Häger, S., Moltz, J., Lessmann, N., Heldmann, S., van Ginneken, B.: Constraining volume change in learned image registration for lung CTs. arXiv preprint arXiv:2011.14372 (2020)
11. Hering, A., Murphy, K., van Ginneken, B.: Learn2Reg Challenge: CT Lung Registration - Training Data, May 2020
12. Isensee, F., Jäger, P.F., Kohl, S.A., Petersen, J., Maier-Hein, K.H.: nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nat. Methods **18**(2), 203–211 (2020)
13. Kipf, T.N., Welling, M.: Semi-supervised classification with graph convolutional networks. In: International Conference on Learning Representations (ICLR) (2017)
14. Liu, X., Qi, C.R., Guibas, L.J.: Flownet3D: Learning scene flow in 3D point clouds. In: International Conference on Computer Vision and Pattern Recognition (CVPR), pp. 529–537 (2019)
15. Liu, Z., Tang, H., Lin, Y., Han, S.: Point-voxel CNN for efficient 3D deep learning. In: Advances in Neural Information Processing Systems (NeurIPS) pp. 965–975 (2019)
16. Mok, T.C.W., Chung, A.C.S.: Large deformation diffeomorphic image registration with Laplacian pyramid networks. In: Martel, A.L., et al. (eds.) MICCAI 2020, Part III. LNCS, vol. 12263, pp. 211–221. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-59716-0_21
17. Murphy, K., et al.: Evaluation of registration methods on thoracic CT: the EMPIRE10 challenge. IEEE Trans. Med. Imaging (TMI) **30**(11), 1901–1920 (2011)
18. Myronenko, A., Song, X.: Point set registration: coherent point drift. IEEE Trans. Pattern Anal. Mach. Intell.(TPAMI) **32**(12), 2262–2275 (2010)
19. Murphy, O., et al.: Structured decision forests for multi-modal ultrasound image registration. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015, Part II. LNCS, vol. 9350, pp. 363–371. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24571-3_44
20. Puy, G., Boulch, A., Marlet, R.: FLOT: scene flow on point clouds guided by optimal transport. In: Vedaldi, A., Bischof, H., Brox, T., Frahm, J.-M. (eds.) ECCV 2020, Part XXVIII. LNCS, vol. 12373, pp. 527–544. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-58604-1_32
21. Qi, C.R., Su, H., Mo, K., Guibas, L.J.: PointNet: deep learning on point sets for 3D classification and segmentation. In: International Conference on Computer Vision and Pattern Recognition (CVPR), pp. 652–660 (2017)
22. Rühaak, J., Polzin, T., Heldmann, S., Simpson, I.J., Handels, H., Modersitzki, J., Heinrich, M.P.: Estimation of large motion in lung CT by integrating regularized keypoint correspondences into dense deformable registration. IEEE Trans. Med. Imaging (TMI) **36**(8), 1746–1757 (2017)

23. Sarlin, P.E., DeTone, D., Malisiewicz, T., Rabinovich, A.: Superglue: earning feature matching with graph neural networks. In: International Conference on Computer Vision and Pattern Recognition (CVPR), pp. 4938–4947 (2020)
24. Sotiras, A., Ou, Y., Glocker, B., Davatzikos, C., Paragios, N.: Simultaneous geometric - iconic registration. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010, Part II. LNCS, vol. 6362, pp. 676–683. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-15745-5_83
25. Sun, D., Yang, X., Liu, M.Y., Kautz, J.: PWC-net: CNNs for optical flow using pyramid, warping, and cost volume. In: International Conference on Computer Vision and Pattern Recognition (CVPR), pp. 8934–8943 (2018)
26. Veličković, P., Cucurull, G., Casanova, A., Romero, A., Liò, P., Bengio, Y.: Graph attention networks. In: International Conference on Learning Representations (ICLR) (2018)
27. Wang, Y., et al.: Dynamic graph CNN for learning on point clouds. ACM Trans. Graph. (TOG) **38**(5), 1–12 (2019)
28. Xu, Z., Niethammer, M.: DeepAtlas: Joint Semi-Supervised Learning Of Image Registration And Segmentation. In: Shen, D., et al. (eds.) MICCAI 2019, Part II. LNCS, vol. 11765, pp. 420–429. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32245-8_47