



Non-rigid point cloud registration for middle ear diagnostics with endoscopic optical coherence tomography

Peng Liu^{1,3} · Jonas Golde^{2,3} · Joseph Morgenstern^{3,4} · Sebastian Bodenstedt¹ · Chenpan Li¹ · Yujia Hu¹ · Zhaoyu Chen¹ · Edmund Koch^{1,2} · Marcus Neudert^{1,4} · Stefanie Speidel^{1,3}

Received: 6 March 2023 / Accepted: 12 May 2023 / Published online: 17 June 2023
© The Author(s) 2023

Abstract

Purpose Middle ear infection is the most prevalent inflammatory disease, especially among the pediatric population. Current diagnostic methods are subjective and depend on visual cues from an otoscope, which is limited for otologists to identify pathology. To address this shortcoming, endoscopic optical coherence tomography (OCT) provides both morphological and functional in vivo measurements of the middle ear. However, due to the shadow of prior structures, interpretation of OCT images is challenging and time-consuming. To facilitate fast diagnosis and measurement, improvement in the readability of OCT data is achieved by merging morphological knowledge from ex vivo middle ear models with OCT volumetric data, so that OCT applications can be further promoted in daily clinical settings.

Methods We propose C2P-Net: a two-staged non-rigid registration pipeline for complete to partial point clouds, which are sampled from ex vivo and in vivo OCT models, respectively. To overcome the lack of labeled training data, a fast and effective generation pipeline in Blender3D is designed to simulate middle ear shapes and extract in vivo noisy and partial point clouds.

Results We evaluate the performance of C2P-Net through experiments on both synthetic and real OCT datasets. The results demonstrate that C2P-Net is generalized to unseen middle ear point clouds and capable of handling realistic noise and incompleteness in synthetic and real OCT data.

Conclusions In this work, we aim to enable diagnosis of middle ear structures with the assistance of OCT images. We propose C2P-Net: a two-staged non-rigid registration pipeline for point clouds to support the interpretation of in vivo noisy and partial OCT images for the first time. Code is available at: https://gitlab.com/nct_tso_public/c2p-net.

Keywords Middle ear · Optical coherence tomography · Point cloud · Deep learning based registration

Introduction

The middle ear consists of the tympanic membrane (TM) and an air-filled chamber containing the ossicle chain (OC) that connects the TM to the inner ear. From a functional

perspective, the middle ear matches the impedance from air to the fluid-filled inner ear [1].

Middle ear disorders contain deformation, discontinuation of TM and OC, effusion, and cholesteatoma in the middle ear. These may occur because of middle ear infection (e.g., acute otitis media (AOM), chronic otitis media (COM), otitis media with effusion (OME)) and trauma [2, 3]. Most commonly, serious and chronic middle ear disorder may lead to conductive hearing loss and inner ear disorder [4]. Current middle ear diagnostics methods or tools cover only aspects of the pathology and cannot determine the origin and site of transmission loss, e.g., otoscopy, audiometry, tympanometry, etc.

As an innovative image technology, endoscopic OCT enables the examination of both the morphology and function of the middle ear in vivo by the non-invasive acquisition of depth-resolved and high-resolution images [5–7]. Neverthe-

✉ Peng Liu
peng.liu@nct-dresden.de

¹ Translational Surgical Oncology, National Center for Tumor Diseases, Dresden 01307, Germany

² Clinical Sensing and Monitoring, TU Dresden, Dresden 01307, Germany

³ Else Kröner Fresenius Center, TU Dresden, Dresden 01307, Germany

⁴ Ear Research Center Dresden, TU Dresden, Dresden 01307, Germany

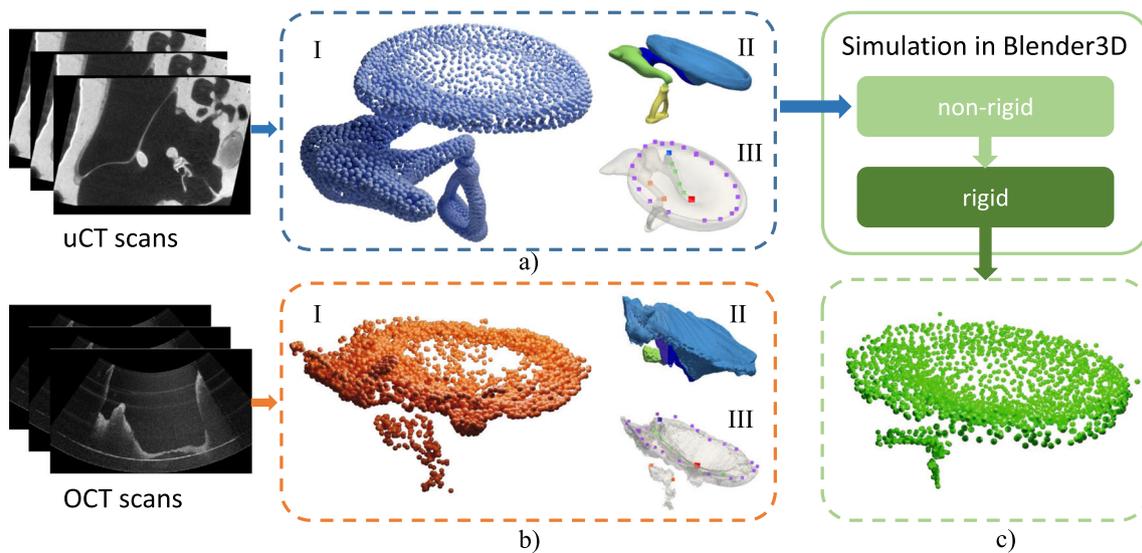


Fig. 1 Overview of data samples. In **a**, I is the ex vivo template point cloud reconstructed from μ CT scans on the left. II is the segmented middle ear model, and each color stands for one structure. III shows the model with sparse labels. Accordingly, I, II and III in **b** present in

vivo OCT point cloud, segmented OCT model and labels. The simulation pipeline on the top left takes a complete ex vivo ear point cloud, performs non-rigid and rigid simulation, and produces synthetic partial and noisy in vivo point cloud shown in **c**

less, multiple sources of noise, e.g., the shadow of preceding structures, reduce the quality of the target structures further away from the endoscopic probe, e.g., ossicles. Therefore, the reconstructed 3D models from the OCT volumetric data are usually noisy and difficult to interpret, especially the identification of the deeper middle ear structures such as incus and stapes (see Fig. 1).

Our aim is to improve the interpretation of the OCT volumetric data by merging the ex vivo middle ear model reconstructed from micro-Computed Tomography (μ CT) scan of isolated temporal bones and the in vivo OCT model. Note that a single μ CT model was used as a template and fitted to all patient-specific OCT scans. We first convert all ex and in vivo middle ear data to point cloud representations for ease of flexible manipulation of middle ear shapes. Nevertheless, finding one-to-one correspondence between such a point cloud pair is still challenging due to the noise and incompleteness of the OCT model, and the difference between the patient data and the template μ CT data. To tackle these issues, we first employ a neural network that searches the sparse correspondences for the source and target points, then a pyramid deformation algorithm to fit the points in a non-rigid fashion based on the predicted correspondences. The neural network is trained on randomly generated synthetic shape variants of the middle ear that contain random noise and only partial points. This enables the generalization of the neural network to new patient data as well as the adaption to noise and outliers. The main contributions of this work are listed as follows:

1. A generation pipeline in Blender3D that simulates synthetic shape variants from a complete ex vivo middle ear model, and simulates noisy and partial point clouds comparable to in vivo data.
2. C2P-Net: a two-staged non-rigid registration pipeline for point clouds. It demonstrates that C2P-Net registers the complete point cloud template to partial point clouds of the middle ear effectively and robustly.

Related work

In this section, we review methods that analyze 3D point clouds and solve registration problems. Learning the geometry of 3D point clouds is the foundation of diverse 3D applications, but is also a challenging task due to the irregularity and asymmetry of point clouds. Recently, approaches that focus on learning point features with neural networks have been widely investigated to overcome such issues. One category of methods is to project the point cloud to a regular representation, e.g., voxel grids [8], 2D images from multiviews [9, 10], or combined [11] that 2/3D convolutional operations can be performed on top of these intermediate data in Euclidean space. Apart from these, PointNet [12] leverages multilayer perceptron (MLP) to obtain pointwise features and a max-pool to aggregate global features but without local information. This drawback is alleviated by PointNet++ [13], which uses PointNet to aggregate local fea-

tures in a multi-scale structure. In addition, various explicit convolution kernels are applied directly on the point clouds to extract encodings [14–16] in which point features are extracted based on the kernel weights.

Conventional methods solve the point cloud registration task as an optimization problem of transformation parameters. Iterative Closest Point (ICP) [17] iteratively calculates a rigid transformation matrix based on updated correspondence set from a last iteration, Non-rigid Iterative Closest Point (NICEP) [18] achieves non-rigid registration by optimizing an energy function including local regularization and stiffness. Coherent Point Drift (CPD) maximizes the probability of Gaussian Mixture Model (GMM) by moving coherent GMM centroids from the two point clouds. Recently, deep learning-based methods using the aforementioned point cloud encoding techniques have been widely investigated [19–22]. PointNetLK [19] extracts semantic point cloud features using an MLP, then performs rigid alignment using the Lucas and Kanade algorithm [23].

RegTR [20] and NgeNet [21] employ KPConv [14] to extract both global and local features from the source and target point clouds. Following this, attention mechanisms are introduced to enable the communication of the extracted features and estimate a global transformation from the predicted correspondences of the MLPs.

In contrast to this direct way, NDP [22] decomposes the global transformation into sub-motions which are iteratively refined using MLPs at each pyramid level. Inspired by this recent work regarding point cloud analysis and registration, we adopted NgeNet [21] and NDP [22] as the main components to construct our registration pipeline.

Methods

The non-rigid registration of the ex and in vivo middle ear point clouds (see Fig. 1) using neural networks is realized in a supervised learning fashion. Due to lack of real OCT data with ground truth, and the difficulties of finding one-to-one correspondences between the general and ex vivo model and the noisy and partial in vivo model, we generated synthetic in vivo middle ear point clouds based on the ex vivo μ CT model using a two-step simulation pipeline. For point cloud registration, we propose C2P-Net, which is a two-stage registration method. We first train a neural network on the synthetic samples to explore the correspondence along with an initial rigid transformation matrix, then based on this, a hierarchical algorithm is performed to estimate the non-rigid deformation.

Synthetic data generation

We start with a complete middle ear model $M_{\mu CT}$ reconstructed from μ CT volume as basis for the simulation in

Blender3D. The non-rigid simulation is performed with the assistance of lattice modifiers attached to each structure with reasonable resolution. The lattice vertices are assigned to various groups according to length, thickness, and width of the global and local shape of a structure. By transforming and rotating various vertex groups with random parameters with boundary conditions, a random shape of a structure can be generated, e.g., a relatively shorter malleus with a longer lateral process. The non-rigid simulation step is formulated as $\tilde{T}_{nr} = S_{nr}(T, p_{nr})$, where T is the template middle ear, p_{nr} are the input parameters, and \tilde{T}_{nr} is the simulated non-rigid shape variants of T . Next, armature modifiers are attached to each structure and connected in an end-to-end fashion at the articulations. Then, rigid simulation is accomplished by transforming and rotating the armature bones. We denote the rigid simulation as S_r . Thus, a random shape variant of the middle ear \tilde{T} can be obtained which is considered as the in vivo ear model of a new patient: $\tilde{T} = S_r(\tilde{T}_{nr}, p_r)$.

In practice, the in vivo OCT model is usually noisy and only partially visible, which limits the registration methods in finding accurate and robust correspondences. To simulate the real data, we extract random partial patches for each structure of the shape variant \tilde{T} by a combined probability, which is calculated as the distance of each vertex to the support point of each structure and random Gaussian noise. Additionally, for the sake of simulating incomplete posterior structure, e.g., stapes, caused by the occlusion and shadow from the anterior structures, e.g., ear canal wall, we determine the number of points of the posterior structures by the distance to the external ear and a random factor. Furthermore, uniform random displacements are applied to the vertices to generate the final noisy and partial in vivo point cloud P_{inv} (see Fig. 1).

C2P-Net

We delineate the architecture of C2P-Net in Fig. 2. It consists of two components dedicated to two stages: initial rigid registration and pyramid non-rigid registration. Given an ex vivo point cloud as a template extracted from a μ CT model: $P_{exv} = \{x_i \in \mathcal{R}^3\}_{i=1,2,\dots,N}$, and a partial point cloud of the simulated in vivo shape variant: $P_{inv} = \{y_j \in \mathcal{R}^3\}_{j=1,2,\dots,M}$, we adopted the Neighborhood-aware Geometric Encoding Network (NgeNet) [21] to solve the initial rigid registration task. This stage is formulated as:

$$\tau, \sigma = \text{NgeNet}(P_{exv}, P_{inv}) \quad \text{where } (u, v) \in \sigma \quad (1)$$

where $\tau \in SE(3)$ is the rigid transformation matrix which aligns P_{exv} with P_{inv} , and $(u, v) \in \sigma$ is the sparse correspondence set where u and v are indices for points in P_{exv} and P_{inv} . Due to the multi-scale structure and a voting mechanism integrating features from different resolutions, NgeNet handles noise well and predicts correspondence robustly.

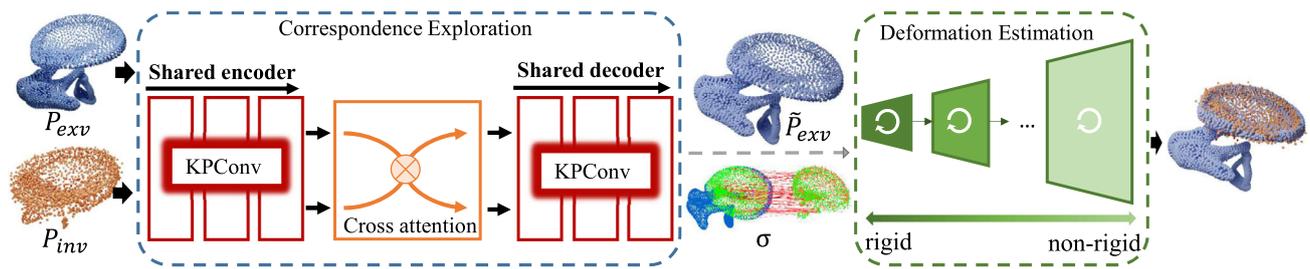


Fig. 2 Schematic architecture of C2P-Net. The input of the pipeline is the ex vivo (blue) and in vivo (orange) point cloud. To explore the correspondence between the two inputs, NgeNet is trained on synthetic samples to extract the features for each point at different levels. These are integrated by a voting algorithm. A feature matching routine is performed to obtain the sparse correspondence set as well as a

rigid transformation matrix. On top of the previous outputs, NDP maps the input point clouds to sinusoidal space and decomposes the global non-rigid deformation at each pyramid level into sub-motions. With the increase in sinusoidal frequency, the non-rigid deformation of each point is continuously refined at each level. The registered point cloud is shown on the right side

Based on the previous predicted correspondence set and the rigidly aligned source and target point clouds, we employ the Neural Deformation Pyramid (NDP) [22] to predict the non-rigid deformation of the given point cloud pair. NDP defines the non-rigid registration problem as a hierarchical motion decomposition problem. At each pyramid level, the input points from last level are mapped to sinusoidal encodings with different frequencies: $\Gamma(p^{k-1}) = (\sin(2^{k+k_0} p^{k-1}), \cos(2^{k+k_0} p^{k-1}))$, k is the current level number, k_0 controls the initial frequency and p^{k-1} is an output point from the last level. Lower frequencies at shallower levels represent rigid sub-motion, while higher frequencies at deep levels emphasize non-rigid deformations. In this way, a sequence of sub-motions is estimated from rigid to non-rigid, and the final displacement field is the combination of such a sequence. Formally, we denote the stage as:

$$\phi_{est} = \text{NDP}_{n,m}(\tilde{P}_{exv}, P_{inv}, \sigma) \quad (2)$$

where \tilde{P}_{exv} is the source point cloud P_{exv} transformed by τ , n is the number of pyramid layers of the NDP neural network, m is the maximal iteration within a single pyramid layer, and ϕ_{est} is the predicted displacement field describing how each point should move to the target. Combined losses are calculated at each iteration, including correspondence loss and regularization loss, and back-propagated to update the weights of each MLP. Of which, the correspondence loss L_{CD} is defined as the Chamfer distance (3) between \tilde{P}_{exv} which is masked by the correspondence σ and P_{inv} .

$$\text{CD}(A, B) = \frac{1}{|A|} \sum_{x_i \in A} \min_{y_j \in B} |x_i - y_j| + \frac{1}{|B|} \sum_{y_j \in B} \min_{x_i \in A} |x_i - y_j| \quad (3)$$

$$L_{CD} = \text{CD}(\tilde{P}_{exv}^\sigma, P_{inv}) \quad (4)$$

$\tilde{P}_{exv}^\sigma = \{\tilde{P}_{exv}[u] \mid \exists v : (u, v) \in \sigma\}$ are the masked ex vivo points that have correspondence in the target in vivo point cloud.

Evaluation

C2P-Net predicts a displacement field to fit the ex vivo point cloud template to the target in vivo point cloud which is noisy and partial. At training time, C2P-Net employs an Adam optimizer and ExpLR scheduler and uses supervised learning to learn on a synthetic dataset with 20,000 in vivo shape variants of the middle ear. During inference, it takes around 3.5 s for C2P-Net to react to a new in vivo shape on RTX 2070 Super. Since the neural network is trained on synthetic data, it is crucial to investigate the generalization of the neural networks on unseen samples as well as the performance on real OCT data. Thus, we conduct two experiments on synthetic data and real OCT data with various metrics: mean displacement error (5), Chamfer distance (3), and landmarks error (6). Furthermore, we compare our results with other popular non-rigid registration methods, e.g., Non-rigid ICP (NIPC) and Coherent Point Drift (CPD).

$$M_{MDE} = \frac{1}{N} \sum_i \|\phi_{est} - \phi_{gt}\| \quad (5)$$

N is the size of the test dataset, ϕ_{est} is the estimated displacement field of a sample, and ϕ_{gt} is the corresponding ground truth.

In Silico A synthetic dataset is generated using the pipeline described in “Synthetic data generation” with the same boundary conditions, and the target shape variants are unseen to the neural network during training. For each target shape, we estimated the displacement field for the identical template point cloud using C2P-Net trained on 20,000 samples. The mean target displacement field for the synthetic

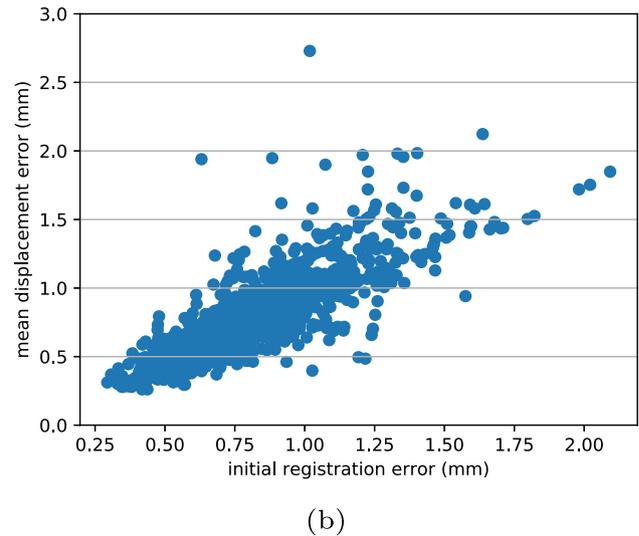
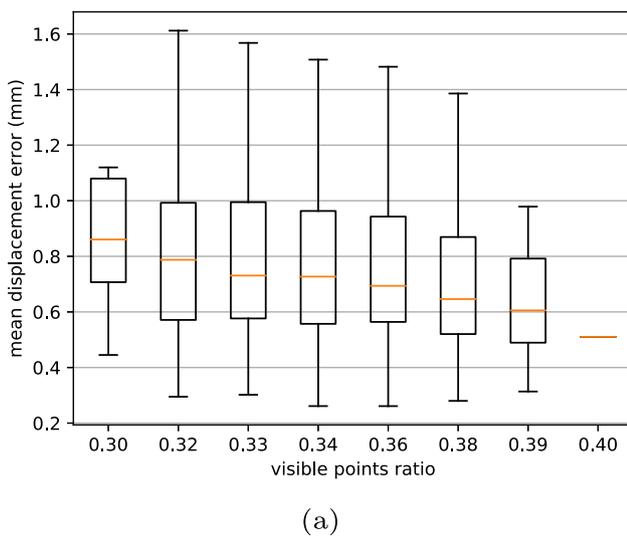


Fig. 3 **a** shows the relation between the MDE of synthetic samples and the visible points ratio, which is calculated as the ratio between the point number of a target point cloud to the corresponding ground truth point cloud: $|P_{inv}|/|T|$. The decreasing trend shows that with more points available in the target in vivo point clouds, the neural network tends

to predict better deformation. **b** Depicts the MDE to the initial rigid registration error which is produced by the NgeNet. The better the initial rigid registration and correspondence set predicted by NgeNet, the lower the global non-rigid registration error from NDP becomes

dataset is 1.54 mm. Figure 3a illustrates that the neural network tends to predict the deformation of the ex vivo point cloud with higher accuracy if it obtains more information about the target in vivo OCT point cloud. With few exceptions, Fig. 3b shows when the initial registration error made by NgeNet is low, the final global non-rigid displacement field from NDP tends to be small.

In Vivo A real OCT dataset with 9 samples of the human middle ear was collected with a handheld OCT imaging system [5] and segmented manually by surgeons within 3D Slicer along with a set of sparse landmarks for each structure of a sample (see Fig. 1). Since there is no ground truth correspondence available for the source and target point clouds to calculate M_{MDE} , we introduce another metric based on the sparse landmarks:

$$M_L = \frac{1}{|L_{exv}|} \sum_{l_{i,k} \in L_{exv}} \min_{l_{j,k} \in L_{inv}} \|l_{i,k} - l_{j,k}\|_{k=1,2,\dots,K} \quad (6)$$

where L_{exv} are the landmarks on the ex vivo point cloud from μ CT, which is also marked manually, L_{inv} are the corresponding landmarks on the target in vivo point cloud, and the landmark points belong to K different segmentations.

Table 1 itemizes the registration results of C2P-Net and baseline models on both datasets. We can observe our registration pipeline outperforms the other methods in terms of M_{MDE} and M_L . However, baseline models tend to have

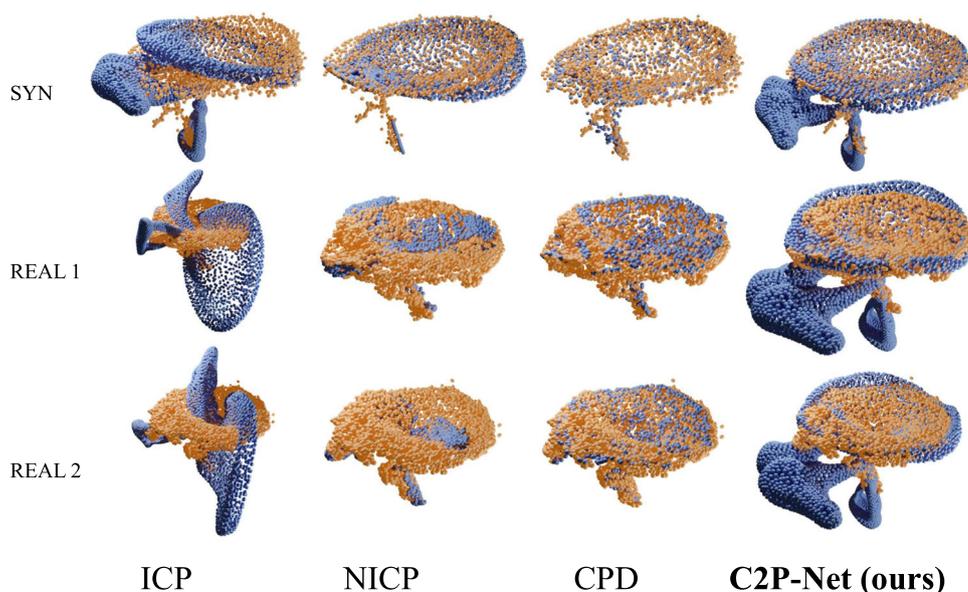
Table 1 Experiments results on synthetic data (SYN) and real OCT data (REAL), including mean displacement error (M_{MDE})[mm], landmarks error (M_L)[mm] and Chamfer distance (M_{CD})[mm]

	SYN		REAL		
	$M_{MDE} \downarrow$	$M_{CD} \dagger$	$M_L \downarrow$	$M_L \downarrow$	$M_{CD} \dagger$
ICP*	2.01	1.85	1.15	4.74	3.03
NICP	4.61	0.22	0.87	4.39	0.68
CPD	8.79	1.07	0.96	2.17	0.059
C2P-Net (ours)	0.781	0.582	0.424	0.808	2.43

The bold numbers indicate the best-performing methods on each metric
 * rigid registration method(s)
 † baseline methods tend to fall into local minima and ignore the geometry, though the CD is lower

smaller M_{CD} , which means the source point clouds are deformed largely to fit the target regardless of the anatomical geometry of the middle ear. Furthermore, this issue is demonstrated visually in Fig. 4, which depicts several registration results from C2P-Net and baseline models on real OCT samples. Hence, it is manifested that C2P-Net is able to register the partial target point cloud without losing the underlying geometry of source point clouds, which is the critical information clinicians want to obtain with registration. On the contrary, the baseline models only focus on the spatial position of local points regardless of the neighboring and global geometry of the middle ear template.

Fig. 4 Registration results of baseline models and C2P-Net on synthetic (SYN) and real OCT samples (REAL). Blue point clouds are the deformed ex vivo ear models, i.e., the shape template, calculated by all the investigated methods. The orange point clouds stand for the registration target, i.e., in vivo point clouds. From the visualization, we can see baseline non-rigid methods, NICP and CPD, tend to squeeze the points from the ex vivo template largely to match the target point clouds, since they only focus on the position of local points, while C2P-Net retains the global geometry of the middle ear due to the learned correspondence knowledge between the ex and in vivo point clouds



Conclusion

In this work, we propose to improve the interpretation of OCT data for surgical diagnosis by fusing the knowledge from ex vivo and general μ CT data with the in vivo noisy and incomplete OCT data. To this end, we propose C2P-Net which is based on NgeNet and Neural Deformation Pyramid. It takes the ex and in vivo 3D point clouds as input and explores the sparse correspondence between the two, then aligns them in a non-rigid fashion. Due to the lack of training data, a fast and effective synthetic simulation pipeline was designed using Blender3D, which produces noisy and partial point clouds as seen in vivo based on the randomly generated shape variants of the middle ear. Our neural network was trained based on 20,000 synthetic samples, and it took around 3.5 s to predict the displacement field for a given in vivo point cloud at inference time.

To assess the performance of C2P-Net, experiments on synthetic and real OCT datasets were carried out. Since the real OCT data are noisy and incomplete, we investigated various metrics including mean displacement error, landmark error, and Chamfer distance and compared C2P-Net to the other baseline models from both statistics and geometry aspects. From Table 1 and Fig. 4, we can see that C2P-Net is able to deal with the partial OCT data and retain the anatomical structure during non-rigid registration, while other non-rigid methods do not understand the global geometry of the point clouds and focus only on local points.

Future work will focus on improving registration accuracy. This can be achieved by improving the simulation pipeline with realistic noise to further bridge the gap between synthetic and real data. Furthermore, the segmentation infor-

mation can be employed by the network to improve the prediction of correspondences. Apart from this, inference time can be further decreased by exploring the optimal configuration of the network for the OCT samples.

Acknowledgements Funded by the Else Kröner Fresenius Centre for Digital Health.

Funding Open Access funding enabled and organized by Projekt DEAL.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Zwislocki J (1982) Normal function of the middle ear and its measurement. *Audiology* 21(1):4–14
- Won J, Monroy GL, Dsouza RI, Spillman DR, McJunkin J, Porter RG, Shi J, Aksamitiene E, Sherwood M, Stiger L, Boppart SA (2021) Handheld briefcase optical coherence tomography with real-time machine learning classifier for middle ear infections. *Biosensors* 11(5):143–156
- Won J, Porter RG, Novak MA, Youakim J, Sum A, Barkalifa R, Aksamitiene E, Zhang A, Nolan R, Shelton R, Boppart SA (2021) In vivo dynamic characterization of the human tympanic membrane

- using pneumatic optical coherence tomography. *J Biophotonics* 14(4):1–12
4. Snow JB, Wackym PA, Ballenger JJ (2009) *Ballenger's otorhinolaryngology: head and neck surgery*. BC Decker, Hamilton, ON
 5. Morgenstern J, Schindler M, Kirsten L, Golde J, Bornitz M, Kemper M, Koch E, Zahnert T, Neudert M (2020) Endoscopic optical coherence tomography for evaluation of success of tympanoplasty. *Otol Neurotol* 41(7):901–905
 6. Monroy GL, Won J, Dsouza R, Pande P, Hill MC, Porter RG, Novak MA, Spillman DR, Boppart SA (2019) Automated classification platform for the identification of otitis media using optical coherence tomography. *NPJ Dig Med* 2(1):1–22
 7. Monroy GL, Shelton RL, Nolan RM, Nguyen CT, Novak MA, Hill MC, McCormick DT, Boppart SA (2015) Noninvasive depth-resolved optical measurements of the tympanic membrane and middle ear for differentiating otitis media. *Laryngoscope* 125(8):276–282
 8. Brock A, Lim T, Ritchie JM, Weston NJ (2016) Generative and discriminative voxel modeling with convolutional neural networks. In: *Neural information processing conference: 3D deep learning*
 9. Su H, Maji S, Kalogerakis E, Learned-Miller E (2015) Multi-view convolutional neural networks for 3d shape recognition. In: *Proceedings of the IEEE international conference on computer vision*, p 945–953
 10. Feng Y, Zhang Z, Zhao X, Ji R, Gao Y (2018) Gvcnn: group-view convolutional neural networks for 3d shape recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, p 264–272
 11. Qi CR, Su H, Nießner M, Dai A, Yan M, Guibas LJ (2016) Volumetric and multi-view cnns for object classification on 3d data. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, p 5648–5656
 12. Qi CR, Su H, Mo K, Guibas LJ (2017) Pointnet: deep learning on point sets for 3d classification and segmentation. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, p 652–660
 13. Qi CR, Yi L, Su H, Guibas LJ (2017) Pointnet++: deep hierarchical feature learning on point sets in a metric space. In: *Advances in neural information processing systems*, vol 30
 14. Thomas H, Qi CR, Deschaud J-E, Marcotegui B, Goulette F, Guibas LJ (2019) Kpconv: flexible and deformable convolution for point clouds. In: *Proceedings of the IEEE/CVF international conference on computer vision*, p 6411–6420
 15. Xu Y, Fan T, Xu M, Zeng L, Qiao Y (2018) Spidernn: deep learning on point sets with parameterized convolutional filters. In: *Proceedings of the European conference on computer vision (ECCV)*, p 87–102
 16. Choy C, Park J, Koltun V (2019) Fully convolutional geometric features. In: *Proceedings of the IEEE/CVF international conference on computer vision*, p 8958–8966
 17. Besl PJ, McKay ND (1992) Method for registration of 3-d shapes. *Sensor fusion IV: control paradigms and data structures*, vol 1611. Spie, Bellingham, pp 586–606
 18. Amberg B, Romdhani S, Vetter T (2007) Optimal step nonrigid icp algorithms for surface registration. In: *2007 IEEE conference on computer vision and pattern recognition*, IEEE, p 1–8
 19. Aoki Y, Goforth H, Srivatsan RA, Lucey S (2019) Pointnetlk: robust & efficient point cloud registration using pointnet. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, p 7163–7172
 20. Yew ZJ, Lee GH (2022) Regtr: end-to-end point cloud correspondences with transformers. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, p 6677–6686
 21. Zhu L, Guan H, Lin C, Han R (2022) Neighborhood-aware geometric encoding network for point cloud registration. *arXiv preprint arXiv:2201.12094*
 22. Li Y, Harada T (2022) Non-rigid point cloud registration with neural deformation pyramid. *arXiv preprint arXiv:2205.12796*
 23. Lucas BD, Kanade T (1981) An iterative image registration technique with an application to stereo vision. In: *IJCAI'81: 7th international joint conference on artificial intelligence*, p 674–679

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.