INAUGURAL-DISSERTATION

zur

Erlangung der Doktorwürde

der

Gesamtfakultät für Mathematik, Ingenieur- und Naturwissenschaften

der

Ruprecht–Karls–Universität

Heidelberg

vorgelegt von

Sara Monji-Azad, M.Sc.

aus Rasht, Guilan, Iran

Tag der mündlichen Prüfung:

Coarse-to-Fine Learning Frameworks for Non-Rigid 3D Point Cloud Registration under Large Deformations

Betreuer: Prof. Dr. Jürgen Hesser

Abstract

Non-rigid point cloud registration is a crucial task for aligning 3D data when objects undergo deformation due to motion, pressure, or biological processes. This is especially important in high-stakes domains such as surgical navigation, where anatomical structures often bend, compress, or stretch in unpredictable ways. Despite recent progress, current methods continue to struggle with large deformations, noisy or partial observations, and generalization to real-world scenarios. Moreover, many approaches fail to integrate both local and global spatial learning or to account for uncertainty in ambiguous regions.

This thesis introduces a multi-stage framework for non-rigid point cloud registration, comprising three progressively refined learning models. First, Robust-DefReg encodes local geometric structures using graph convolutions to build deformation-aware descriptors. Next, DefTransNet incorporates global learning through a hybrid Transformer–Graph architecture that explicitly resolves feature ambiguity via cross-attention between source and target point clouds. Finally, Learning-to-Refine introduces a probabilistic iterative refinement strategy that regularizes deformation prediction using KL divergence over learned feature distributions, addressing uncertainty in ambiguous and partially observed regions. Together, these models directly respond to the core research questions on local representation, global context, and uncertainty modeling posed in this dissertation.

To enable reproducible benchmarking across diverse deformation levels, two datasets were developed: SynBench, a synthetic dataset with controlled and progressively increasing deformation levels; and DeformedTissue, a real-world dataset based on simulated anatomical tissue deformation. Additionally, all methods were evaluated on two widely used public benchmarks, ModelNet40 and 4DMatch, to validate generalization across domains. Experimental results reveal a clear progression in performance across the proposed methods. DefTransNet outperforms state-of-the-art baselines by achieving high accuracy and stability under severe deformations, while Learning-to-Refine introduces probabilistic refinement that further improves convergence and consistency. Evaluation across synthetic, real-world, and public datasets confirms the generalizability of the framework. Notably, the deformation–robustness plots indicate that the performance of our proposed methods remains stable even under extreme deformation levels, suggesting that, within the scope of this study, the core challenge of deformation in non-rigid point cloud registration has been effectively addressed.

Zusammenfassung

Die nicht-starre Punktwolkenregistrierung ist eine entscheidende Aufgabe für die Ausrichtung von 3D-Daten, wenn Objekte aufgrund von Bewegung, Druck oder biologischen Prozessen Verformungen erfahren. Dies ist besonders wichtig in Bereichen mit hohem Risiko, wie beispielsweise der chirurgischen Navigation, wo anatomische Strukturen oft auf unvorhersehbare Weise gebogen, komprimiert oder gedehnt werden. Trotz jüngster Fortschritte haben aktuelle Methoden weiterhin Schwierigkeiten mit großen Verformungen, verrauschten oder unvollständigen Beobachtungen und der Generalisierung auf reale Szenarien. Darüber hinaus versäumen es viele Ansätze, sowohl lokales als auch globales räumliches Lernen zu integrieren oder Unsicherheiten in mehrdeutigen Bereichen zu berücksichtigen.

Diese Arbeit stellt ein mehrstufiges Framework für die nicht-starre Punktwolkenregistrierung vor, das drei zunehmend verfeinerte Lernmodelle umfasst. Zunächst codiert Robust-DefReg lokale geometrische Strukturen mithilfe von Graph-Faltungen, um verformungsbewusste Deskriptoren zu erstellen. Als Nächstes integriert DefTransNet globales Lernen durch eine hybride Transformer-Graph-Architektur, die Mehrdeutigkeiten durch Cross-Attention zwischen Quell- und Zielpunktwolken explizit auflöst. Schließlich führt Learning-to-Refine eine probabilistische iterative Verfeinerungsstrategie ein, die die Verformungsvorhersage mithilfe der KL-Divergenz über gelernte Merkmalsverteilungen reguliert und so Unsicherheiten in mehrdeutigen und nur teilweise beobachteten Bereichen berücksichtigt. Zusammen beantworten diese Modelle direkt die zentralen Forschungsfragen zu lokaler Repräsentation, globalem Kontext und Unsicherheitsmodellierung, die in dieser Dissertation gestellt werden.

Um ein reproduzierbares Benchmarking über verschiedene Verformungsgrade hinweg zu ermöglichen, wurden zwei Datensätze entwickelt: SynBench, ein synthetischer Datensatz mit kontrollierten und progressiv ansteigenden Verformungsgraden, und DeformedTissue, ein realer Datensatz, der auf simulierten anatomischen Gewebeverformungen basiert. Zusätzlich wurden alle Methoden anhand von zwei weit verbreiteten öffentlichen Benchmarks, ModelNet40 und 4DMatch, evaluate, um die Generalisierbarkeit über verschiedene Domänen hinweg zu validieren. Die experimentellen Ergebnisse zeigen eine deutliche Leistungssteigerung. DefTransNet übertrifft den aktuellen Stand der Technik durch hohe Genauigkeit und Stabilität unter starken Verformungen, während Learning-to-Refine eine probabilistische Verfeinerung einführt, die die Konvergenz und Konsistenz weiter verbessert. Die Bewertung anhand synthetischer, realer und öffentlicher Datensätze bestätigt die Verallgemeinerbarkeit des Frameworks. Insbesondere zeigen die Verformungs-Robustheits-Diagramme, dass die Leistung unserer vorgeschlagenen Methoden auch unter extremen Verformungsgraden stabil bleibt, was darauf hindeutet, dass im Rahmen dieser Studie die zentrale Herausforderung effektiv gelöst wurde.

Acknowledgment

First and foremost, I would like to express my deepest gratitude to Prof. Dr. Jürgen Hesser for giving me the opportunity to pursue my PhD under his supervision. I am especially thankful for his continuous support, guidance, and encouragement to always push beyond my limits.

I am also sincerely grateful to Prof. Dr. med. Claudia Scherl and Dr. med. David Männle for providing the medical data and datasets that were essential for this work.

To my colleagues at the Mannheim Institute for Intelligent Systems in Medicine (MIISM), thank you for providing such a friendly and supportive environment. Dr. Nikolas Löw, I truly appreciate your help and guidance at the beginning of this journey. Dr. Katharina Jerg, thank you for your thoughtful reviews and comments on my thesis and papers, and for your constant support and cherished friendship. Marlen Runz, thank you for your kind presence throughout both the good and difficult days, in and beyond the office. Dr. Tobias Meißner, your feedback on my thesis and papers has been extremely valuable. Thank you for your time and support. Marvin Kinz, thank you for your valuable contributions during your Bachelor's and Master's projects, which supported the development of this dissertation. I am also grateful to all my students, whose teamwork and contributions have shaped many parts of this thesis.

A heartfelt thank you to Astrid D'Alessandro and Lei Zheng. You were always there to help me with official and technical matters, making everything feel easier whenever I turned to you.

To my beloved parents: Your lifelong dedication, unconditional love, and unwavering support have made me who I am today. Thank you for encouraging me to follow my own path and for the sacrifices you made to support my future, even when it meant letting your child live far away. To my dear sisters, thank you for your constant love and encouragement. And finally, Christian, thank you for your valuable feedback on this thesis, but most importantly, thank you for being by my side as the right person at the right time.

„Die Entdeckung der Wahrheit wird wirksamer verhindert, nicht durch die falsche Erscheinung der Dinge, die zur Irrung verleitet, nicht direkt durch die Schwäche des Verstandes, sondern durch vorgefaßte Meinung, durch Vorurteil.“

Attributed to Arthur Schopenhauer, *Über die vierfache Wurzel des Satzes vom zureichenden Grunde*, Dissertation, 1813

Contents

List of Figures

List of Tables

Abbreviations

3D	Three-Dimensional
CD	Chamfer Distance
CNN	Convolutional Neural Network
CPD	Coherent Point Drift
DGCNN	Dynamic Graph Convolutional Neural Network
EMD	Earth Mover's Distance
FEM	Finite Element Method
GAN	Generative Adversarial Network
GCNN	Graph Convolutional Neural Network
GMM	Gaussian Mixture Model
HD	Hausdorff Distance
ICP	Iterative Closest Point
IR	Inlier Ratio
KL	Kullback–Leibler
LBP	Loopy Belief Propagation
LiDAR	Light Detection and Ranging
MLP	Multilayer Perceptron
MSE	Mean Squared Error
PCR	Point Cloud Registration
RANSAC	Random Sample Consensus
RMSE	Root Mean Squared Error
RRE	Relative Rotation Error
RTE	Relative Translation Error
T-Net	Spatial Transformation Network
TPS	Thin Plate Spline
VAE	Variational Autoencoder

Chapter 1

Introduction

Understanding and modeling the physical world is a cornerstone of progress across science, engineering, and medicine. From designing intelligent robots that navigate dynamic environments to guiding surgeons through constantly changing anatomy during operations, the ability to accurately interpret and align 3D information is central to reliable decision-making [2]. Yet, many real-world scenarios involve objects that are not static but deform; they bend, stretch, compress, or shift over time. Capturing and learning about these deformations is essential for making technologies safer, more adaptive, and more context-aware in high-stakes applications [3].

In many real-world applications, scenarios are encountered in which objects undergo complex deformations. For example, during surgery, soft tissues are deformed due to manipulation or the insertion of instruments [4]. Similarly, in industrial settings, flexible materials or manufactured parts may be bent or warped under pressure. In such cases, it becomes essential that 3D data representing these deformable objects be tracked and aligned across time or under varying acquisition conditions, a task referred to as non-rigid point cloud registration (PCR) [5, 6].

PCR is a fundamental task in computer vision, aimed at estimating spatial transformations that align corresponding points across two or more point clouds [7]. Accurate registration underpins a wide range of applications, including 3D reconstruction [8], autonomous driving [9], augmented and virtual reality [10], LiDAR-based mapping [11], robotic perception, and quality control in manufacturing [12]. In medical imaging and surgical interventions, PCR is particularly critical for aligning anatomical structures captured at different time points or under different conditions (e.g., before and after tissue deformation) [13].

The goal of registration is to minimize geometric differences between the source and target point clouds, thereby estimating the underlying transformation that aligns them. While rigid registration assumes that the structure of objects remains constant (translation and rotation only), non-rigid registration is significantly more complex as it must account for de-

formable objects whose shape changes due to motion, pressure, or biological processes [14]. This is especially important in soft tissue applications, where anatomical variations and intraoperative deformations are inevitable [15].

1.1 Motivation

To effectively model and register deformations occurring in the physical world, such as those seen in surgical procedures or flexible manufacturing, accurate 3D surface capture is essential. Among the available representations, point clouds offer a particularly compelling choice. Unlike volumetric grids or surface meshes, which require uniform resolution or pre-defined connectivity, point clouds provide a direct and flexible means of representing real-world surfaces as collections of discrete 3D samples. Their ability to natively handle sparse, incomplete, and irregular data makes them well-suited for capturing deformable objects in real-time, sensor-driven settings. This suitability is especially critical in high-stakes environments like the operating room, where surfaces deform unpredictably and data acquisition may be constrained by time, occlusions, or motion.

Given these advantages, point cloud-based methods have become increasingly prominent for capturing and aligning deformable surfaces across time or varying acquisition conditions. In particular, non-rigid PCR has emerged as a central task in this context, aiming to estimate spatially varying transformations that align a source point cloud to a target one. Unlike rigid registration, which assumes a single global transformation, non-rigid registration must account for localized, nonlinear displacements arising from bending, stretching, or compression. These challenges become especially severe under large deformations, where significant geometric and topological discrepancies exist between the point clouds.

Traditional optimization-based methods such as Coherent Point Drift (CPD) [16] and Thin Plate Splines (TPS) [17] have long been used for non-rigid PCR. However, they typically rely on smoothness assumptions and good initial alignment, limiting their performance under noise, occlusions, or large displacements [18]. In response, learning-based approaches have gained traction, using data-driven models to estimate correspondences or deformation fields. Convolutional Neural Networks (CNN) models [19] operate on voxelized or projected views but suffer from discretization artifacts. Graph CNN (GCNNs) [20] preserve local geometric relations but struggle with long-range interactions. Transformer-based architectures [21, 22] leverage attention mechanisms to model global context but can be computationally expensive and less robust to incomplete data. Despite these advances, several critical challenges remain unresolved in the field of non-rigid PCR:

- ***Uncertainty modeling is often neglected.*** Most learning-based methods generate deterministic displacement fields without explicitly modeling uncertainty arising from

ambiguous input data or regions with missing correspondences. This lack of uncertainty awareness can lead to overconfident predictions and inaccurate estimates, particularly in ill-posed areas of the point cloud.

- ***Robustness to real-world conditions is limited.*** Many approaches perform well on clean synthetic benchmarks but degrade significantly in the presence of noise, outliers, partial observations, or low-resolution scans, conditions that are common in real-world medical and robotics applications.

- ***Feature ambiguity remains unresolved.*** In deformable objects, different regions may exhibit locally similar geometry (e.g., symmetric or flat regions), leading to incorrect matches. Without a strong global context or cross-cloud learning, networks often fail to distinguish between these similar structures.

- ***Lack of generalization and data dependency.*** Many deep learning models require large, labeled datasets for training, which are difficult to obtain in domains such as surgical navigation or biomedical imaging. In addition, models trained on one dataset often fail to generalize to different anatomical structures or deformation types.

- ***Insufficient evaluation protocols.*** A number of methods are not evaluated under comprehensive or standardized benchmarks. This limits fair comparison and makes it difficult to assess robustness across diverse deformation scales, noise levels, or sensor artifacts.

- ***Single-pass pipelines dominate.*** Many models rely on one-shot inference and do not incorporate feedback mechanisms or iterative refinement to correct residual errors. As a result, small misalignments accumulate and reduce the final registration quality.

Together, these limitations highlight the need for non-rigid PCR frameworks that not only extract meaningful geometric features but also reason across scales, handle noisy and partial data, and explicitly model prediction uncertainty. Addressing these challenges is essential for reliable deployment in high-stakes domains such as surgical assistance, soft tissue analysis, and autonomous robotics.

1.2 Research Questions and Hypotheses

To address the gaps outlined in the previous section, particularly the lack of realistic datasets, limited model generalizability, and the absence of robust refinement strategies, this dissertation introduces a comprehensive framework for non-rigid PCR under large and complex

deformations. The research is structured around four key questions that span data generation, model architecture, and optimization strategies:

- ***RQ1: Dataset design and benchmarking.*** How can synthetic datasets be designed to realistically simulate soft-body deformation, noise, and partiality to reduce the simulation-to-reality gap in evaluating non-rigid PCR methods?
- ***RQ2: Local geometric representation.*** How can local relationships between points in a point cloud be efficiently leveraged for PCR to achieve improved accuracy and robustness?
- ***RQ3: Resolving feature ambiguity.*** How can the feature ambiguity problem arising from long-range dependencies in geometrically complex point clouds be efficiently addressed?
- ***RQ4: Uncertainty modeling and probabilistic refinement.*** How can the deformation vector field be effectively represented and utilized during the PCR process to enhance convergence and robustness under uncertainty?

In response to these research questions, the following hypotheses are formulated:

- ***H1:*** Designing synthetic datasets using physics-based simulation with controlled deformation parameters and ground-truth correspondences enables realistic modeling of soft-body behavior and reduces the simulation-to-reality gap.
- ***H2:*** Encoding local geometric structures via graph-based models improves robustness to deformation and noise compared to pointwise or voxel-based approaches.
- ***H3:*** Modeling global interactions with transformer-based attention reduces feature ambiguity and improves registration accuracy in complex geometries.
- ***H4:*** A probabilistic prior over deformation fields, integrated into an iterative refinement process, enhances convergence stability and robustness under uncertainty.

To address the formulated research questions and validate the corresponding hypotheses, this dissertation introduces contributions along three core dimensions: Simulation-based data generation, deep learning model development, and systematic performance evaluation. The main contributions are summarized as follows:

- ***A simulation toolkit for realistic deformation synthesis.*** A simulation framework, SimTool [23], is developed to generate synthetic point clouds of soft-body objects undergoing controlled and reproducible deformations. This tool enables fine-grained

manipulation of deformation type and magnitude, while preserving ground truth correspondences, facilitating the creation of synthetic data that closely mirrors real-world conditions and supporting rigorous benchmarking of non-rigid PCR algorithms.

- ***Two benchmark datasets bridging simulation and reality.*** To comprehensively evaluate non-rigid PCR under diverse and realistic conditions, two benchmark datasets are introduced: The synthetic dataset SynBench [24], generated using SimTool, and the real-world dataset DeformedTissue [25, 26], derived from actual soft tissue deformations. Both datasets provide the ground truth correspondences and span a wide range of deformation levels, noise and outlier levels, and partial data, enabling robust and generalizable evaluation protocols.

- ***Three deep learning architectures for non-rigid PCR.*** Three progressively refined deep learning models are proposed to tackle the key challenges identified in this dissertation:

  – Robust-DefReg [27], a graph-based model that captures local geometric context via graph convolutions, providing resilience to moderate deformations and spatial noise.

  – DefTransNet [22], a transformer-based model that leverages global cross-attention to resolve feature ambiguity in repetitive or geometrically similar regions, especially under severe non-local deformations.

  – Learning-to-Refine, an iterative refinement model that incorporates uncertainty modeling through a KL divergence-based probabilistic loss, enabling progressive correction of ambiguous correspondences and improved convergence stability.

The papers resulting from the aforementioned contributions have been published in peer-reviewed journals, and the corresponding code is publicly available. A list of the published papers based on this dissertation is provided in Appendix A.

1.3 Structure of the Thesis

The remainder of this thesis is structured as follows. Chapter 2 presents the theoretical foundations and related work in the field of non-rigid PCR. It introduces essential concepts such as rigid versus non-rigid alignment, feature-based versus robustness-oriented methods, and commonly used loss functions and optimization strategies. Furthermore, it

provides a comprehensive review of recent learning-based approaches, categorized by architectural paradigms, like CNNs, GCNNs, and Transformers. Chapter 3 details the proposed methodology, including the simulation framework, the SynBench and DeformedTissue datasets, and the three non-rigid PCR models: Robust-DefReg, DefTransNet, and the iterative Learning-to-Refine strategy. Chapter 4 presents quantitative and qualitative evaluations of the proposed methods under varying conditions such as deformation scale, noise, and partial data, using standard metrics and comparing performance against baseline and state-of-the-art methods. Chapter 5 discusses the findings in light of the research questions and hypotheses, analyzes architectural contributions, and reflects on limitations and failure modes. Finally, Chapter 6 concludes the thesis by summarizing key contributions and outlining future research directions.

Chapter 2

Theoretical Foundations and Related Work

Point cloud registration (PCR) involves estimating a spatial transformation that aligns two or more 3D point sets into a common coordinate frame. While the core goal is geometric alignment, the problem is inherently complex due to challenges such as noise, outliers, occlusions, varying point densities, and, in the non-rigid case, spatially varying deformations. Over the years, a wide range of methodologies have been proposed to tackle different aspects of this task, evolving from traditional geometric techniques to modern learning-based solutions.

This chapter provides a comprehensive overview of the theoretical foundations and existing approaches to PCR. Section 2.1 formally defines the PCR problem. Section 2.2 introduces the main methodological dimensions, including rigid and non-rigid paradigms, coarse-to-fine pipelines, feature engineering, robustness mechanisms, search space formulations, loss functions, and optimization strategies. Section 2.3 summarizes standard evaluation metrics for assessing registration accuracy and robustness. Finally, Section 2.4 reviews representative learning-based approaches for non-rigid PCR, categorized by architectural families such as CNNs, GCNNs, and Transformers.

2.1 Problem Definition of Point Cloud Registration

Let $\mathbf{X} = \{\mathbf{x}_i \in \mathbb{R}^n \mid i = 1, \ldots, N\}$ and $\mathbf{Y} = \{\mathbf{y}_j \in \mathbb{R}^n \mid j = 1, \ldots, M\}$ be two finite point sets, referred to as the source and target point clouds, respectively. These point sets are defined in a metric space $(\mathbb{R}^n, d)$, where $d : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}_{\geq 0}$ is a distance function satisfying the properties of a metric (non-negativity, identity of indiscernibles, symmetry, and triangle inequality). A common choice for d is the Euclidean distance.

We define a point-set discrepancy as a function $D : \mathcal{P}(\mathbb{R}^n) \times \mathcal{P}(\mathbb{R}^n) \to \mathbb{R}_{\geq 0}$, where

$\mathcal{P}(\mathbb{R}^n)$ denotes the set of finite point clouds in $\mathbb{R}^n$. The discrepancy $D(\mathbf{X}, \mathbf{Y})$ quantifies the dissimilarity between $\mathbf{X}$ and $\mathbf{Y}$ based on the pointwise metric d. For instance, Chamfer distance is computed as a sum or average of nearest-neighbor distances using d, while Earth Mover's Distance (EMD or Wasserstein-1) and Sinkhorn distance define optimal transport costs between point sets or distributions.

The task of PCR is to find a transformation $\mathcal{T}_{\boldsymbol{\theta}} : \mathbb{R}^n \to \mathbb{R}^n$, parameterized by $\boldsymbol{\theta} \in \Theta$, that aligns the source point cloud $\mathbf{X}$ with the target point cloud $\mathbf{Y}$ by minimizing a discrepancy measure between them. The transformed source point cloud is denoted $\mathcal{T}_{\boldsymbol{\theta}}(\mathbf{X}) = \{\mathcal{T}_{\boldsymbol{\theta}}(\mathbf{x}_i) \mid i = 1, \ldots, N\}$. The registration problem is formulated as

$$\min_{\boldsymbol{\theta} \in \Theta} D\left(\mathcal{T}_{\boldsymbol{\theta}}(\mathbf{X}), \mathbf{Y}\right), \tag{2.1}$$

where

- $\mathcal{T}_{\boldsymbol{\theta}}$ is a transformation function chosen from a parameterized family $\{\mathcal{T}_{\boldsymbol{\theta}} \mid \boldsymbol{\theta} \in \Theta\}$,
- $D(\cdot, \cdot)$ is a discrepancy measure or distance function between two finite point sets, such as Chamfer Distance, EMD, or a learned alignment loss,
- $\boldsymbol{\theta}$ denotes the parameters of the transformation (e.g., neural network weights or rigid-body parameters).

Rigid point cloud registration. In the special case of rigid registration, the transformation $\mathcal{T}$ is constrained to consist of a rotation and translation:

$$\mathcal{T}(\mathbf{x}_i) = \mathbf{R}\mathbf{x}_i + \mathbf{t}, \tag{2.2}$$

where

- $\mathbf{R} \in SO(n)$ is a rotation matrix such that $\mathbf{R}^\top \mathbf{R} = \mathbf{I}$ and $\det(\mathbf{R}) = 1$,
- $\mathbf{t} \in \mathbb{R}^n$ is a translation vector,
- $\mathcal{T} \in SE(n)$ is a rigid-body transformation.

Rigid registration is then formulated as the following constrained optimization problem:

$$\min_{\mathbf{R} \in SO(n),\, \mathbf{t} \in \mathbb{R}^n} \sum_{i=1}^{N} d(\mathbf{R}\mathbf{x}_i + \mathbf{t}, \mathbf{y}_{j(i)}), \tag{2.3}$$

where $\mathbf{y}_{j(i)} \in \mathbf{Y}$ denotes the corresponding point to $\mathbf{x}_i$, either known a priori or estimated via correspondence algorithms (e.g., nearest neighbors or soft matching). Note that while the squared distance is often used in practice under the Euclidean metric for optimization convenience, it is not required in the general formulation, where d is any valid metric.

Non-rigid point cloud registration. In the more general case of non-rigid registration, the transformation $\mathcal{T}_{\boldsymbol{\theta}}$ is allowed to deform the geometry in a spatially varying, potentially nonlinear manner. To motivate the formulation of the objective function, we adopt a probabilistic view of registration.

Assume we model the deformation process using a posterior distribution over the transformation given the observed target point cloud $\mathbf{Y}$, denoted as $p(\mathcal{T}_{\boldsymbol{\theta}}(\mathbf{x}) \mid \mathbf{Y})$. The goal is to estimate the most probable transformation under this posterior, which can be expressed as a maximum a posteriori (MAP) problem:

$$\boldsymbol{\theta}^* = \arg\max_{\boldsymbol{\theta}} p(\mathcal{T}_{\boldsymbol{\theta}}(\mathbf{x}) \mid \mathbf{Y}). \tag{2.4}$$

To make this tractable, we take the negative logarithm and convert the maximization into a minimization:

$$\boldsymbol{\theta}^* = \arg\min_{\boldsymbol{\theta}} -\log p(\mathcal{T}_{\boldsymbol{\theta}}(\mathbf{x}) \mid \mathbf{Y}). \tag{2.5}$$

Using Bayes' theorem, the log-posterior can be decomposed into a log-likelihood and a log-prior:

$$\log p(\mathcal{T}_{\boldsymbol{\theta}}(\mathbf{x}) \mid \mathbf{Y}) = \log p(\mathbf{Y} \mid \mathcal{T}_{\boldsymbol{\theta}}(\mathbf{x})) + \log p(\mathcal{T}_{\boldsymbol{\theta}}(\mathbf{x})). \tag{2.6}$$

Substituting Equation 2.6 into Equation 2.5, the optimization objective becomes:

$$\min_{\boldsymbol{\theta}\in\Theta} -\log p(\mathbf{Y} \mid \mathcal{T}_{\boldsymbol{\theta}}(\mathbf{x})) - \log p(\mathcal{T}_{\boldsymbol{\theta}}(\mathbf{x})). \tag{2.7}$$

We now identify the components of Equation 2.7 with the terms in our non-rigid registration formulation:

- The negative log-likelihood term, $-\log p(\mathbf{Y} \mid \mathcal{T}_{\boldsymbol{\theta}}(\mathbf{x}))$, corresponds to a discrepancy measure $D(\mathcal{T}_{\boldsymbol{\theta}}(\mathbf{X}), \mathbf{Y})$, which quantifies alignment quality.
- The negative log-prior term, $-\log p(\mathcal{T}_{\boldsymbol{\theta}}(\mathbf{x}))$, is modeled as a regularization functional $\mathcal{R}(\boldsymbol{\theta})$, penalizing implausible or overly complex deformations.

Thus, the resulting optimization objective can be expressed as:

$$\min_{\boldsymbol{\theta}\in\Theta} D\left(\mathcal{T}_{\boldsymbol{\theta}}(\mathbf{X}), \mathbf{Y}\right) + \lambda \mathcal{R}(\boldsymbol{\theta}), \tag{2.8}$$

where

- $\mathcal{R}(\boldsymbol{\theta})$ encodes prior assumptions over the deformation field (e.g., smoothness or statistical regularity),

- $\lambda \in \mathbb{R}_{\geq 0}$ controls the trade-off between alignment quality and regularization strength.

This probabilistic formulation makes it explicit that the discrepancy term models data likelihood, accounting for deformation, noise, and artifacts, while the regularization term reflects prior knowledge about the deformation space.

In this thesis, we focus on the case $n = 3$, where point clouds are embedded in 3D Euclidean space $\mathbb{R}^3$. However, the formulation is general and can be extended to other metric spaces and higher-dimensional embeddings as required. Similarly, the underlying pointwise metric d can be adapted from the Euclidean norm to alternatives such as Wasserstein or Sinkhorn distances. In learning-based settings, the transformation $\mathcal{T}_{\theta}$ is typically implemented as a neural network that predicts a deformation field to align $\mathbf{X}$ with $\mathbf{Y}$.

2.2 Overview of Various Methodologies for Point Cloud Registration

Existing registration approaches can be broadly classified into distinct yet interconnected categories, such as rigid vs. non-rigid registration, coarse vs. fine registration, feature-based methods, robustness-oriented techniques, search-space strategies, loss-function-driven approaches, and various optimization techniques [6]. This section provides an overview of the fundamental concepts and definitions underlying these categories. The subsequent sections examine each classification in detail, highlighting how modern registration methods often integrate multiple strategies, thereby bridging these distinctions.

2.2.1 Rigid and Non-Rigid Registration Approaches

PCR methods can be broadly categorized into rigid and non-rigid techniques, depending on whether the underlying transformation involves local deformation. Let us denote by $X = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$ the source point cloud and by $Y = \{\mathbf{y}_1, \ldots, \mathbf{y}_M\}$ the target point cloud. The goal is to find a suitable transformation that maps points in X to corresponding points in Y.

Rigid registration assumes only a global transformation (rotation and translation) without changing the relative distances within each point cloud. Under this assumption, each point cloud is treated as a rigid body with no internal distortion. One of the most commonly used rigid registration algorithms is the Iterative Closest Point (ICP) method [28,29]. ICP iteratively finds correspondences between points in X and Y, and updates the transformation parameters to minimize alignment error. A typical ICP objective can be expressed as

$$\min_{\mathbf{R},\mathbf{t}} \sum_{i=1}^{N} \left\| \mathbf{y}_i - \left(\mathbf{R}\,\mathbf{x}_i + \mathbf{t} \right) \right\|^2, \tag{2.9}$$

where $\mathbf{R}$ is a rotation matrix, $\mathbf{t}$ is a translation vector, $\mathbf{x}_i \in X$, and $\mathbf{y}_i \in Y$ are corresponding points. Several variants of ICP exist, such as point-to-point and point-to-plane ICP [30], which improve accuracy by incorporating geometric constraints (e.g., surface normals).

Rigid registration is particularly effective for scenarios in which objects maintain a fixed shape, including robot navigation and autonomous driving (e.g., LiDAR-based SLAM), 3D object recognition, and industrial quality inspection. However, if there are significant shape changes, rigid alignment alone is insufficient.

Non-rigid registration accommodates local deformations such as bending, stretching, and scaling, making it suitable for cases where the shape of the source cloud $\mathbf{X}$ differs from that of the target cloud $\mathbf{Y}$. Applications include human body scans, soft tissue deformation in medical imaging, and character animation.

A common approach to non-rigid registration involves using deformation models like Thin Plate Splines (TPS) [31]. TPS seeks a smooth transformation $f(\cdot)$ that maps $\mathbf{x}_i \in \mathbf{X}$ to $\mathbf{y}_i \in \mathbf{Y}$. A TPS deformation can be written as

$$f(\mathbf{x}) = \boldsymbol{\alpha}_0 + \boldsymbol{\alpha}_1 x + \boldsymbol{\alpha}_2 y + \sum_{j=1}^{K} w_j\, \phi\big(\|\mathbf{x} - \mathbf{c}_j\|\big), \tag{2.10}$$

where $\phi(r) = r^2 \log(r^2)$, $\{\mathbf{c}_j\}_{j=1}^{K}$ are control points, and w_j are corresponding weights. Other approaches leverage machine learning, including deep neural networks, to capture complex deformations more efficiently [32]. One major challenge is avoiding overfitting or physically unrealistic deformations. Consequently, regularization terms that enforce smoothness and plausible transformations are frequently introduced.

Hybrid methods combine both techniques by first applying a rigid alignment to match the global structure of $\mathbf{X}$ and $\mathbf{Y}$, followed by a non-rigid refinement to capture local deformations. This strategy is common in medical imaging, where scans of the human body are first coarsely aligned via rigid transformations and then refined to model soft tissue or anatomical variations accurately.

2.2.2 Feature-Based Registration Methods

In the context of machine learning and PCR, a feature refers to a representation or descriptor that captures relevant information about a point, its local neighborhood, or the global shape of a 3D object. Features can be handcrafted, such as curvature, surface normals, or spatial coordinates, or learned automatically using neural networks, particularly in deep learning-

based approaches [33].

Features play a central role in point cloud processing tasks, including correspondence estimation, transformation prediction, and classification. Local features describe geometric or contextual properties of a point relative to its immediate surroundings, while global features encode holistic structural information about the entire object or scene. In learning-based frameworks, particularly those using graph-based or transformer-based architectures, features are typically represented as high-dimensional vectors produced through successive layers of convolution, attention, or aggregation mechanisms [33].

In the context of PCR, the term feature correspondence denotes the association between two points (or regions) from different point clouds that are assumed to represent the same physical location or structure in 3D space. Such correspondences are established by comparing their feature descriptors, which may encode geometric shape, spatial location, or contextual cues. Accurate feature correspondence is especially critical in non-rigid registration, where deformation between point clouds can be spatially varying and non-uniform. In these cases, relying solely on raw spatial proximity is insufficient; robust and informative features are essential to guide the alignment process effectively [34].

Feature-based registration is a commonly used approach for aligning point clouds [35, 36], particularly beneficial when significant initial misalignment makes direct point-to-point techniques prone to failure [37]. By focusing on distinctive geometric structures, these methods maintain robustness in challenging conditions involving noise, occlusion, or partial overlap [38, 39]. They are often employed in a coarse alignment phase [40], which can then be refined by algorithms such as the ICP method [28].

Figure 2.1 illustrates a typical pipeline for feature-based registration. First, keypoints are detected in each point set [36, 39], followed by the computation of local descriptors to capture geometric properties [38, 41]. The next step involves comparing these descriptors to establish correspondences between the two clouds [42]. Finally, solving a system of equations yields the rigid transformation parameters [43]. Depending on the application, optional stages, such as noise filtering or outlier rejection, may be introduced. RANSAC and its adaptations [44–48] are frequently used to discard erroneous matches and ensure reliable correspondences.

While some registration methods bypass explicit feature extraction entirely [49], identifying salient points often proves advantageous in complex or cluttered environments. Popular descriptors include Fast Point Feature Histograms (FPFH), which encodes local curvature and normal information [38], and 3D-SIFT, which detects distinctive keypoints that remain stable across varying scales [41, 50]. Alternatives like Harris 3D [39] or Intrinsic Shape Signatures (ISS) [51] emphasize corner-like or highly repeatable regions that persist under different viewpoints and transformations [49].

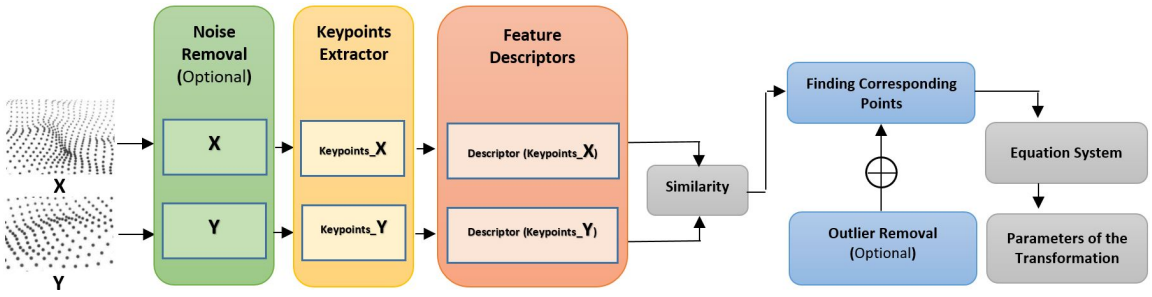


Figure 2.1. Overview of feature-based approaches for point cloud registration [35]

Once features are computed, descriptor matching typically employs a nearest-neighbor search or hashing to pair similar feature vectors [42, 52]. Randomized strategies such as RANSAC [44] then filter spurious correspondences by iteratively estimating transformations from randomly sampled subsets. After robust matches have been confirmed, transformation parameters can be found using methods like Singular Value Decomposition (SVD) [43] or least-squares optimization, aligning the two point clouds [53]. A subsequent fine registration stage, for instance, through ICP [28], may further refine this alignment and reduce the remaining errors.

Although feature-based registration tends to be more computationally intensive than direct point-to-point alignment, it significantly improves resilience to large pose differences and accommodates various real-world constraints [35, 37]. Consequently, it sees broad use in fields such as autonomous driving (e.g., Lidar-based SLAM) [54], 3D object recognition [55], medical imaging fusion [56], and augmented reality [57]. Continuous advances in parallel computing and machine learning have further enhanced the efficiency and accuracy of feature-based pipelines [58], reinforcing their centrality in modern 3D vision tasks.

2.2.3 Coarse-to-Fine Registration Approaches

PCR is often carried out through a coarse-to-fine approach, which applies two successive stages to maximize both efficiency and accuracy. First, a coarse alignment step estimates an initial transformation for two point clouds that may be significantly misaligned. Second, a fine alignment step refines this transformation to achieve high-precision registration.

When point clouds exhibit large offsets in position or orientation, attempting fine-grained alignment directly can be computationally expensive and prone to local minima. As a solution, coarse alignment relies on robust feature extraction and matching to compute an approximate transformation [38, 39, 59]. Feature descriptors such as Fast Point Feature Histograms (FPFH) [38] and keypoint detectors like SIFT-3D [59] or Harris 3D [39] help identify reliable correspondences.

Random Sample Consensus (RANSAC) [44] is frequently employed at this stage to

refine candidate transformations while discarding outliers. Advanced global methods, such as Go-ICP [60] and Super4PCS [61], further improve efficiency by searching for similar geometric structures in both point sets.

Once a coarse estimate is obtained, fine alignment techniques minimize local discrepancies between the two point clouds. As discussed before, the ICP algorithm iteratively finds correspondences and updates the transformation to reduce error. For scenarios involving non-rigid deformations, methods like Coherent Point Drift (CPD) [16] treat the data probabilistically. CPD aligns two point sets X and Y by modeling X as centroids of a Gaussian Mixture Model (GMM) and maximizing the data likelihood

$$p(\mathbf{Y} \mid \mathbf{X}, \theta) = \prod_{n=1}^{N} \sum_{m=1}^{M} \frac{1}{M} \mathcal{N}\big(\mathbf{y}_n \mid \mathbf{x}_m + \mathbf{v}_m,\ \sigma^2 \mathbf{I}\big), \tag{2.11}$$

where $\mathbf{v}_m$ is a displacement vector for $\mathbf{x}_m$, σ^2 is the variance of the isotropic covariance, and $\theta = \{\mathbf{v}_1, \ldots, \mathbf{v}_M, \sigma^2\}$. This approach iteratively updates the displacement vectors $\mathbf{v}_m$ and handles smooth, non-rigid transformations.

Combining these two stages, the coarse-to-fine registration pipeline efficiently tackles large misalignments first and then refines for precision.

2.2.4 Robustness-Oriented Methods

Robustness-oriented approaches in PCR aim to improve alignment reliability when confronted with real-world challenges such as noise, outliers, limited overlap, and uneven point density [7, 62]. Real-world data often contains sensor-related inaccuracies, environmental artifacts, and occlusions [63], making these approaches vital for achieving consistent and accurate registration in adverse conditions. In contrast to methods that assume ideal input data, robustness-oriented techniques incorporate specialized strategies (e.g., statistical filtering, adaptive weighting, and robust estimation) to mitigate errors and maintain stable alignment [64].

A primary obstacle in PCR is noise, which introduces small but disruptive deviations in point positions [6]. To mitigate this, robust preprocessing pipelines frequently incorporate smoothing or denoising techniques, such as Gaussian filtering or Moving Least Squares (MLS) [65].

In the context of point clouds, Gaussian filtering operates by re-estimating each point's position as a weighted average of its neighbors, where weights decrease with spatial distance, analogous to convolution in image domains. This reduces local fluctuations while preserving the overall structure. MLS, on the other hand, fits a smooth surface to a local neighborhood around each point and projects the point onto this surface, providing better

preservation of underlying geometry.

In addition to denoising, feature-based strategies enhance robustness to noise by focusing on local geometric descriptors (e.g., FPFH or Wavelet-based features) rather than raw coordinates [38, 51]. These features encode more stable characteristics such as curvature and neighborhood structure, enabling more reliable matching under noisy conditions.

Outliers constitute another significant issue, often caused by sensor errors or moving objects in the scene [44, 66]. RANSAC is widely used for dealing with outliers, iteratively sampling candidate correspondences, estimating a transformation, and evaluating its validity across the remaining data before discarding mismatches [44, 46]. More advanced variants, including M-estimators [67] and Expectation-Maximization (EM) approaches [68], adjust correspondence weights according to their inlier probability, thereby reducing the impact of spurious measurements on the final alignment.

Partial overlap poses further difficulties, particularly when only a subset of each point cloud represents the same region [69]. Standard algorithms like ICP often struggle in these scenarios [28]. Robust methods counter this by employing global feature matching techniques that do not require extensive one-to-one correspondences. Probability-based frameworks such as CPD and Gaussian Mixture Models (GMM) [16, 70] allow for soft correspondences and accommodate partial overlaps, while graph-based algorithms like Super4PCS [61] exploit shared geometric structures to identify matching segments even when overlap is minimal.

Point density variations can also compromise registration quality, especially in large-scale scans where certain areas appear dense and others appear sparse [71]. Multi-scale feature matching or adaptive weighting functions help normalize the influence of correspondences in high- and low-density regions [72]. Furthermore, machine learning and deep neural networks have become instrumental in learning robust representations that are invariant to noise levels, density shifts, or overlap fractions [35].

By integrating these robustness-focused techniques, PCR becomes more dependable in applications such as autonomous navigation (noisy, incomplete LiDAR data), medical imaging (partially overlapping scans), and robotics (dynamic, cluttered environments). Advanced filtering, statistical models, and adaptive algorithms collectively ensure that even under challenging conditions, registration retains both stability and accuracy, an essential requirement for modern 3D perception systems [35].

2.2.5 Local and Global Registration

In PCR, search space approaches refer to how algorithms navigate the space of possible transformations, such as rotation, translation, or deformation, to align two point sets. These

strategies are commonly categorized into global and local registration methods [62, 73, 74], based on their initialization requirements and search behaviors. Global methods operate without prior knowledge of the correct alignment and are therefore suited for cases with large initial misalignments. In contrast, local methods assume that a coarse initial transformation is available and aim to refine this estimate efficiently [37].

Global registration techniques search the entire parameter space without assuming a priori alignment. They are valuable when point clouds have substantial misalignment or unknown initial positions [60, 61]. Feature-based matching is a frequent strategy, using descriptors like FPFH or SIFT to establish correspondences [38, 41]. Algorithms such as 4-Point Congruent Sets (4PCS) and their extension Super4PCS [61] detect rigid transformations by identifying congruent point sets, handling significant noise and partial overlap [75]. Another prominent global method, Go-ICP, integrates a branch-and-bound search into the ICP framework [60], ensuring global optimality without needing an initial guess. Although robust, global strategies tend to be computationally demanding due to their exhaustive search process [35].

Local registration methods refine an existing coarse alignment by iteratively minimizing alignment errors [28]. They excel when an approximate transformation is already at hand, often converging quickly and accurately [37]. A well-known local algorithm is ICP, which repeatedly pairs the nearest points and updates the transformation to minimize their distances [28]. Variants such as point-to-plane ICP exploit surface normals for faster, more accurate convergence in structured environments [76]. However, local methods are ill-suited for large initial misalignments or limited overlap; in these cases, a poor initial guess risks convergence to an incorrect solution or local minimum [77].

Hybrid registration methods combine both approaches to balance global robustness with local efficiency [78]. Typically, a global algorithm provides a coarse alignment, which is then refined by a local technique for higher accuracy [37, 62]. This two-stage pipeline is common in many applications, such as Lidar-based SLAM, multi-modal medical image fusion, and 3D object reconstruction, where reliability, speed, and precision are all paramount [7, 79, 80].

Overall, the choice between global and local registration (or a hybrid of the two) significantly impacts accuracy, runtime, and tolerance to noise or overlap constraints [35, 73]. Global methods provide a more reliable solution from scratch but can be slow, whereas local methods converge quickly when a decent initial guess is known [37]. By tailoring these approaches to the application's data characteristics and computational requirements, optimal performance can be achieved across diverse real-world scenarios [7].

2.2.6 Loss Function-Oriented Approaches

Loss function-oriented approaches in PCR emphasize optimizing specific objective functions to evaluate and minimize alignment errors between two point clouds. Selecting a suitable loss function critically affects registration accuracy, robustness, and convergence. These methods include geometric distance-based metrics, probabilistic models, and robust statistical measures.

Geometric distance-based loss functions are popular due to computational efficiency and effectiveness in structured scenarios. For instance, the classical ICP algorithm minimizes the sum of squared Euclidean distances between corresponding points in the two point clouds, as shown in 2.9. A more refined approach, point-to-plane ICP, minimizes the perpendicular distance from a source point to the tangent plane at the target surface, improving convergence for structured surfaces [76]

$$L_{ptp}(\mathbf{R}, \mathbf{t}) = \sum_{i=1}^{N} \left((\mathbf{R}\mathbf{x}_i + \mathbf{t} - \mathbf{y}_i)^\top \mathbf{n}_i \right)^2, \tag{2.12}$$

where $\mathbf{n}_i$ denotes the surface normal at point $\mathbf{y}_i$. Such geometric methods are widely adopted in robotic mapping, LiDAR-based SLAM, and 3D reconstruction tasks [6].

Probabilistic frameworks provide alternative loss definitions, particularly useful when correspondences are uncertain or data are noisy or incomplete. EM is a prominent probabilistic method, modeling registration as a likelihood maximization problem [16]. CPD utilizes EM to represent the source point cloud as a GMM and iteratively maximizes the likelihood for alignment

$$\mathcal{L}EM(\theta) = \sum i = 1^N \log \left(\sum_{j=1}^{M} p(\mathbf{x}_i | \mathbf{y}_j, \theta) \right), \tag{2.13}$$

where θ represents model parameters. EM-based probabilistic methods excel in non-rigid registration scenarios, particularly in medical imaging and dynamic tracking applications [16, 81].

Robust statistical loss functions address the sensitivity of conventional squared-error metrics to outliers and noisy data. Examples include the Huber loss, Tukey's biweight function, and Cauchy loss, which diminish the influence of large residual errors while maintaining sensitivity to accurate alignments. The Huber loss function, for example, is defined as

$$L_{Huber}(r) = \begin{cases} \frac{1}{2} r^2 & \text{if } |r| \leq \delta, \\ \delta \left(|r| - \frac{1}{2}\delta \right) & \text{otherwise}, \end{cases} \tag{2.14}$$

where r is the residual alignment error, and δ is a predefined threshold parameter [82].

Recent developments in deep learning have significantly influenced loss function design by employing neural network-based alignment prediction. Learning-based methods integrate geometric consistency, feature similarity, and regularization terms into neural-network-driven loss functions. These approaches have demonstrated superior performance in complex and large-scale environments, including autonomous driving and real-time scene understanding [83, 84].

2.2.7 Optimization Strategies

Optimization strategies in PCR are employed to refine a transformation by minimizing an objective function that quantifies the misalignment between two point sets. These strategies play a critical role in ensuring convergence to an accurate solution, particularly after a suitable initialization has been provided. Although optimization typically begins from an initial guess, this initialization, often referred to as pre-alignment, is conceptually distinct from the optimization process itself. Pre-alignment methods aim to estimate a rough transformation to help avoid poor local minima, while optimization methods operate on a well-defined cost function to achieve precise alignment.

Once an initial estimate is available, various iterative optimization algorithms can be applied to minimize the registration error. A widely used method is the Levenberg-Marquardt (LM) algorithm [85], which combines aspects of gradient descent and Gauss-Newton methods [86]. LM iteratively refines the transformation parameters by minimizing the following non-linear least squares objective

$$\theta^* = \arg\min_{\theta} \sum_{i=1}^{N} \left\| \mathbf{R}(\theta)\mathbf{x}_i + \mathbf{t}(\theta) - \mathbf{y}_i \right\|^2, \tag{2.15}$$

where θ represents the transformation parameters, $\mathbf{x}_i$ and $\mathbf{y}_i$ are corresponding points, and $\mathbf{R}, \mathbf{t}$ denote rotation and translation components, respectively.

Other deterministic optimizers, such as Gauss-Newton [87], Limited-memory BFGS (LBFGS) [88], and gradient descent, are also employed in PCR tasks, particularly in formulations where the registration problem is cast as continuous optimization over transformation parameters. These solvers are chosen based on the characteristics of the objective function and the desired balance between convergence speed, memory efficiency, and robustness. While their performance may vary in terms of runtime and sensitivity to local minima, they generally produce comparable alignment results when the objective is well-formulated and the initialization is sufficiently accurate.

In addition to deterministic solvers, probabilistic optimization strategies such as the EM algorithm are also used. Methods like CPD [16] frame registration as a probabilistic match-

ing problem, where correspondences are modeled as probability distributions. The EM algorithm iteratively updates both the correspondence probabilities and the transformation to maximize the alignment likelihood.

In fact, the choice of optimization strategy can influence convergence speed and robustness, but it is often the problem formulation and quality of the initial estimate that determine the success of the registration process. Optimization algorithms refine this estimate by minimizing a well-defined objective, completing the alignment in a precise and reliable manner.

2.3 Evaluation Metrics for Point Cloud Registration

Assessing PCR involves examining both quantitative metrics and robustness criteria. Quantitative metrics measure how accurately the source point cloud aligns to the target, while robustness criteria evaluate an algorithm's performance under challenging conditions like noise or partial overlap. Generally, these evaluation metrics are categorized into three main groups: Distance-based metrics, transformation-based metrics, and matching accuracy metrics. Each of these categories provides different insights into the overall effectiveness of a registration method.

2.3.1 Quantitative Metrics

Quantitative metrics measure alignment accuracy between two point clouds. They can be further divided into distance-based and transformation-based metrics, with an additional focus on matching accuracy before the transformation is applied.

I) Distance-Based Metrics

Distance-based metrics capture how closely the transformed source point cloud aligns with the target point cloud, revealing whether correspondences are near-perfect or significantly deviating.

a) Chamfer Distance (CD) computes the mean of the shortest distances between points in one cloud and points in the other, penalizing misalignment.

$$CD(X,Y) = \sum_{x \in \mathbf{X}} \min_{y \in \mathbf{Y}} \|x - y\|^2 \; + \; \sum_{y \in \mathbf{Y}} \min_{x \in \mathbf{X}} \|y - x\|^2, \tag{2.16}$$

where $\mathbf{X}$ is the set of points in the source cloud, and $\mathbf{Y}$ is the set of points in the target cloud. The symbols x and y represent individual points in their respective sets.

b) Hausdorff Distance (HD) focuses on the greatest point-to-point deviation between

two clouds, enforcing strict alignment requirements.

$$HD(X,Y) = \max\left\{ \sup_{x\in\mathbf{X}} \inf_{y\in\mathbf{Y}} \|x-y\|,\ \sup_{y\in\mathbf{Y}} \inf_{x\in\mathbf{X}} \|y-x\| \right\}, \tag{2.17}$$

c) Mean Squared Error (MSE) averages the squared distance between corresponding points, revealing overall registration error.

$$MSE = \frac{1}{N}\sum_{i=1}^{N} \|x_i - y_i\|^2, \tag{2.18}$$

where x_i is the i-th point in the transformed source cloud, and y_i is the corresponding i-th point in the target cloud. N denotes the total number of corresponding points. The notation $\|\cdot\|$ represents the Euclidean norm.

d) Root Mean Squared Error (RMSE), the square root of MSE, reduces the influence of large individual errors and provides a more intuitive measure of overall deviation.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N} \|x_i - y_i\|^2}, \tag{2.19}$$

e) Earth Mover's Distance (EMD) interprets the registration task as a transport problem, calculating the minimal cost required to move points in one distribution to match the other.

$$EMD(X,Y) = \min_{\phi: X\to Y} \sum_{x\in X} \|x - \phi(x)\|, \tag{2.20}$$

The notation ϕ represents a bijection (one-to-one mapping) between points in X and Y.

f) Geodesic Distance (GD) measures the distance along a curved surface rather than a direct Euclidean line, crucial for shapes where surface geometry is significant.

$$GD(x,y) = \min_{\gamma\in\Gamma} \int_0^1 \|\gamma'(t)\|\, dt, \tag{2.21}$$

where x and y are points on the surface, and Γ is the set of all possible paths connecting x and y. The notation $\gamma'(t)$ denotes the derivative (tangent vector) of the path γ at parameter t.

II) Transformation-Based Metrics

Transformation-based metrics compare the estimated transformation to the ground truth transformation, highlighting how precisely the rotation and translation parameters are recovered.

a) Relative Rotation Error (RRE) calculates the discrepancy between the estimated rotation matrix and the true one, indicating the accuracy of rotational alignment.

$$RRE = \|\mathbf{R_{est}} - \mathbf{R_{gt}}\|, \tag{2.22}$$

where $\mathbf{R_{est}}$ is the estimated rotation matrix, and $\mathbf{R_{gt}}$ is the ground truth rotation matrix.

b) Relative Translation Error (RTE) measures how closely the estimated translation vector matches the actual translation, revealing translational accuracy.

$$RTE = \|t_{est} - t_{gt}\|, \tag{2.23}$$

where t_{est} is the estimated translation vector, and t_{gt} is the ground truth translation vector.

c) End-Point Error (EPE) averages the Euclidean distance between each transformed source point and its corresponding target point, providing a direct measure of registration quality.

$$EPE = \frac{1}{N}\sum_{i=1}^{N} \|\mathbf{T}(x_i) - y_i\|, \tag{2.24}$$

where $\mathbf{T}$ is the estimated transformation (encompassing both rotation and translation), x_i is the i-th point in the source cloud, and y_i is the corresponding i-th point in the target cloud. N represents the total number of corresponding points.

III) Matching Accuracy Metrics

These metrics focus on how accurately the algorithm identifies correct correspondences before a transformation is computed.

a) Inlier Ratio (IR) indicates the proportion of correctly matched point pairs relative to the total matches, spotlighting the quality of initial feature or keypoint correspondence.

$$IR = \frac{\text{Number of Correct Matches}}{\text{Total Number of Matches}}, \tag{2.25}$$

where Number of Correct Matches is the count of correspondences that are truly correct, and Total Number of Matches is the total set of correspondences proposed by the algorithm.

b) Feature Matching Recall (FMR) shows what fraction of true correspondences is detected correctly among all possible features, gauging how reliably features overlap.

$$FMR = \frac{\text{Number of Correct Feature Matches}}{\text{Total Number of Features}}, \tag{2.26}$$

where Number of Correct Feature Matches is the count of correctly identified matches, and Total Number of Features represents the total features detected in the point clouds.

c) Precision and Recall widely used in classification and detection tasks. These metrics measure correspondence quality from different angles, precision reflects how many selected matches are correct, while recall shows how many of the true matches have been found.

$$\text{Precision} = \frac{TP}{TP + FP}, \quad \text{Recall} = \frac{TP}{TP + FN}, \tag{2.27}$$

where TP denotes true positives (correctly identified matches), FP denotes false positives (incorrectly identified matches), and FN denotes false negatives (missed matches).

2.3.2 Robustness Evaluation

Real-world PCR is frequently affected by imperfections and uncertainties, which challenge the stability and accuracy of registration algorithms. To systematically evaluate robustness, we consider five common sources of degradation: Noise, outliers, partial overlap, data density variation, and deformation level, and provide formal definitions for each.

Noise. Noise refers to random perturbations affecting the position of observed points due to sensor inaccuracies or environmental effects. In PCR, noisy input causes the observed point cloud to deviate from the ideal surface.

Let $\mathbf{X} = \{\mathbf{x}_i \in \mathbb{R}^n\}_{i=1}^N$ be the clean source point cloud. Each point is corrupted by an additive random vector $\boldsymbol{\varepsilon}_i$, drawn from a probability distribution $p(\boldsymbol{\varepsilon})$, leading to

$$\tilde{\mathbf{x}}_i = \mathbf{x}_i + \boldsymbol{\varepsilon}_i, \quad \boldsymbol{\varepsilon}_i \sim p(\boldsymbol{\varepsilon}).$$

Common distributions include Gaussian noise $\mathcal{N}(0, \sigma^2\mathbf{I})$, Poisson noise, or uniform noise within bounded domains.

Outliers. Outliers are erroneous points in the point cloud that do not belong to the underlying object or surface. They often arise from reflections, occlusions, or artifacts in the sensor data.

Let the observed cloud be $\mathbf{X}_{\text{obs}} = \mathbf{X}_{\text{in}} \cup \mathbf{X}_{\text{out}}$, where $\mathbf{X}_{\text{in}}$ are inlier points and $\mathbf{X}_{\text{out}}$ are outliers. A point $\mathbf{x}_i$ is considered an outlier if

$$d(\mathbf{x}_i, \mathcal{M}) > \tau,$$

where $\mathcal{M}$ is the true surface manifold and τ is a threshold distance.

Partial overlap. Partial overlap occurs when only a portion of the source point cloud has a matching region in the target point cloud. This situation is common in multi-view scanning or occluded scenes.

Given source $\mathbf{X}$ and target $\mathbf{Y}$, define the overlapping subset as

$$\mathbf{X}_{\text{ov}} = \{\mathbf{x}_i \in \mathbf{X} \mid \exists\, \mathbf{y}_j \in \mathbf{Y} \text{ such that } \|\mathbf{x}_i - \mathbf{y}_j\| < \delta\}\,,$$

and the overlap ratio as

$$\rho = \frac{|\mathbf{X}_{\text{ov}}|}{|\mathbf{X}|}.$$

Lower ρ values correspond to more limited overlap.

Data density variation. Density variation refers to inconsistencies in the sampling resolution of different regions within a point cloud, caused by scanning angles, distance, or sensor limitations.

For a point $\mathbf{x}_i \in \mathbf{X}$, define local point density as

$$\rho(\mathbf{x}_i) = |\{\mathbf{x}_j \in \mathbf{X} \mid \|\mathbf{x}_j - \mathbf{x}_i\| < r\}|\,,$$

where r is the neighborhood radius. Significant variation in $\rho(\mathbf{x}_i)$ across points indicates non-uniform density, which can affect correspondence estimation.

Deformation level. Deformation level quantifies the extent of non-rigid transformation between two point clouds. It measures how much individual points have been displaced due to bending, stretching, or other deformations.

Let each $\mathbf{x}_i \in \mathbf{X}$ correspond to $\mathbf{y}_{j(i)} \in \mathbf{Y}$. The displacement vector is

$$\mathbf{u}_i = \mathbf{y}_{j(i)} - \mathbf{x}_i.$$

The average deformation level is defined as

$$\Delta = \frac{1}{N} \sum_{i=1}^{N} \|\mathbf{u}_i\|_2.$$

Larger Δ values represent more significant shape changes between the source and target.

By introducing these mathematically grounded definitions, we establish clear criteria for evaluating the robustness of registration methods under varying real-world conditions. These formulations are used consistently throughout this work. Figure 2.2 provides a comparative summary of robustness performance in recent learning-based non-rigid registration methods.

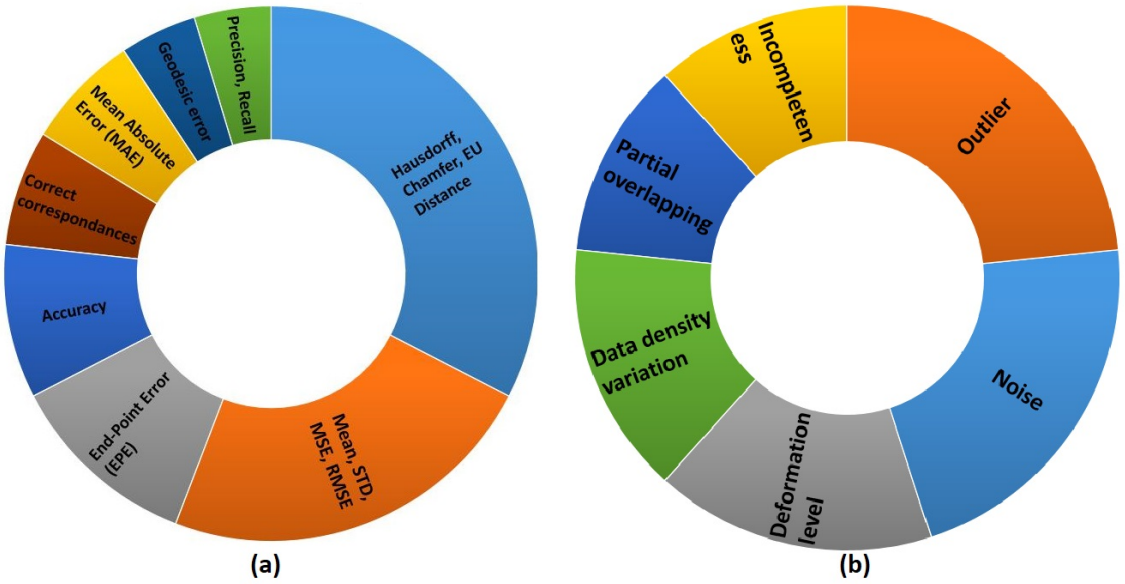


Figure 2.2. Quantitative evaluation metrics and robustness to common challenges for learning-based and non-rigid 3D point cloud registration methods published between 2017 and 2021: (a) Assessment metrics (b) Robustness analysis [35]

2.4 Overview of Learning-Based Non-Rigid Point Cloud Registration Methods

Learning-based methods have become increasingly important for non-rigid PCR because they can capture complex deformations and underlying data distributions more effectively than traditional optimization-based approaches [5]. Traditional methods, while powerful for small or structured deformations, can struggle with large or highly non-linear transformations. In contrast, deep learning models can be trained on diverse sets of shape variations to automatically learn robust correspondences and handle significant geometric changes. By leveraging large amounts of training data, these models adapt to real-world variations such as missing data, noise, and partial overlaps, often resulting in improved registration accuracy and generalization [89].

Another key advantage of learning-based methods lies in their inference speed and scalability. Once a model is trained, registration of new point clouds can be executed in near-real-time, as opposed to iterative optimization-based techniques, which can be time-consuming for large-scale or high-resolution data [89]. Furthermore, many learning-based pipelines can be designed to work end-to-end, integrating multiple tasks, such as segmentation, feature extraction, and correspondence prediction, into a single unified framework. This holistic approach often leads to more coherent and consistent transformations, reducing the need for separate, potentially error-prone post-processing steps. Learning-based solutions offer a compelling way to tackle the complexity of non-rigid PCR, enabling more robust, effi-

cient, and versatile 3D shape analysis [35]. Learning-based registration approaches commonly employ various network architectures, including CNNs, Recurrent Neural Networks (RNNs), GCNNs, and Multi-Layer Perceptron (MLP). This section surveys several notable non-rigid registration techniques built on these architectures, emphasizing those that are frequently cited and featured in reputable journals.

It is important to recognize that many of the methods discussed here can be categorized into multiple classes introduced earlier. Nevertheless, for clarity, we organize these learning-based methods according to their respective network architectures. An overview of the techniques examined is provided in Table 2.1.

2.4.1 Convolutional Neural Networks

CNNs have shown remarkable success in computer vision tasks and have been adapted for PCR. The core objective is to learn robust and discriminative features directly from 3D data so that correspondences between point clouds can be determined effectively. Nonetheless, unlike 2D images, point clouds are inherently unstructured and irregular, making it non-trivial to apply standard 2D convolutions. Below, we discuss three common strategies for leveraging CNNs in PCR [90].

Volumetric representations. One common approach is to convert the point cloud into a structured voxel grid, allowing standard 3D convolutions to be applied. By treating the 3D space as a voxelized volume, CNNs can extract features that capture both local geometry and global context [91,92]. Here, each voxel in the grid can hold information about whether it is occupied by one or more points (e.g., binary occupancy) or store aggregated features such as point densities or intensities. Convolution operations then proceed in 3D just as they do in 2D CNNs, although with higher memory requirements. This allows powerful feature extraction in a structured manner, but can suffer from quantization artifacts and a significant computational burden when dealing with high-resolution grids.

Multi-view projections. Another method projects the 3D data onto multiple 2D views (e.g., using spherical or cylindrical projections) so that 2D CNNs can be utilized. This approach leverages the maturity of 2D CNN architectures but may lose some 3D details during projection. Carefully designed view aggregation strategies can mitigate this issue and improve overall registration accuracy [93, 94].

Point-based convolutions. Recent advancements in point-based methods enable the definition of convolution-like operators directly on point sets. These networks (sometimes referred to as “point CNNs”) learn local neighborhood features without relying on regular grids, preserving the original geometric information. Techniques such as KPConv and other variants introduce specialized layers and kernels that capture local context and aggregate

features in a hierarchical manner [95].

In the following, we present several well-known approaches that use CNNs to address registration problems.

ProRegNet [96] is a notable CNN-based method that incorporates biomechanical constraints to enhance registration accuracy. In a clinical setting, aligning Magnetic Resonance (MR) and Transrectal Ultrasound (TRUS) images of the prostate can be challenging because the prostate may deform differently in each modality. To tackle this issue, a non-rigid registration framework has been proposed that leverages CNNs for segmentation, as well as a specialized point cloud matching network for the final alignment. First, two separate CNNs segment the prostate in MR and TRUS images, respectively. Each segmentation yields a volumetric mask that delineates the prostate boundary, effectively isolating the region of interest in both modalities. Following segmentation, tetrahedron meshing is applied to the prostate masks to create 3D point clouds representing the prostate geometry in each image. Instead of relying on iterative optimization techniques for non-rigid registration, the authors introduce a dedicated point-cloud-based network to directly match the 3D structures. This network is trained using deformation fields derived from finite element analysis (FEA). Finite element models simulate realistic tissue deformations under various forces, embedding biomechanical constraints into the network weights. Consequently, the network not only finds correspondences in a data-driven manner but also encourages solutions that align with plausible prostate deformations rather than arbitrary warpings.

In another approach, called the Volume-To-Surface Registration Network (V2S-Net) [97], the CNN takes two 3D meshes as input: (1) a preoperative liver volume (from a CT scan) and (2) a partial intraoperative surface (extracted from a laparoscopic stereo video). The network then outputs a displacement field that warps the preoperative organ mesh to align with the intraoperative surface. Unlike traditional registration pipelines that separately handle correspondence estimation and deformation modeling, this CNN simultaneously learns to resolve surface correspondence and ensure biomechanically realistic transformations. To achieve this, synthetic "organ-like" meshes are generated and deformed using Finite Element Method (FEM) simulations, creating a wide range of realistic shape variations for training. By exposing the network to diverse, physically plausible deformations and deliberately adding noise to mimic real surgical conditions, V2S-Net acquires the capability to generalize to new patient data without retraining. Through this end-to-end formulation, the CNN effectively manages the sparse, noisy partial point clouds obtained intraoperatively, providing a robust and rapid solution for non-rigid liver registration.

Point Registration Neural Network (PR-Net) [98] tackles point set registration by learning a parametric mapping that directly predicts a spatial transformation, rather than relying on iterative optimization. PR-Net's pipeline is organized into three main components. In

the first stage, learning shape descriptor tensor, a grid-reference structure is introduced to handle arbitrarily structured data and extract robust feature representations from each point set. Next, in the learning shape correlation tensor stage, an all-to-all point-wise computation is performed on these descriptor tensors to formulate a detailed correlation relationship between the source and target point sets. Finally, the learning of transformation parameters stage leverages a CNN as a functional regression model. By mapping the shape correlation tensor to the transformation parameters, PR-Net pinpoints the best geometric alignment in one forward pass, without iterative refinement.

In [99], the authors leverage a CNN to learn feature embeddings that reflect the geometry and local context of points on 3D human shapes, both unclothed and clothed. Specifically, they modify a standard classification CNN (based on AlexNet [100]) to output feature descriptors for sub-regions of the human body rather than simply class labels. By incorporating a multi-segmentation technique on the training shapes, the network is encouraged to produce smooth embeddings, meaning that points on the surface of a human body that are geodesically close will map to nearby positions in the learned feature space. This CNN-based approach captures the variations in pose and clothing across the different training sets, enabling dense correspondence matching between partial scans, depth maps, and full 3D models. Rather than relying on iterative or optimization-based registration methods, the learned descriptors allow for a single-step matching of points to their counterparts, substantially reducing outliers and increasing efficiency. As a result, the method can handle complex poses and various clothing styles, making it more versatile than traditional non-rigid registration pipelines.

DispVoxNets [101] is a neural network–based approach that performs non-rigid point set registration by converting point sets into regular 3D voxel grids and then regressing displacement fields within this voxel space. Because point sets can vary in their number and ordering of points, the authors circumvent the need for a fixed input size by turning each input into a uniformly sampled voxel representation. The core engine is a 3D CNN that processes these voxel grids to predict per-voxel displacement vectors, effectively warping the “template” shape to the “reference” shape. By training on collections of deformable objects (such as clothes, human bodies, and faces) with known intra-state correspondences, DispVoxNets learn category-specific deformation priors as well as constraints like weak topology preservation. The 3D CNN architecture also helps maintain robustness against large deformations, noise, and clustered outliers, while offering fast inference compared to traditional iterative methods.

In [102], Parallel Frames CNNs (PFCNNs), a method for applying CNNs directly on surface meshes, is introduced, which typically poses difficulties for CNNs due to the non-Euclidean nature of surface meshes. Rather than representing a shape as a point cloud or

volumetric grid, the authors define a parallel frame field on the surface, consistently aligning local tangent spaces. They then map each local patch onto a flat, Euclidean-like coordinate system, thereby enabling standard convolution operations analogous to those in regular 2D CNNs. A key component of PFCNNs is the use of locally flat connections, which is a concept from discrete differential geometry to enforce parallel alignment across neighboring tangent planes. This alignment is encoded through a pointwise tangential N-direction frame field, making it possible to systematically handle the absence of a canonical axis on a curved surface. For each local patch, the network employs regular grids in the tangent space and applies typical convolutional filters, thereby leveraging existing 2D CNN designs. By effectively preserving local geometry and aligning features in tangent space, PFCNNs can capture fine-grained details on meshes without resorting to specialized, handcrafted surface features. In experiments, PFCNNs exhibit robustness and high performance across tasks such as classification, segmentation, and registration on both deformable and rigid surfaces. Because this framework closely mirrors conventional CNNs, it can incorporate efficient architectures (e.g., residual blocks, dilated convolutions) used in 2D image processing.

2.4.2 Graph Convolutional Neural Networks

GCNNs provide a powerful framework for learning features directly from the intrinsic geometric relationships in point clouds, making them highly effective for registration tasks [103, 104]. Rather than using an unstructured set of points or regular voxel grids, a GCNN constructs a graph whose vertices correspond to points (or patches) in the cloud, while edges capture local neighborhood relationships (e.g., via k-nearest neighbors or a radius-based approach). This graph representation enables the network to propagate and aggregate information among spatially connected points, yielding context-sensitive descriptors that can be more robust than simpler point-wise features. In the context of PCR, GCNNs are typically used in one of two ways:

Descriptor learning and correspondence estimation. A GCNN first learns a latent descriptor for each node in the graph. Points with similar local neighborhoods in the source and target clouds receive similar descriptors, facilitating correspondence matching. A transformation (rigid or non-rigid) can then be estimated from these correspondences. Unlike classical handcrafted descriptors, GCNN-derived embeddings can adapt to training data distributions and remain robust under noisy, partially overlapping conditions [105].

End-to-end registration. In a fully end-to-end pipeline, GCNNs can learn not only pointwise features but also a direct alignment strategy. For instance, one might combine a GCNN-based encoder with a learnable transformation regression module (e.g., predicting

rotation, translation, or more complex deformations). The underlying graph structure enforces local geometric consistency, helping the network infer valid global transformations even for large deformations or cluttered scenes.

By capturing local connectivity patterns and global context via repeated graph convolutions, GCNNs naturally handle irregular sampling densities and preserve topology better than purely pointwise networks. Consequently, GCNN-based approaches excel in complex registration scenarios, including those with significant shape deformation, noise, and incomplete data [106].

In [106], it is discussed how these networks process data structured as graphs, learning geometric features based on the relationships between neighboring nodes. This approach is particularly effective for point clouds, where the structure of the data is inherently graph-like. In the medical domain, GCNNs have shown promise for tasks like 3D lung registration [107], where edge convolutions are used to extract geometric features, and Loopy Belief Propagation (LBP) regularizes displacements on a k-nearest neighbor graph. Additionally, [108] introduces a dynamic GCNN approach for PCR, which refines correspondences probabilistically using the CPD algorithm. These approaches demonstrate the versatility of GCNNs in handling various PCR challenges, particularly when dealing with complex deformable structures.

Continuing this line of research, [109] introduces NrtNet, an unsupervised transformer-based network designed for non-rigid PCR. This method leverages self-attention mechanisms to extract feature correspondences between large deformations. NrtNet's three main components, a feature extraction module, a correspondence matrix generation module, and a reconstruction module, work together to align point clouds by learning and normalizing correspondence probabilities. This approach is designed to handle large-scale deformations, a significant challenge in non-rigid PCR. Extending this work, [20] proposes GraphSCNet, a network that tackles outlier correspondence pruning, particularly in non-rigid PCR. GraphSCNet addresses the challenge of local rigidity in non-rigid deformations by using a local spatial consistency measure to evaluate correspondence compatibility, ensuring better outlier discrimination and improving the overall accuracy of registration.

Building on the concept of structural alignment, [110] presents NIE, a method for embedding the vertices of point clouds into a high-dimensional space to preserve intrinsic structural properties. This technique is particularly useful for aligning point clouds sampled from deformable shapes, which often lack explicit structural information. NIE forms the foundation for a weakly-supervised framework for non-rigid PCR, avoiding expensive preprocessing steps and the reliance on ground-truth correspondence labels.

SyNoRiM [111] is a deep learning-based framework for multiway, non-rigid PCR that employs a fully connected graph representation for each shape and learns functional bases

directly from raw point cloud data. Rather than relying on predefined operators such as the Laplace–Beltrami basis, SyNoRiM uses GCNN to extract smooth, band-limited representations of local geometry and global context, encouraging nearby source points to stay close under deformation. This learned basis allows SyNoRiM to establish coherent pairwise correspondences between arbitrarily sampled, partial, or noisy scans and subsequently enforces cycle-consistency across all scans through a functional map synchronization step. By integrating these ideas into a single pipeline, SyNoRiM discovers a latent canonical shape and refines all pairwise deformations accordingly, achieving robust, end-to-end multiway registration that can handle non-isometric deformations, occlusions, and other real-world complexities.

The advantages of using graph structures for point cloud representation also justify a closer look at the EdgeConv operator. EdgeConv, introduced by [104], is a neural network module specifically designed for point cloud analysis. It effectively captures both local neighborhood geometry and global context through edge-based feature aggregation.

EdgeConv operates on a set of points represented as a matrix $X = \{\mathbf{x}_i \in \mathbb{R}^d : i = 1, 2, ..., n\}$, where n is the number of points and d is the dimensionality of each point. It also takes as input a matrix of neighbor indices that defines the graph structure. For each pair of neighboring points $(\mathbf{x}_i, \mathbf{x}_j)$, the edge feature $\mathbf{e}_{ij}$ is computed using a nonlinear function h_Θ, parameterized by learnable weights Θ. This function combines the relative position $\mathbf{x}_j - \mathbf{x}_i$ and the central coordinate $\mathbf{x}_i$ to capture both local and contextual information. The edge feature computation is given by

$$e'_{ijl} = \text{ReLU}\left(\theta_l \cdot (\mathbf{x}_j - \mathbf{x}_i) + \phi_l \cdot \mathbf{x}_i\right), \tag{2.28}$$

where θ_l and ϕ_l are learnable parameters. The aggregated feature for point $\mathbf{x}_i$ is obtained by applying max pooling over its neighborhood

$$x'_{il} = \max_{j:(i,j)\in\varepsilon} e'_{ijl}. \tag{2.29}$$

The output is an m-dimensional feature representation for each point, preserving the original point count. Multiple EdgeConv layers can be stacked to iteratively enrich the representation. EdgeConv has demonstrated strong performance in capturing local geometric structures and improving accuracy in various 3D vision tasks [104].

2.4.3 Multilayer perceptrons and PointNet-Based Methods

MLPs and PointNet-inspired architectures have become pivotal in handling unordered point cloud data for registration tasks. MLPs typically process each point or local neighborhood

independently through fully connected layers, followed by a permutation-invariant pooling operation. This design naturally accommodates irregular input sizes and directly outputs either pointwise displacements or global transformation parameters. Meanwhile, PointNet [112] introduces an influential framework that encodes each point's features via a shared MLP, then aggregates them using a symmetric pooling function (e.g., max pooling). PointNet++ [113] enhances this approach with a hierarchical feature extraction scheme, capturing local geometric structures at multiple scales. In registration pipelines, both MLPs and PointNet-based encoders often pair with additional network components that either iteratively refine alignment or directly learn transformations in an end-to-end manner.

PointNetLK [114] merges PointNet's global feature encoding with the classical Lucas-Kanade (LK) algorithm to solve rigid registration. Here, the source and target point clouds are passed through the PointNet encoder, producing two global feature vectors. These features are compared iteratively using an LK-style gradient update to converge on the optimal rotation and translation. A later refinement, detailed in [115], proposes a decomposition of the Jacobian matrices involved in the LK step, improving convergence stability for larger motions or noisier data.

In the domain of non-rigid registration, GP-Aligner [116] tackles groupwise alignment by learning Group Latent Descriptors (GLD). The approach couples these latent descriptors with MLP-based prediction modules to encode and transfer coherent deformations across multiple shapes. By modeling a shared representation of shape variations within a dataset, GP-Aligner excels at ensuring consistent alignment for collections of related objects, rather than merely tackling pairwise registration.

Another notable framework, Neural Deformation Pyramid (NDP) [117], exploits hierarchical MLPs with sinusoidal encodings to approximate non-rigid deformations at multiple scales. This pyramid-like decomposition lets the network capture both coarse global transformations and fine local details. The multi-resolution design accelerates convergence and improves robustness when dealing with large deformations or diverse shape variations.

CPD-Net [118] draws inspiration from the CPD algorithm (see Section 2.2.3), avoiding iterative registration by training a network to predict point-level displacements in a single forward pass. This network relies on PointNet-like feature extraction to generate pointwise embeddings, which are then processed by MLP layers to produce coherent drifts. By encoding coherence constraints in the learned features, CPD-Net achieves accurate alignment while maintaining topology and preserving local structure.

FlowNet3D [119] adapts the idea of optical flow estimation to 3D point sets. It uses a PointNet++ encoder-decoder architecture to predict scene flow, which can be interpreted as a dense motion field guiding registration. However, performance in dynamic or complex scenarios led to the development of FlowNet3D++ [120], which integrates refined cost-

volume computations and additional grouping strategies to better capture intricate motions. Although these frameworks focus primarily on scene flow, they illustrate how point-level MLPs and hierarchical pooling can address registration by directly estimating per-point 3D motion vectors.

2.4.4 Transformer-Based Methods

Transformers are widely recognized in deep learning for their ability to model long-range dependencies via attention mechanisms, initially gaining prominence in natural language processing (NLP) tasks such as machine translation [121]. By discarding the need for recurrent or convolutional structures, transformers leverage parallelization and global context, which has led to extensive success in NLP domains and subsequent adaptations in computer vision. This adaptability naturally extends to 3D point cloud processing, where irregular, unordered data often poses difficulties for traditional neural networks.

Early works on transformers for 3D data focused on tasks like classification and segmentation. For instance, methods such as [122] and [123] present transformer-based architectures that capitalize on self-attention to capture local and global geometric relationships in unordered point sets, demonstrating strong performance in shape classification, semantic segmentation, and normal estimation. Although initially geared toward perception-related tasks, the application of transformers in PCR has more recently gained attention. Transformers can facilitate PCR by leveraging their robust attention mechanisms to learn correspondences and transformations between sets of 3D points. Below are the primary ways in which transformers are employed in this domain:

Global context and permutation invariance. Unlike convolutional or recurrent models that rely on fixed local receptive fields or sequential ordering, transformers compute pairwise "attention" across all points [121]. This attention-based mechanism naturally accommodates the unordered nature of point clouds, allowing the network to capture both local and long-range relationships without imposing a rigid input structure.

Learned correspondences and feature matching. A core step in many registration pipelines is to identify reliable correspondences between source and target clouds. Transformer layers, equipped with multi-head attention, can learn discriminative features that align points of similar geometric context [122, 123]. By focusing attention on overlapping regions, these methods can handle partial scans, occlusions, or noisy inputs, often circumventing the need for explicit keypoint detection.

End-to-end deformation estimation. Some transformer-based approaches regress motion parameters directly from attention-derived correspondences, obviating additional post-processing or iterative methods [84]. By integrating a final layer or small module to es-

timate the overall (rigid or non-rigid) transformation, the entire alignment process can be made end-to-end, reducing complexity and inference time.

Coarse-to-fine hierarchies. Transformers can be utilized across multiple resolutions to refine matches incrementally [124]. A coarse alignment stage may operate on downsampled "superpoints" or graph nodes, while subsequent fine-scale transformer layers refine local correspondences. This hierarchical approach lowers computational overhead, focusing attention on critical regions for improved alignment accuracy.

Position encoding and spatial cues. Originally developed for sequential data, transformers rely on positional encodings to provide a notion of ordering [121]. In the 3D context, adapted encodings embed coordinates or distance metrics that capture geometric relationships among points [125]. More advanced schemes employ tailored encodings that handle varying reference frames, occlusions, or partial overlaps, thereby improving the model's resilience to real-world sensor noise and viewpoint changes.

These features, attention-driven global context, flexible positional encodings, and direct transformation predictions, permit transformers to manage complex registration scenarios. Their strengths include accommodating partial data, substantial motion, and heterogeneous point densities, making them potent alternatives or complements to traditional geometry-based and earlier learning-based methods in 3D applications.

While transformers have found extensive application in point cloud classification and segmentation, their use in registration tasks remains comparatively underexplored. One of the earlier transformer-based methods for rigid PCR is Deep Closest Point [84], which relies on a three-stage pipeline: (1) extracting features from the input point clouds using an embedding network, (2) estimating combinatorial matches between points via an attention-based pointer-generation layer, and (3) determining the rigid transformation with a differentiable SVD layer. Building upon this direction, [21] introduces a Geometric Transformer that sidesteps traditional keypoint extraction by matching so-called superpoints, downsampled point clusters, using the overlap of their neighborhoods. This superpoint-based matching not only enhances robustness under low-overlap conditions but also achieves invariance to rigid transformations. In a similar vein, CoFiNet (Coarse-to-Fine Network) [124] omits keypoint detection and instead adopts a hierarchical registration strategy. At a coarse level, the method identifies downsampled nodes with higher overlap via a weighting scheme and narrows down the areas of interest before refining the matches at a finer scale. Notably, a density-adaptive matching module helps handle both differing point densities and overlapping segments, thereby illustrating ongoing advancements in the efficiency and accuracy of PCR under real-world constraints.

An additional notable approach, OIF-PCR [126], targets alignment challenges caused by inconsistent reference frames. Its main innovation is to identify a single inlier correspon-

dence using a differentiable optimal transport layer, then normalize each point's position based on that inlier for subsequent encoding. This design reduces ambiguity and bolsters spatial consistency, and by incorporating an iterative optimization scheme, progressively refines the registration outcome. In parallel, [127] proposes an end-to-end solution using transformer layers to directly estimate correspondences. The model's self- and cross-attention mechanisms obviate the need for classic steps like feature matching and RANSAC filtering, allowing the network to learn correspondences in a unified pipeline and compute the rigid transformation without additional post-processing.

Finally, the partial point cloud scenario is addressed by Lepard [125], which integrates a KPFCN feature extractor, transformer modules, and differentiable matching algorithms. Central to this method are three key innovations that reinforce the role of 3D positional cues: disentangling the feature and position representations, introducing a specialized positional encoding for relative 3D distances, and employing a repositioning module that adjusts inter-point distances across partial scans. These enhancements yield improved robustness and precision in partial registration settings, underlining transformer-based models' growing potential to tackle complex point cloud data across a wide range of 3D registration tasks.

2.4.5 Other network architectures (GAN, RNN, ResNet, T-Net)

While CNNs, GCNNs, and Transformer-based models remain dominant in PCR, there are additional architectures that, though less common, have shown promise in specific scenarios. These include Generative Adversarial Networks (GANs), RNNs, and Residual Networks (ResNet).

1) Generative Adversarial Networks

Generative Adversarial Networks (GANs) can be adapted to PCR by framing the alignment task as one of learning a transformation that fools (fooling operator) a discriminator into believing that a deformed source shape matches the target shape [128, 129]. Below is a concise overview of this process:

Generator (transformation network). The generator is conditioned on the source point cloud (and optionally additional context) and produces a transformation, often formulated as a per-point displacement field or deformation parameters, and when applied to the source, yields a shape resembling the target. Through adversarial training, the generator learns to warp the source such that it becomes indistinguishable from the target.

Discriminator (real vs. generated alignment). Operating in parallel, the discriminator receives either the real target point cloud or the generator's transformed output, classifying it as either real or fake. This adversarial loss [128] drives the generator to produce transfor-

mations that deceive the discriminator, thereby promoting closer alignment of the source to the target.

Loss functions and constraints. In addition to the adversarial objective, geometric or smoothness constraints are frequently imposed [129]. Examples include penalties on extreme stretching or compression and explicit alignment objectives (e.g., pointwise distances) to preserve local fidelity. Such constraints steer the generator toward more realistic and spatially consistent deformations.

Training and inference. Training proceeds by iteratively updating both the generator and the discriminator in an adversarial loop: The generator strives to align the source to the target more convincingly, while the discriminator attempts to refine its ability to detect misalignments. Inference requires only the trained generator, enabling direct prediction of the deformation needed to align a new source point cloud to the target without iterative optimization during testing.

GAN-based registration leverages high-capacity neural networks to capture complex, non-linear deformations, embedding data-driven shape priors to avoid unnatural warping. It is often robust to noise and partial data since the discriminator penalizes outputs deviating significantly from realistic geometries. Using GANs shifts registration from a purely geometric optimization problem to an adversarially trained framework, which can excel in scenarios involving large or intricate shape variations. For example, conditional GANs [128] have been employed to learn geometric transformations for non-rigid registration [129]. In this setup, the discriminator attempts to distinguish between transformed source shapes and genuine target shapes, while the generator (conditioned on the source shape) learns to produce transformations that map the source to the target. This adversarial training scheme captures complex, high-dimensional deformation patterns that might be difficult to model with direct optimization approaches. By leveraging robust adversarial objectives, GAN-based methods can alleviate issues such as mode collapse or overfitting to specific deformation types, making them suitable for non-rigid registration tasks where shape variation is high.

2) Recurrent Neural Networks

Although RNNs, including Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) variants, are typically associated with temporal sequence modeling, they can also be adapted for PCR by treating the alignment task as a sequential process. In this view, each step refines either the transformation parameters or intermediate features, gradually converging to an accurate registration.

Sequential alignment strategy. One approach is to impose a pseudo-sequential ordering on the points in a cloud and feed them into an RNN, allowing the hidden state to integrate partial registration cues over multiple time steps [130]. For example, an RNN can iteratively

refine a set of transformation parameters by comparing intermediate alignment outcomes with the next subset of points. Each iteration makes small adjustments to reduce alignment error, effectively distributing the registration burden over multiple recurrent steps.

Iterative update of features and transformations. Instead of directly outputting a global transformation in one pass, an RNN can maintain an internal representation of the alignment status. At each step, it processes local features from the next point (or small cluster of points) in the pseudo-ordered set, updates the transformation estimates (e.g., rotation, translation, or non-rigid offsets), and outputs refined alignment parameters. This iterative process continues until sufficient convergence is reached. The recurrent structure naturally tracks alignment progress over time (i.e., steps of the iteration), mitigating large jumps that might destabilize training.

Balancing permutation invariance and temporal ordering. Because point clouds are inherently unordered, a key challenge is preserving the benefits of permutation invariance while still defining a sequence for the RNN to process [130]. Potential solutions include: (1) randomizing the order of points between training epochs to avoid overfitting to any single ordering. (2) using spatial heuristics (e.g., sweeping from nearest to farthest) to create a more geometrically meaningful sequence. (3) combining RNNs with global pooling or attention mechanisms to capture global context despite sequential input.

These strategies help maintain robustness to different sampling densities and viewpoints. RNN-based pipelines benefit from incremental refinement, allowing for a flexible trade-off between computation time and registration accuracy. However, they can be more sensitive to point ordering than architectures like PointNet or Transformers, and potentially require more careful engineering to ensure stability, particularly when dealing with large or noisy point clouds. Despite these considerations, RNNs remain a viable option for registration, especially when sequential thinking (e.g., iterative correction) aligns well with the desired application.

3) Residual Networks

Residual Networks (ResNets) [131] are commonly associated with 2D image tasks, yet they also lend themselves well to PCR, particularly for non-rigid alignment. By integrating skip connections, ResNets maintain stable gradient flow across many layers, which is critical when dealing with the large parameter spaces and complex deformations inherent to PCR.

Skip connections in deformation layers. A representative method, ResNet-LDDMM [132], weaves skip connections into the layers responsible for computing deformation fields. Each residual block refines the estimated transformation from the previous stage (e.g., rotation, translation, or a per-point offset), combining new updates with the prior state. This incremental refinement strategy ensures that errors do not accumulate excessively, mitigat-

ing vanishing or exploding gradients and leading to more robust convergence.

Iterative transformation refinement. Residual architectures naturally suit iterative deformation strategies. Rather than predicting the full alignment in one pass, a ResNet can sequentially adjust transformation parameters, effectively "adding" corrections to the current estimate. This setup is especially useful for handling large or complex deformations, as each layer can focus on local adjustments relative to the latest deformation field instead of attempting to solve the entire registration in a single step.

Benefits for high-dimensional spaces. Because PCR often operates in high-dimensional parameter spaces (e.g., non-rigid motion, large shape variability), stable training is vital. Skip connections help preserve gradient signals, allowing deeper networks to capture more nuanced geometric features without suffering from convergence difficulties. Consequently, ResNet-based models can learn more detailed deformation mappings, improving alignment quality across diverse shapes or complex real-world data.

4) T-Net: Spatial transformation network

T-Net is one of the most important modules to understand and study separately when working with point cloud data. Although it is not explicitly designed for PCR, it plays a critical role in many point cloud-based applications by learning to align and normalize data, thereby improving model invariance to geometric transformations such as rotation and scaling.

T-Net was originally introduced in the context of Spatial Transformer Networks by [133], and later adapted in PointNet [112] as a module that learns a transformation matrix to align input point clouds into a canonical space. The core idea is to apply a data-driven affine transformation to the input, making the subsequent feature extraction more invariant to spatial perturbations.

The architecture of T-Net consists of the following stages:

- ***Feature extraction via MLP:*** The input point cloud is processed by a shared multi-layer perceptron (MLP), typically implemented as a sequence of 1D convolutional layers with output channel sizes of 64, 128, and 1024. Each convolutional layer is followed by instance normalization and a ReLU activation. This stage extracts local features for each point.

- ***Global feature aggregation:*** A symmetric function, in this case, max-pooling, is applied across all point features to obtain a global feature vector of size 1024. This vector summarizes the global structure of the entire point cloud.

- ***Transformation regression:*** The global feature vector is passed through another MLP with fully connected layers of sizes 512, 256, and 9. The final layer outputs a flattened

3×3 matrix (for 3D point clouds), which represents the affine transformation to be applied. The output matrix is reshaped into a 3×3 matrix T, which is used to transform the input points

$$\mathbf{x}_i^{\text{aligned}} = T \cdot \mathbf{x}_i. \tag{2.30}$$

- ***Regularization term (optional):*** To encourage the predicted matrix T to be close to a valid rotation matrix and avoid degenerate solutions, a regularization loss such as orthogonality loss may be added during training

$$\mathcal{L}_{\text{reg}} = \|TT^T - I\|_F^2, \tag{2.31}$$

where I is the identity matrix and $\|\cdot\|_F$ denotes the Frobenius norm.

T-Net is typically used in two places: (1) to align raw input point clouds, and (2) to align intermediate feature representations. In both cases, the goal is to make the model more robust to transformations in the data, such as varying orientations or scaling. This property is particularly valuable in registration tasks, where misaligned coordinate frames can lead to poor correspondence matching. T-Net serves as a learnable pre-alignment module that improves generalization and invariance in downstream tasks involving unordered 3D point sets.

Table 2.1. Overview of some learning-based non-rigid point cloud registration methods based on Transformers and DGCNN

Methods	Year	Network Architecture	Robustness	Experimental Data
PPFNet [134], PPF-FoldNet [135]	2018	PointNet [112], MLP	Partial, Noise	3DMatch [136], Synthesis [137], 7-Scenes [138], SUN3D [139], RGB-DScenesv.2 [140], SpinImages [141], SHOT [142], FPFH [38],USC [143]
[108]	2019	DGCNN	Noise, outliers	Medical dataset
PRNet [144]	2019	DGCNN, Transformer	Noise, partial	ShapeNetCore [145], ModelNet40 [92]
[104]	2019	EdgeConv, CNN	Partial	ShapNet [145]
3DSmoothNet [146]	2019	CNN	Partial, Noise	3DMatch [136], ETH Dataset [147]
DispVoxNets [101]	2019	CNN	Noise, deformation levels, outlier	FLAME [148], Dynamic FAUST (DFAUST) [149], cloth [150]
CoFiNet [124]	2021	Transformer	Outlier, partial	odometryKITTI [151], 3Dmatch [136], 3DLoMatch [152]
[107]	2021	DGCNN	-	Medical dataset
NrtNet [109]	2022	DGCNN, Transformer	Deformation levels	SURREAL [153], SHREC'19 [154], MIT [155]
[21]	2022	Transformer	Outlier, partial	odometryKITTI [151], 3Dmatch [136], 3DLoMatch [152]
[127]	2022	Transformer	Outlier, partial	3Dmatch [136], 3DLoMatch [152], ModelNet [92]
Lepard [125]	2022	Transformer	Outlier, partial, deformation levels	3Dmatch [136], 3DLoMatch [152], 4DMatch [125]
OIF-PCR [126]	2022	Transformer	Outlier, partial	odometryKITTI [151], 3Dmatch [136], 3DLoMatch [152]
[156]	2022	DGCNN	Noise, outlier, partial, data density variation, deformation levels	ModelNet [92], TOSCA [157], Human motion [158]
SyNoRiM [111]	2022	CNN	Noise, partial, outlier, data density variation	[159], Clothcap [160], 4dcomplete [161], Deepdeform [162], SAPIEN [163]
NDP [117]	2022	MLP	Outlier, partial, deformation levels	4DMatch [125]
GraphSCNet [20]	2023	GCNN	Outliers, deformation levels, partial	4DMatch [125], CAPE [160], DeepDeform [162]
NIE [110]	2023	DGCNN	Noise, partial	SURREAL [153], FAUST [164], SCAPE [165]
MAFNet [166]	2024	Transformer	Noise, partial	7-Scenes [138], ModelNet [92]

Chapter 3

Material and Methods

This chapter presents the methodologies employed in this study, along with the generated datasets and published simulation tools, organized into Material (Section 3.1) and Methods (Section 3.2). The Material section introduces the resources used, including SimTool [23], a toolset for soft-body simulation based on NVIDIA Flex and Unreal Engine, as well as two datasets developed as part of this work: The synthetic soft tissue dataset SynBench [27] and the real-world dataset DeformedTissue [25, 26], both introduced in this dissertation. In addition, the configuration of two widely used benchmark datasets, ModelNet [92] and 4DMatch [125], is presented.

The Methods section outlines the proposed approaches, Robust-DefReg [27] and Def-TransNet [22], along with their iterative refinement strategies. These network architectures are specifically designed to evaluate and improve non-rigid PCR in the context of soft tissue deformation.

The author of this dissertation contributed to the development of all tools, datasets, and methods mentioned above, including the SimTool framework [23], the SynBench [27] and DeformedTissue [25, 26] datasets, the registration methods Robust-DefReg [27] and Def-TransNet [22], and the iterative refinement framework integrated into both methods.

3.1 Material

In recent years, the generation of synthetic datasets for evaluating computer vision methods has gained significant traction, primarily because such datasets can provide accurate ground truth data. This is particularly valuable in domains like surgical procedures, where collecting real-world data is often challenging. In this section, we present a simulation toolset called SimTool [23] and a synthetic benchmark dataset named SynBench [27]. SimTool is designed to simulate soft tissue deformation during resection surgery, and SynBench is a dataset generated using this tool. In addition to SynBench, a real-world dataset named

DeformedTissue has also been created and published [25, 26].

3.1.1 SimTool: Soft Body Simulation

Simulation tools for soft body deformation are essential for developing and benchmarking non-rigid PCR methods, particularly when real data is limited or lacks ground truth. A suitable simulation framework for this purpose must fulfill several key requirements:

- ***Customizability:*** The tool should support flexible control over object shapes, deformation parameters, and material properties.
- ***Physical realism:*** It must produce plausible non-rigid deformations that approximate real-world soft tissue or elastic material behavior.
- ***Surface-level output:*** The simulation should allow extraction of high-quality surface point clouds from 3D scenes.
- ***Ground truth availability:*** It must support tracking and output of ground truth correspondences for evaluation.
- ***Open-source accessibility:*** The tool should be publicly available to support reproducibility and benchmarking.

While various soft body simulators exist (e.g., for computer graphics and gaming [167–169]), they are often application-specific, closed-source, or lack the necessary support for data generation and control needed for machine learning research in non-rigid PCR. To address this gap, we developed SimTool, an open-source simulation framework designed to generate realistic and controllable soft body deformations for use in registration benchmarks and learning-based model training. SimTool satisfies the above requirements through a hybrid integration of:

- ***3D modeling and mesh processing:*** This module is responsible for generating and manipulating the geometric models used in simulation. It supports both procedurally created shapes and pre-defined anatomical or synthetic models. Basic operations such as mesh smoothing, resampling, slicing, and remeshing are included, allowing the user to control surface complexity, resolution, and object topology prior to deformation.
- ***Physically-based soft body simulation:*** SimTool simulates realistic non-rigid behavior using a position-based dynamics framework. Objects can undergo a wide range of

deformations, including compression, stretching, bending, twisting, or cutting. Material properties such as elasticity, mass, damping, and structural constraints are configurable, enabling the simulation of soft tissues, elastic components, or other flexible materials under physical interaction or external forces.

- ***Surface capture and point cloud generation:*** Following deformation, the object's surface is sampled to produce point clouds. These can be extracted under varying conditions, such as different viewpoints, occlusions, or simulated sensor settings. The tool supports exporting both the deformed point cloud and the ground truth correspondence or deformation field, enabling supervised evaluation of registration accuracy under controlled deformation levels.

The SimTool workflow consists of three main stages: Random shape generation, deformation and slicing, and surface capturing. These components are described in detail in our peer-reviewed publication [23]. They will not be elaborated upon in this thesis in order to maintain consistency and integration with the current document, and because this is not the primary focus of the present work.

3.1.2 SynBench and DeformedTissue Datasets

Evaluating a PCR method typically requires testing under multiple scenarios to assess the robustness and generalization capabilities of the proposed approach. A comprehensive evaluation of non-rigid PCR demands datasets that incorporate key challenges such as large deformations, noise, outliers, and incompleteness. Although several datasets exist for deformable PCR, none provide a complete benchmark that encompasses all these challenges, making fair comparison across different methods difficult.

In the following, our proposed datasets will be introduced: (1) SynBench, a synthetic soft tissue dataset, and (2) DeformedTissue, a real-world soft tissue dataset.

1) SynBench: Synthetic Soft Tissue Dataset

In the previous section, we introduced SimTool [23], a toolbox designed to simulate soft body deformation and generate deformable point clouds. In this work, SimTool is utilized to create a benchmark for non-rigid PCR, named SynBench, which serves as an evaluation framework for PCR methods.

The following sections outline the dataset development process, highlighting its adaptability for various applications. A more detailed discussion on dataset generation can be found in our published paper [24] and the link to download the dataset is [170]. To emphasize the novelty of our proposed dataset, its main contributions are summarized below:

- ***A benchmark of customizable objects.*** Although the dataset is generated using soft body deformation simulations, it can be adapted for broader applications. Unlike datasets with predefined objects, such as animals or human bodies, SynBench provides a flexible framework for training machine learning models on any non-rigid object.

- ***Challenges and robustness.*** The dataset introduces various challenges in PCR, including different deformation levels, varying noise intensities, outlier ratios, and data incompleteness. This comprehensive design facilitates an effective assessment of method robustness under diverse conditions.

- ***Ground truth for corresponding points.*** SynBench includes ground truth correspondences for both pre- and post-deformation objects and slices. This information is valuable not only for evaluating registration accuracy but also for other 3D point cloud applications.

The proposed dataset is derived from 30 primitive objects generated using SimTool. These objects undergo different deformation levels to create diverse test scenarios. Additionally, the dataset includes challenges such as varying outlier ratios, noise levels, and data incompleteness, along with ground truth point correspondences.

Since all challenges are generated under small to large deformation levels, SynBench allows users to select subsets based on their method's capabilities and assess robustness to complex scenarios. The dataset comprises five main subsets: "Data," representing the 30 primitive objects, and four challenge categories, "Deformation Level," "Incompleteness," "Noise," and "Outlier", containing 5,297, 26,485, 21,188, and 26,485 object samples, respectively. Each sample consists of a source and target point cloud pair. The higher file count in certain challenge categories arises from applying each challenge across multiple deformation levels while varying key parameters. These parameters for each challenge are discussed in the following sections.

Different deformation levels. To simulate varying deformation levels in a controlled and reproducible way, two complementary strategies are implemented. Each is chosen with respect to different requirements, physical realism in one case and mathematical control in the other.

The first strategy involves physically simulated deformation using a soft body physics engine. By varying gravitational force and simulation time, a range of deformation intensities is applied to 30 primitive objects, resulting in a dataset that reflects real-world object behavior under physical stress [23]. This physically grounded approach is particularly valuable when simulating realistic, plausible interactions such as those in surgical or mechanical

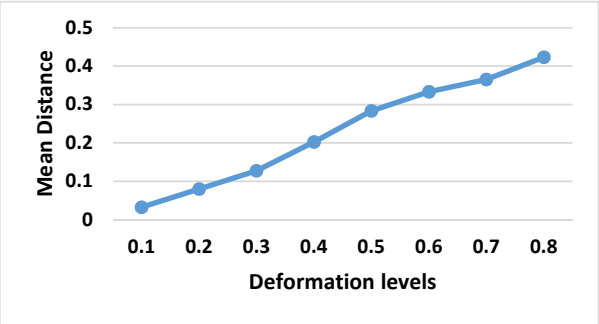


Figure 3.1. The average mean distance for each deformation level. As deformation levels increase, the mean distance between source and target point clouds rises, indicating greater deformation [27].

environments. However, due to limited control over exact deformation patterns and the lack of point-to-point ground truth, an alternative strategy is introduced.

The second strategy is based on Thin-Plate Splines (TPS) [68], a kernel-based interpolation method widely used in non-rigid registration. TPS was selected for its ability to generate smooth, globally coherent deformations by separating affine and non-affine components. Unlike spectral or eigenfunction-based methods, such as those relying on Laplacian eigenmaps or eigenfaces, TPS operates directly on point sets and does not require mesh topology or consistent connectivity. This makes it especially suitable for use with unstructured 3D point clouds.

The TPS deformation of a point x is defined as

$$f(x) = A \cdot x + b + \sum_{i=1}^{k} w_i \cdot \phi(\|x - c_i\|), \tag{3.1}$$

where $x \in \mathbb{R}^3$ is a point in the source point cloud, $A \in \mathbb{R}^{3\times3}$ is the affine transformation matrix, $b \in \mathbb{R}^3$ is the translation vector, $c_i \in \mathbb{R}^3$ are the control points, $w_i \in \mathbb{R}^3$ are the weights associated with each control point, and $\phi(r)$ is the radial basis function defined as

$$\phi(r) = r^2 \log(r), \tag{3.2}$$

with $r = \|x - c_i\|$ denoting the Euclidean distance between point x and control point c_i. This specific RBF was chosen because it minimizes the bending energy of the deformation field, ensuring smooth transitions and natural-looking deformations.

To generate different deformation levels, we apply controlled displacements to the control points c_i. The control points are initially sampled uniformly across the bounding region of the object. Each control point is then perturbed by a displacement vector drawn from a zero-mean isotropic Gaussian distribution

$$c_i' = c_i + \mathcal{N}(0, \sigma^2), \tag{3.3}$$

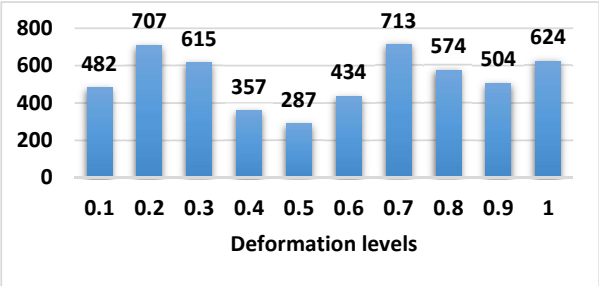


Figure 3.2. The number of source and target point cloud pairs at different deformation levels [27].

where σ is the standard deviation that controls the magnitude of the deformation. This stochastic perturbation is not intended to simulate measurement noise, but rather to introduce controlled variability into the deformation field. The choice of a Gaussian distribution provides spatially unbiased, symmetric perturbations around each control point, with a single parameter (σ) that allows for continuous adjustment of deformation strength. Alternative distributions such as uniform or Poisson were considered, but Gaussian was preferred due to its continuous support and natural ability to localize deformation while preserving global smoothness. Uniform distributions do not concentrate around the origin and are less appropriate for localized, smooth shifts, while Poisson distributions are generally used for modeling discrete events rather than continuous vector fields.

The final deformation field is then computed by evaluating the TPS transformation based on the displaced control points c_i'. The resulting transformation includes: Affine and non-affine. The affine component is define as

$$A \cdot x + b, \tag{3.4}$$

where A governs linear transformations such as rotation, scaling, and translation, while b represents a translation vector. The non-affine component, responsible for non-linear deformations, is given by

$$\sum_{i=1}^{k} w_i \cdot \phi(\|x - c_i\|), \tag{3.5}$$

which models deformations using radial basis functions.

Three key factors influence the severity and complexity of the resulting deformation:

1. The number of control points k: More control points allow finer-grained deformation.
2. The standard deviation σ: Governs the displacement magnitude of control points.
3. The spatial spread of the radial basis function $\phi(r)$: Larger distances r reduce influence, producing localized deformation.

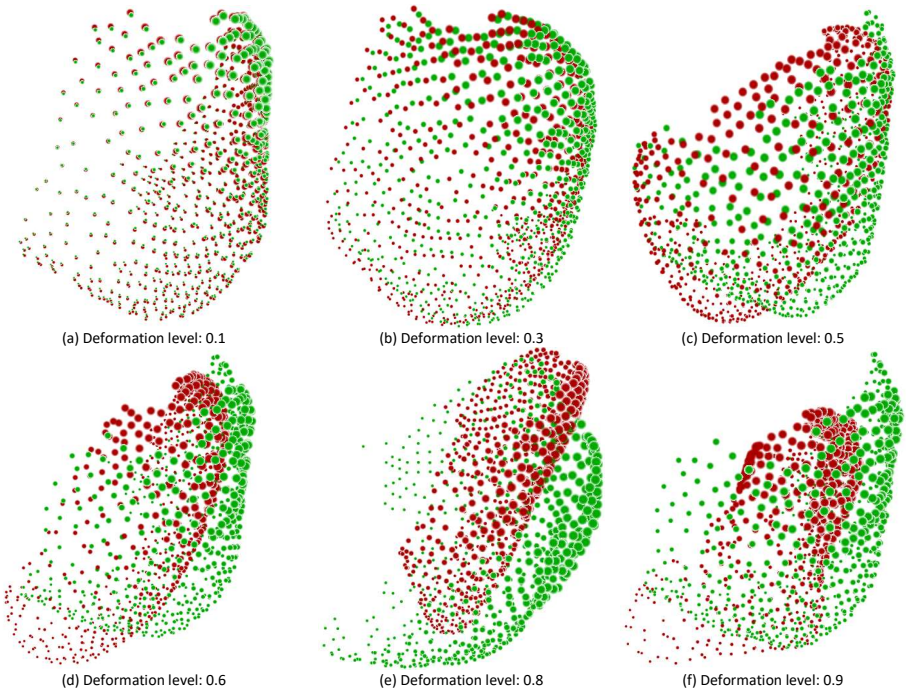


Figure 3.3. Visualization of the proposed SynBench dataset used in this study. Subfigures (a), (b), (c), (d), (e), and (f) depict different deformation levels (0.1, 0.3, 0.5, 0.6, 0.8, and 0.9) within the SynBench dataset. In these point cloud visualizations, the source point clouds are shown in green, while the target point clouds are shown in red, illustrating varying degrees of deformation [27].

No explicit constraints are imposed on the topology of the object. The TPS-based deformation operates on unstructured point sets and does not require connectivity or mesh information. This topology-agnostic property is essential for PCR scenarios where the object geometry may be irregular or incomplete.

To quantitatively measure deformation levels, we compute the average Euclidean distance between corresponding source and target points. Let $X = \{x_1, x_2, \ldots, x_n\} \subset \mathbb{R}^3$ and $Y = \{y_1, y_2, \ldots, y_n\} \subset \mathbb{R}^3$ denote the original and deformed point clouds, respectively. The distance between corresponding points is

$$d(x_i, y_i) = \sqrt{(x_i^1 - y_i^1)^2 + (x_i^2 - y_i^2)^2 + (x_i^3 - y_i^3)^2}, \tag{3.6}$$

and the mean deformation level is defined as

$$D_{\text{mean}} = \frac{1}{n} \sum_{i=1}^{n} d(x_i, y_i) = \frac{1}{n} \sum_{i=1}^{n} \sqrt{(x_i^1 - y_i^1)^2 + (x_i^2 - y_i^2)^2 + (x_i^3 - y_i^3)^2}. \tag{3.7}$$

Visualizations of point clouds at different deformation levels are shown in Figure 3.3. Corresponding quantitative statistics of mean distance and dataset distribution are presented

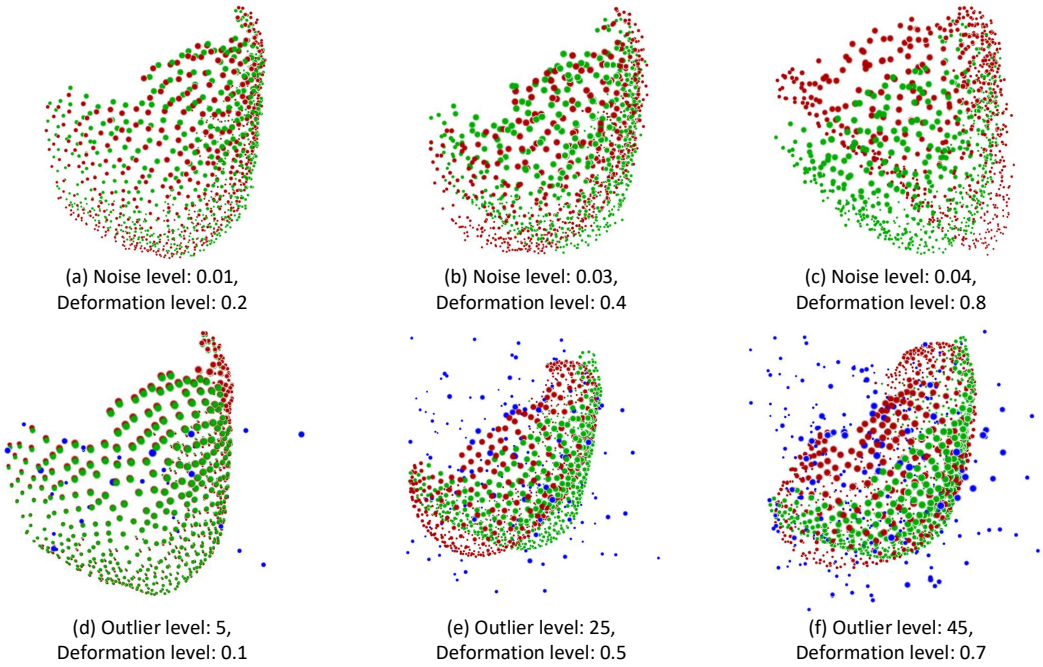


Figure 3.4. Visualization of the SynBench (synthetic) datasets under different noise and outlier levels. Sub-figures (a), (b), and (c) illustrate varying deformation and noise levels within the SynBench dataset, while sub-figures (d), (e), and (f) depict different deformation and outlier levels. In the point cloud visualizations, the source point clouds are displayed in green, and the target point clouds are displayed in red, representing varying degrees of deformation [27]. Outliers are shown as blue points.

in Figures 3.1 and 3.2, respectively.

Different levels of noise. Evaluating the robustness of methods to noise is a common practice in the literature [7]. Since the SynBench dataset is synthetically generated, the initial point clouds are noise-free. To better simulate real-world scenarios, varying levels of synthetic noise are introduced. In this study, Gaussian noise is added to the point sets with zero mean and varying standard deviations. Gaussian noise, commonly used in research, follows a normal distribution $\mathcal{N}(0, \sigma^2)$, where $\mu = 0$ ensures the noise is centered around zero, and σ determines its magnitude. Larger values of σ result in noisier data. Reported values in the literature typically range between 0.01 and 0.04, representing small to large noise levels [114, 171]. Based on this, four noise categories are generated in the dataset: $\sigma = 0.01$, $\sigma = 0.02$, $\sigma = 0.03$, $\sigma = 0.04$.

For each point $x_i = (x_i^1, x_i^2, x_i^3)$ in the original noise-free source point cloud X, a noisy point x_i' is generated as

$$x_i' = x_i + \mathcal{N}(0, \sigma^2), \tag{3.8}$$

where Gaussian noise is independently added to each coordinate of x_i. The noisy dataset categories, corresponding to different noise levels, are illustrated in Figure 3.4.

Different levels of outliers. Outliers pose a significant challenge in point cloud processing, often arising from sensor inaccuracies, environmental disturbances, or misalignment during data acquisition. In this context, an outlier is defined as a point that deviates substantially from the spatial distribution of the main object surface.

To systematically evaluate robustness under varying levels of contamination, synthetic outliers are introduced into the dataset. These outlier points are generated by sampling from a Gaussian distribution $\mathcal{N}(\mu, \sigma^2)$, where $\mu \in \mathbb{R}^3$ specifies the mean location relative to the main point cloud, and σ controls the spatial spread of the outliers. By adjusting μ, outliers can be positioned at varying distances from the object surface, while larger values of σ produce more dispersed outlier patterns.

The number of outliers added to each point cloud is determined as a proportion of the original point set size. Inspired by prior work [172, 173], we generate five variants of each dataset, corresponding to outlier levels of 5%, 15%, 25%, 35%, and 45%.

Let $X = \{x_1, x_2, \ldots, x_n\}$ denote the original point cloud. The augmented version X' is constructed by appending $m = \frac{p}{100} \cdot n$ outlier points, sampled independently from the specified Gaussian distribution. The resulting point cloud size is given by

$$|X'| = n + \frac{p}{100} \cdot n, \tag{3.9}$$

where $p \in \{5, 15, 25, 35, 45\}$ denotes the outlier percentage. Each outlier point o_j is sampled as

$$o_j \sim \mathcal{N}(\mu, \sigma^2). \tag{3.10}$$

This procedure enables controlled generation of varying outlier densities and spatial distributions, allowing systematic evaluation of the robustness of registration methods. It is important to note that the choice of outlier magnitude and density may vary depending on application-specific constraints or performance benchmarks under investigation.

2) DeformedTissue: Real-world Soft Tissue Dataset

Tissue deformation, also known as tissue shift, is a significant challenge in soft-tissue surgeries. It results in the displacement of anatomical landmarks, complicating navigation within soft tissues. Such deformations occur after the surgical opening due to the release of tension, changes in patient positioning, or removal of tissue, and are influenced by the tissue's texture and shape. This phenomenon has been well documented in neurosurgery [174]. However, tissue displacement is also a concern in head and neck surgeries, particularly during tumor resections, where vital structures lie close together and must be preserved.

Following tumor removal, the pathological TNM classification plays a key role in guid-

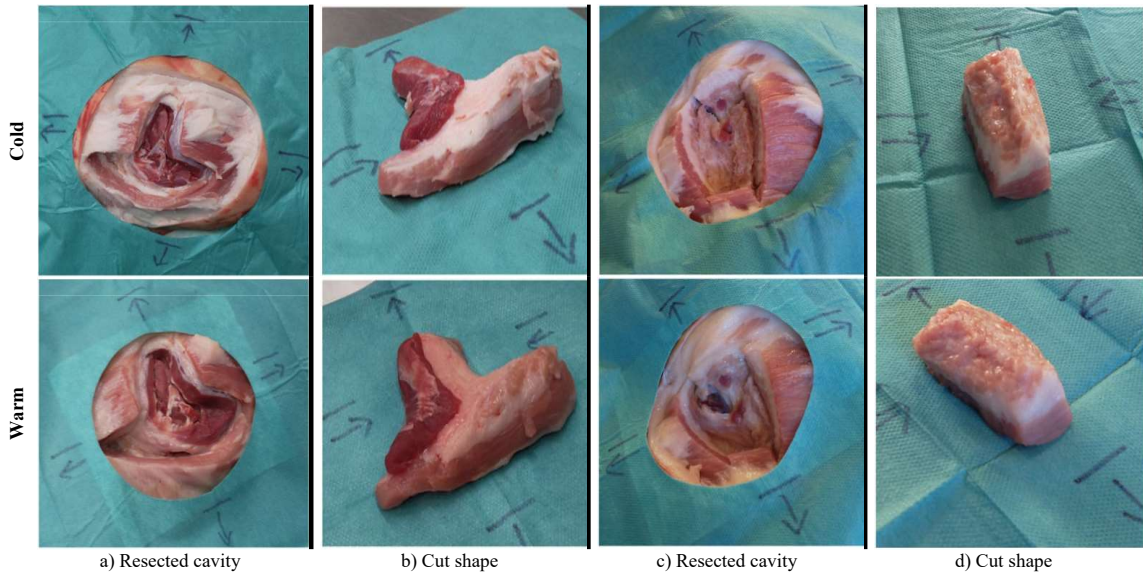
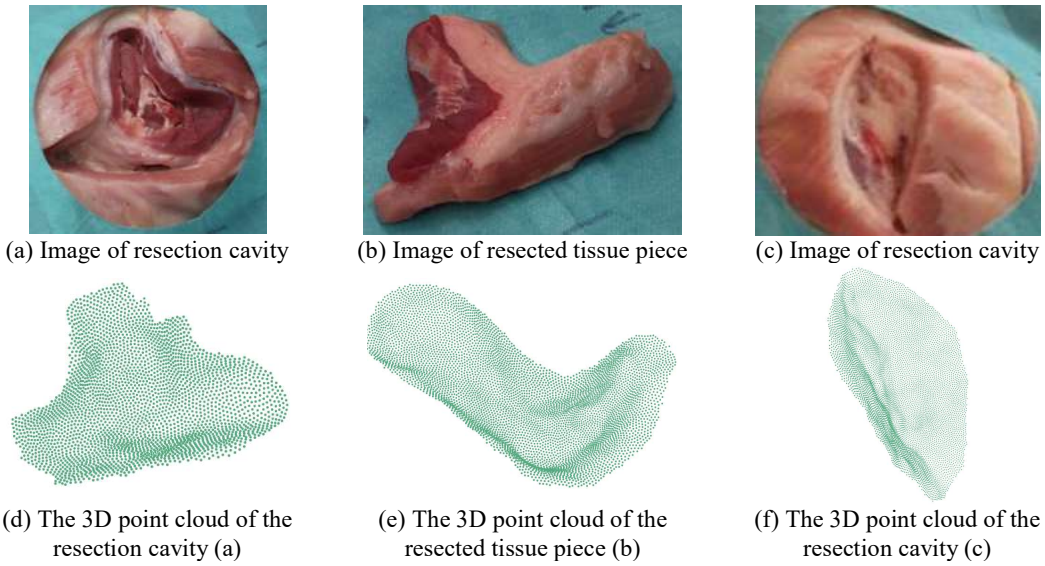


Figure 3.5. 3D camera and head-mounted display images of cut tissue shapes (CTs) and their corresponding resection cavities (RCs). Anatomical directions, cranial, caudal, rostral, and occipital, are indicated with arrows and lines to guide 3D imaging and photogrammetry. (a) T-shape (branched geometry), resection cavity; (b) T-shape, excised tissue piece; (c) Rectangular shape (compact geometry), resection cavity; (d) Rectangular shape, excised tissue piece. Tissue temperatures: Cold = $7.91 \pm 4.1\,^{\circ}\mathrm{C}$, warm = $36.37 \pm 1.28\,^{\circ}\mathrm{C}$. [26].

ing treatment decisions and predicting outcomes. Accurately determining the pathological T stage requires pathological examination of the tumor, which becomes more complex when tissue is malformed. Variations in tumor shape and tissue displacement further complicate the assessment of frozen sections for both surgeons and pathologists [26].

This experimental study utilized 45 pig head cadavers (Schradi Frischfleisch GmbH, Mannheim, Germany) and was approved by the Mannheim Veterinary Office (DE 08 222 1019 21). The use of cadavers enabled the generation of large datasets suitable for training deep learning models. In contrast, real tumor specimens are scarce and not reproducible, which motivated the cadaver-based approach. To capture tissue morphology before and after controlled heat-induced deformation, 3D cameras and head-mounted displays were employed. Examples of the tissue images, both pre- and post-heating, are shown in Figure 3.5. The data were further processed using tools such as Meshroom, MeshLab, and Blender to produce and analyze 2½D mesh models. The outcomes of this study have been published in peer-reviewed journals, and additional technical details can be found in [25, 26].

After generating the natural deformations, we adopted the same methodology as in the SynBench dataset to construct a large-scale dataset for training and evaluating neural networks. It includes 5,126 samples for the ”Deformation Level” challenge (ranging from 0.1 to 0.7), 20,504 for ”Noise”, and 25,630 for ”Outlier”, ensuring compatibility with SynBench for comparative evaluation. Example images captured using HoloLens 2 and ArtecEva,



(a) Image of resection cavity (b) Image of resected tissue piece (c) Image of resection cavity

(d) The 3D point cloud of the resection cavity (a) (e) The 3D point cloud of the resected tissue piece (b) (f) The 3D point cloud of the resection cavity (c)

Figure 3.6. Head-mounted display images of resection cavities and cut tissue pieces are shown in the first row and the second row demonstrates the extracted 3D point cloud. [22].

along with the corresponding point clouds, are shown in Figure 3.6. The dataset is publicly available at [175].

3.1.3 ModelNet10 Dataset

ModelNet10 [92] is a widely used benchmark dataset consisting of clean, synthetic 3D CAD models. It is a subset of the larger ModelNet40 collection and includes 4,899 shapes across 10 object categories, such as bathtub, bed, chair, desk, dresser, monitor, nightstand, sofa, table, and toilet. Each object is pre-aligned to a canonical orientation, which facilitates comparison across registration methods. The dataset is split into 3,991 training samples (80%) and 908 test samples (20%). To simulate more realistic variations during training, we apply random rotations of up to 45 degrees around the z-axis to the point clouds. ModelNet10 is particularly useful for evaluating registration methods under clean, rigid transformations, serving as a baseline before introducing more complex non-rigid scenarios. To extend this baseline to non-rigid cases, we adapted the ModelNet10 dataset by introducing varying levels of deformation, following the same procedure described in the previous section for generating our SynBench dataset. The dataset is available at: [92].

3.1.4 4DMatch/4DLoMatch Dataset

4DMatch [125] is a dynamic point cloud dataset derived from the 4DComplete dataset [161], designed to benchmark registration and correspondence estimation in scenes undergoing temporal geometric changes. It includes dense ground-truth correspondences between par-

tial scans, with a low-overlap variant referred to as 4DLoMatch.

Each sample includes a source point cloud X, deformation array D, a target point cloud Y, a rotation matrix R, a translation vector t, an overlap ratio, and point correspondences. To align with our method's requirements and maintain consistency, the target point cloud is regenerated using the transformation

$$Y = t^T + (X + D)R^T \tag{3.11}$$

This transformation is applied only to points with valid correspondences, ensuring the structural integrity of overlapping regions. Point clouds are resampled to have equal cardinality, which is necessary for our method. Based on overlap ratios, samples are categorized into 4DMatch (> 0.45) and 4DLoMatch (< 0.45). The dataset contains 47,738 training pairs, 6,400 validation pairs, and a test set of 10,327 4DMatch and 4,590 4DLoMatch samples. Link to download the dataset: [125].

3.2 Methods

In this chapter, we present the proposed methods for non-rigid PCR: Robust-DefReg [27], a graph-based coarse-to-fine registration network; DefTransNet [22], a Transformer-based architecture designed to handle complex deformations; and Learning-to-Refine, an iterative refinement strategy that improves registration accuracy across architectures. Each approach is designed to address a specific challenge identified in the research questions (RQ2–RQ4) and to validate the corresponding hypotheses (H2–H4) (see Section 1.2).

To address RQ2 and test H2, we introduce Robust-DefReg [27], a graph-based non-rigid registration method built on a coarse-to-fine strategy. It encodes local geometric relationships by constructing a graph over the point cloud and applying graph convolutional layers to learn neighborhood-aware features. The model integrates a spatial transformer network (T-Net) for pre-alignment and employs a Loopy Belief Propagation (LBP) module to enforce local smoothness in the estimated displacement field. This design ensures resilience to deformation and spatial noise by explicitly leveraging local geometric structures, a critical factor in achieving robust registration under challenging conditions.

To address RQ3 and validate H3, we propose DefTransNet [22], a Transformer-based non-rigid PCR network that enhances Robust-DefReg by incorporating global attention mechanisms. While graph-based methods excel at modeling local geometry, they may struggle in regions with repetitive or ambiguous structures. DefTransNet mitigates this issue by combining EdgeConv layers for local feature encoding with a Transformer module that models long-range dependencies between source and target point clouds. This enables the

network to resolve feature ambiguity and establish correspondences in complex, non-local deformation scenarios. The architecture also includes a T-Net for affine pre-alignment, aiding in convergence and structural consistency.

To address RQ4 and test H4, we introduce Learning-to-Refine, an iterative refinement strategy applied atop both Robust-DefReg and DefTransNet. This model-agnostic framework improves registration by incorporating a probabilistic perspective into the learning process. Specifically, it refines displacement predictions over multiple iterations using a combined loss that includes geometric alignment and Kullback-Leibler (KL) divergence terms. The latter imposes a prior over the deformation distribution, enabling uncertainty modeling that enhances convergence stability and reduces overfitting. This approach is particularly beneficial in scenarios with ambiguous or incomplete correspondences, where deterministic single-pass models often fail to generalize.

3.2.1 Robust-DefReg: GCNN-Based Method

Accurate non-rigid PCR under large deformation and noise remains an open problem, especially in applications like soft tissue simulation, where local geometric variations and sparse correspondences introduce substantial complexity. Traditional methods often fall short due to the following limitations:

- ***Sensitivity to deformation:*** Voxel-based or pointwise methods often lose geometric details and cannot generalize well under non-rigid deformation.
- ***Vulnerability to noise and outliers:*** Sparse keypoint matching and global models can easily be disrupted by local noise or partial occlusions.
- ***Lack of local geometric awareness:*** Many registration pipelines ignore spatial relationships between neighboring points, leading to inconsistencies in the predicted deformation field.

To address these limitations, and to answer RQ2 and validate H2 (see Section 1.2), our method aims to:

- Explicitly model local geometric relationships using graph-based learning.
- Improve spatial coherence in the predicted displacement field via regularization over a neighborhood graph.
- Enhance robustness under deformation and noise through message-passing-based refinement.

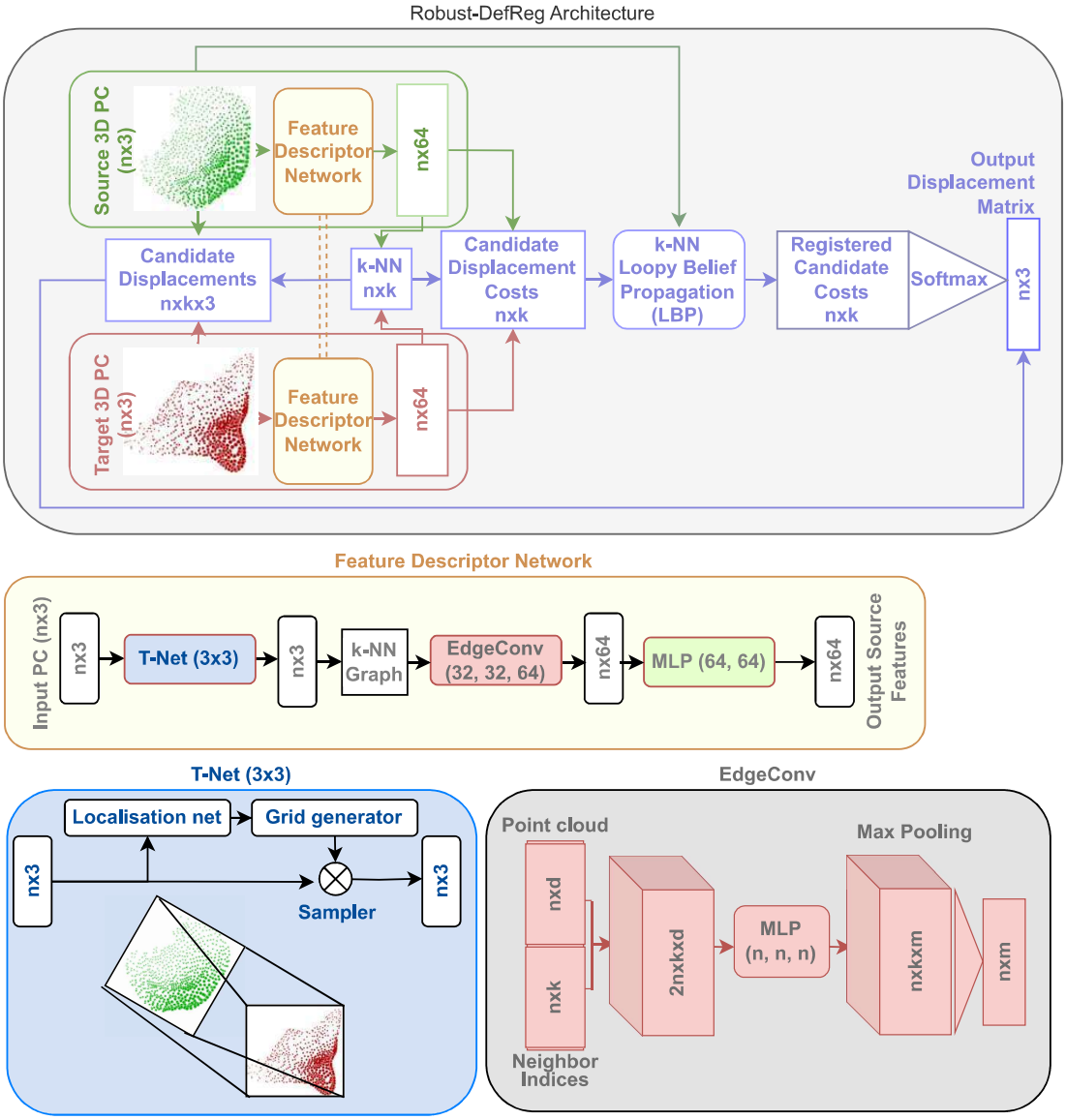


Figure 3.7. The proposed network architecture of Robust-DefReg [27].

We propose Robust-DefReg, a graph-based learning method that models local geometric structure via GCNNs and enforces spatial regularity using LBP. Considering $\mathbf{X}$ and $\mathbf{Y}$ as source and target point cloud, the goal is to estimate a continuous displacement field $D = \{\mathbf{d}_i\}_{i=1}^{N}$, where $\mathbf{d}_i \in \mathbb{R}^3$ aligns $\mathbf{x}_i$ with a corresponding point in $\mathbf{Y}$.

Rather than relying on hard point-to-point correspondences, our method leverages a soft, weighted approach to displacement estimation. For each source point $\mathbf{x}_i$, we identify k candidate correspondences $\{\mathbf{c}_i^p\}_{p=1}^{k}$ in the target cloud based on feature similarity. The unary cost of assigning candidate $\mathbf{c}_i^p$ to $\mathbf{x}_i$ is computed as

$$d_i^p = \|f(\mathbf{x}_i) - f(\mathbf{c}_i^p)\|_2^2, \tag{3.12}$$

where $f(\cdot)$ is the learned feature embedding function.

To enforce spatial coherence, a pairwise regularization cost is defined between two neighboring source points $(\mathbf{x}_i, \mathbf{x}_j) \in E$ and their respective displacement candidates.

$$r_{ij}^{pq} = \left\| (\mathbf{c}_i^p - \mathbf{x}_i) - (\mathbf{c}_j^q - \mathbf{x}_j) \right\|_2^2 . \tag{3.13}$$

This cost encourages neighboring points to undergo similar displacements, promoting a smooth deformation field.

These costs are integrated into a pairwise graphical model, and inference is performed using LBP. Each node (point) iteratively exchanges messages with its neighbors to refine its belief over candidate displacements. The message update rule at iteration t is

$$m_{i\to j}^t(q) = \min_p \left(d_i^p + \alpha \cdot r_{ij}^{pq} + \sum_{h \in \mathcal{N}(i)\setminus j} m_{h\to i}^{t-1}(p) \right), \tag{3.14}$$

where

- d_i^p is the unary cost from Eq. 3.12,
- r_{ij}^{pq} is the pairwise regularization term from Eq. 3.13,
- α controls the influence of regularization,
- $m_{h\to i}^{t-1}$ are messages from other neighbors at the previous iteration.

Messages are initialized to zero and updated for a fixed number of iterations. After convergence, the final belief distribution over candidates is normalized using Softmax

$$w_i^p = \frac{\exp(-d_i^p)}{\sum_q \exp(-d_i^q)}. \tag{3.15}$$

The final displacement vector $\mathbf{d}_i$ is then computed as the expected displacement

$$\mathbf{d}_i = \sum_{p=1}^{k} w_i^p \cdot (\mathbf{c}_i^p - \mathbf{x}_i). \tag{3.16}$$

Our model architecture consists of two main stages:

- ***Feature descriptor network:*** A shared T-Net [133] (see Section 2.4.5) aligns both input point clouds to a common canonical space, ensuring rotation invariance. Graphs are constructed using k-NN, and EdgeConv [104] (see Section 2.4.2) layers extract local shape-aware features. These features capture both positional and relational information essential for deformation-aware registration.

Table 3.1. Overview of the main processing steps in the Robust-DefReg method

- **Feature Descriptor Network.**
  - **Shared T-Net (3×3):** Both point clouds are initially aligned to a common reference frame using a shared T-Net, enhancing rotational invariance of the features.
  - **k-NN Graph Construction:** Graphs are built from the aligned point clouds by connecting each point to its k nearest neighbors.
  - **Shared EdgeConv Layers (32, 32, 64):** Multiple EdgeConv layers are applied to extract local geometric features from the constructed graphs.
  - **Shared MLP Layers (64, 64):** A shared multilayer perceptron comprising two 1D convolutional layers (each of size 64) with instance normalization between them refines the learned features.
- **Learning Displacement Field.**
  - **Feature Embedding:** Each point in the input point clouds is embedded into a 64-dimensional feature vector using the feature descriptor network.
  - **Neighborhood Matching:** For every point in the source point cloud, the k closest points in the target cloud are identified using the squared L2 distance between their feature vectors.
  - **Displacement Estimation:** Displacement vectors are computed from each source point to its selected neighboring target points.
  - **Cost Computation:** For each candidate displacement, a cost is computed based on the average squared Euclidean distance between the feature vectors.
  - **LBP over k-NN Graph:** A k-NN graph is formed for the source cloud, and Loopy Belief Propagation (LBP) is employed to iteratively refine the displacement costs through message passing.
  - **Softmax Aggregation:** The refined costs are passed through a Softmax layer to produce weights, which are used to compute the final displacement vector for each point as a weighted sum of its candidate displacements.

- ***Displacement field estimation:*** For each point, candidate displacements are estimated using feature similarity. Costs are computed as in Eq. 3.12 and refined through LBP using Eq. 3.14. Softmax aggregation (Eq. 3.16) produces the final deformation field.

This formulation emphasizes local geometric preservation and spatially-consistent cost refinement through graph-based message passing, enabling more robust alignment under deformation and noise. An overview of the method is shown in Figure 3.7, and Table 3.1 further illustrates the proposed pipeline. Inspired by [107], our method incorporates a novel feature descriptor network to enhance robustness without compromising accuracy or computational efficiency. While [107] focuses on registering key points in deformed lungs, we build on this foundation by extending Robust-DefReg into a general-purpose registration framework applicable to various point cloud types, beyond medical data. Additionally, our approach is designed to ensure robustness to rotation and other structural challenges.

3.2.2 DefTransNet: Transformer-Based Method

While Robust-DefReg leverages local geometric features to achieve robust registration under noise and deformation, it is inherently limited in resolving feature ambiguity, a situation where geometrically similar regions lead to incorrect correspondences. These limitations become particularly pronounced in scenarios involving symmetric objects, repetitive structures, or partial overlap. To overcome this, we introduce DefTransNet, which integrates global contextual learning through a Transformer-based joint embedding. The following key limitations motivate our design:

- ***Ambiguous local descriptors:*** Local similarity across different regions can lead to incorrect matches.
- ***Limited long-range learning:*** Graph convolutions operate on neighborhood graphs and lack the capacity to model cross-cloud interactions.
- ***Independent encoding:*** Treating source and target separately restricts the ability to align semantically corresponding structures.

To address these limitations and answer RQ3 and validate H3 (see Section 1.2), our method aims to:

- Improve contextual distinctiveness by encoding long-range spatial dependencies using Transformer attention.
- Jointly model source and target representations to disambiguate structurally similar regions.
- Enhance global consistency and robustness to deformation, partial visibility, and noise through sequence-level embeddings.

We propose DefTransNet, a Transformer-based framework for non-rigid PCR. Given input point clouds $\mathbf{X}$ and $\mathbf{Y}$, similar to Robust-DefReg, our goal is to estimate a dense displacement field $D = \{\mathbf{d}_i\}_{i=1}^{N}$ that aligns $\mathbf{X}$ with $\mathbf{Y}$.

We begin with a shared T-Net transformation (see Section 2.4.5) and EdgeConv (see Section 2.4.2) layers for initial alignment and local feature encoding, respectively. The extracted features are then passed into a Transformer encoder-decoder that jointly embeds and updates both source and target point clouds. To enhance the learned features with long-range dependencies and contextual information, we integrate a Transformer encoder-decoder structure. This is a major addition compared to Robust-DefReg and enables DefTransNet to reason jointly over source and target point clouds in a unified feature space. Unlike methods that treat source and target clouds independently, we model them together

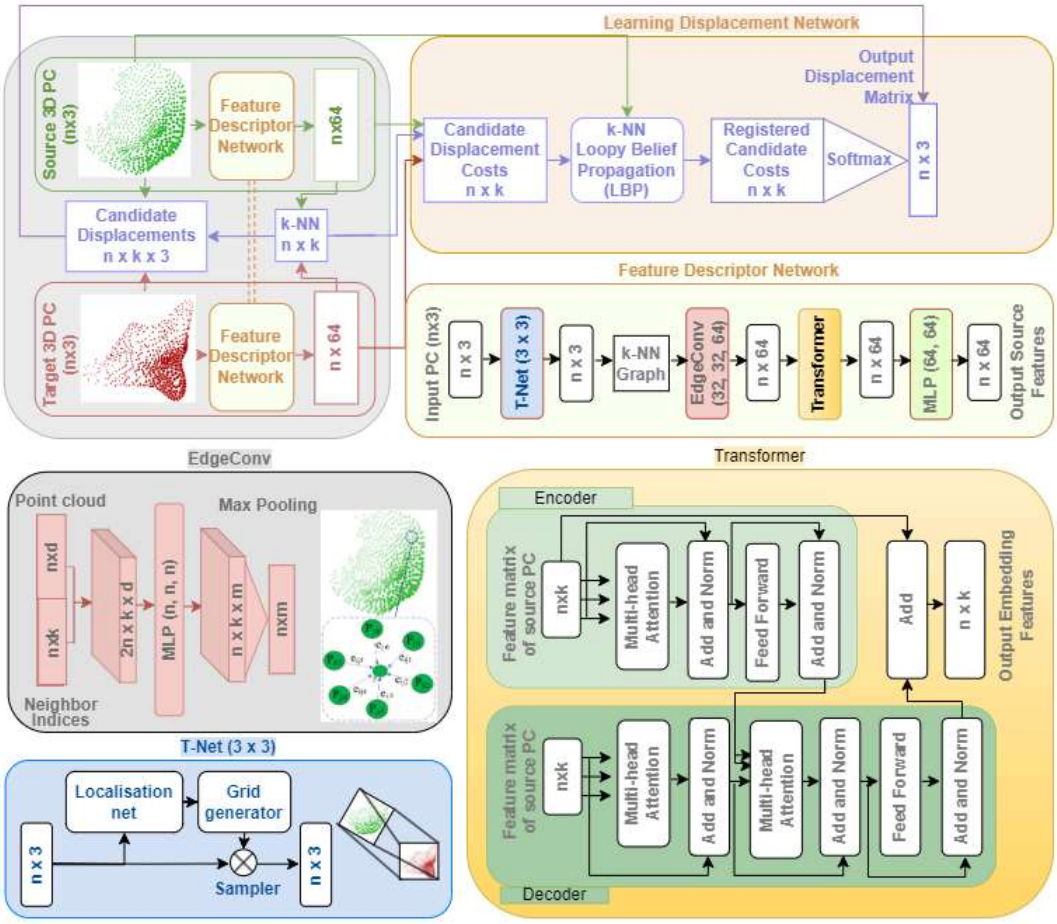


Figure 3.8. The proposed network architecture of DefTransNet [22].

using an encoder-decoder Transformer. Both the encoder and decoder are built from identical blocks composed of multi-head self-attention and feedforward networks. Each block contains multi-head self-attention and a feedforward network. The self-attention mechanism evaluates inter-point similarity via

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V, \tag{3.17}$$

where Q, K, and V are linear projections of the input features and d_k is the dimension of the keys. Multi-head attention expands this representation

$$\text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_1, ..., \text{head}_h)W^O, \tag{3.18}$$

$$\text{head}_i = \text{Attention}(QW_i^Q, KW_i^K, VW_i^V). \tag{3.19}$$

The decoder applies cross-attention to enable feature exchange between clouds. Cross-attention in the decoder enables the model to match features between clouds, improving

correspondence estimation. The resulting features are updated symmetrically

$$\begin{aligned} \Phi_{\mathcal{X}} &= \mathcal{F}_{\mathcal{X}} + \phi(\mathcal{F}_{\mathcal{X}}, \mathcal{F}_{\mathcal{Y}}), \\ \Phi_{\mathcal{Y}} &= \mathcal{F}_{\mathcal{Y}} + \phi(\mathcal{F}_{\mathcal{Y}}, \mathcal{F}_{\mathcal{X}}), \end{aligned} \tag{3.20}$$

where $\phi(\cdot)$ represents the cross-attention transformation that resolves ambiguity by considering mutual context.

To compute the displacement field, the enhanced features are passed into a regression module. For each source point $\mathbf{x}_i$, we find its k-nearest candidate displacements $\mathbf{c}_i^p$ in the target point cloud and evaluate their costs

$$d^p i = |f(\mathbf{x}i) - f(\mathbf{c}_i^p)|_2^2 . \tag{3.21}$$

A spatial regularization term ensures smoothness between neighboring displacements

$$r_{ij}^{pq} = |(\mathbf{c}i^p - \mathbf{x}i) - (\mathbf{c}j^q - \mathbf{x}j)|_2^2 . \tag{3.22}$$

Following the approach of Robust-DefReg, we use Loopy Belief Propagation (LBP) to iteratively refine the displacement beliefs. The message-passing formulation propagates updates across neighboring points. For details on this inference step, we refer the reader to Section 3.2.1. The final displacements are computed by applying a softmax weighting over the refined candidate scores, yielding a smooth and accurate deformation field.

Our architecture is structured in three stages:

- ***Feature encoding:*** T-Net aligns inputs; EdgeConv extracts local descriptors.
- ***Contextual embedding:*** Transformer attention models long- and cross-cloud dependencies.
- ***Displacement estimation:*** LBP-refined costs are fused to predict smooth displacements.

This formulation addresses RQ3 and supports H3 by resolving feature ambiguity through globally informed, joint embedding, as demonstrated in Figure 3.8.

3.2.3 Learning-to-Refine: Iterative Refinement Approach

Despite the advances introduced in the preceding methods, Robust-DefReg for local geometric encoding and DefTransNet for long-range semantic learning, non-rigid PCR remains challenged by two persistent limitations:

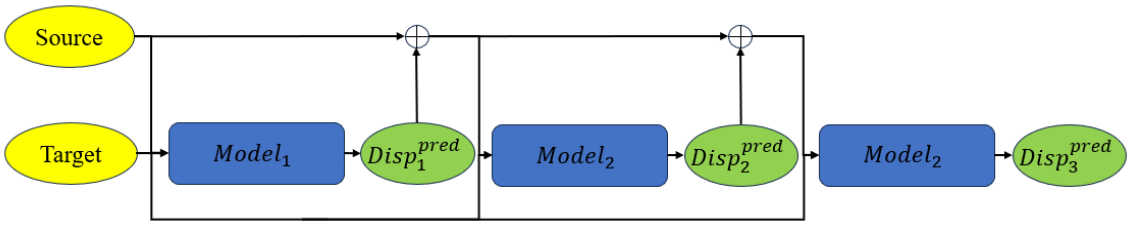


Figure 3.9. Iterative probabilistic refinement framework (Learning-to-Refine). The initial prediction is performed by Model_1, which can be instantiated by either DefTransNet or Robust-DefReg. The same model is reused as Model_2 in subsequent refinement iterations. From the second iteration onward, the loss function is extended with a KL divergence term to incorporate uncertainty modeling. Each pass predicts residual displacements that are added cumulatively to the previous estimation. This iterative self-training framework improves alignment progressively while penalizing implausible deformation fields.

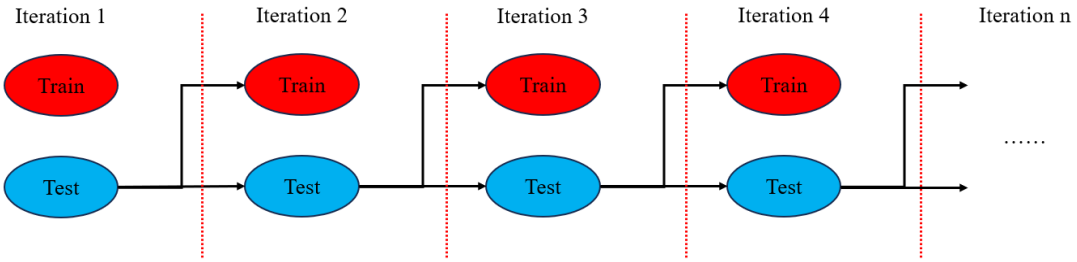


Figure 3.10. Progressive self-training with dynamic dataset splits across multiple iterations. In the first iteration, the model is trained on an initial training set and evaluated on a separate test set. After inference, a portion of the test set is pseudo-labeled and incorporated into the training set for the next iteration. The remaining half of the test set becomes the new test set. This cyclical process allows the model to progressively expand its training data while adapting to increasingly diverse and challenging samples.

- ***Residual misalignment.*** Even after a single forward pass, both methods often fail to fully resolve alignment in regions with high deformation, noise, or topological ambiguity. The output transformation is final and not self-correcting.

- ***Lack of uncertainty awareness.*** Predictions are generated deterministically, without accounting for ambiguity in the alignment. The models produce pointwise outputs without regard to the global distributional differences between source and target.

To address research question 4 (RQ4) and hypothesis 4 (H4) (see Section 1.2), we investigate whether regularizing the registration process through a probabilistic prior over deformation distributions leads to improved convergence and robustness compared to purely deterministic models. One natural strategy to introduce uncertainty is to learn a probabilistic displacement model, e.g., using Gaussian predictions with per-point variance and KL divergence regularization to a prior. However, this requires architectural changes and can lead

to instabilities when the model is overconfident in incorrect predictions. Another approach is to inject noise or dropout and use ensemble techniques to approximate uncertainty, but these are computationally expensive and do not enforce global consistency.

Instead, we propose a third approach: To treat the entire point cloud as a probability distribution and formulate the registration task as the alignment of two distributions in 3D space. This formulation naturally allows the use of distribution-level divergence measures, such as KL divergence, to regularize the output toward a globally consistent structure.

Considering $\mathbf{X}_k$ denotes the refined source point cloud at iteration k, and let $\mathbf{Y}$ denote the fixed target point cloud. We interpret both as empirical distributions over 3D space:

$$P(\mathbf{X}_k) = \frac{1}{N}\sum_{i=1}^{N}\delta(\mathbf{x}_{k,i}), \quad P(\mathbf{Y}) = \frac{1}{M}\sum_{j=1}^{M}\delta(\mathbf{y}_j), \tag{3.23}$$

where $\delta(\cdot)$ is a Dirac delta at each point location. We then seek to iteratively transform $\mathbf{X}_{k-1} \to \mathbf{X}_k$ such that:

1. The transformed source aligns closely with the target, as measured by a distance-based loss (e.g., Chamfer distance).

2. The global distribution of the source matches that of the target, as measured by a divergence $D(P(\mathbf{X}_k)\|P(\mathbf{Y}))$.

The loss function at each iteration k is defined as:

$$\mathcal{L}_k = \mathcal{L}_{\text{dist}}(\mathbf{X}_k, \mathbf{Y}) + \lambda_k \,\text{KL}(P(\mathbf{X}_k) \,\|\, P(\mathbf{Y})), \tag{3.24}$$

where $\mathcal{L}_{\text{dist}}$ is the symmetric Chamfer distance:

$$\mathcal{L}_{\text{dist}}(\mathbf{X}, \mathbf{Y}) = \frac{1}{N}\sum_{\mathbf{x}\in\mathbf{X}}\min_{\mathbf{y}\in\mathbf{Y}}\|\mathbf{x}-\mathbf{y}\|_2^2 + \frac{1}{M}\sum_{\mathbf{y}\in\mathbf{Y}}\min_{\mathbf{x}\in\mathbf{X}}\|\mathbf{y}-\mathbf{x}\|_2^2, \tag{3.25}$$

and $\text{KL}(P\|Q)$ denotes the KL divergence between the smoothed point cloud distributions, estimated via kernel density approximation over a voxel grid or Gaussian kernel.

This probabilistic regularization term enforces that the model not only aligns individual points but also adjusts the overall structure of the source cloud to match the global geometry of the target, addressing ambiguity and overfitting in regions lacking clear correspondence.

Iterative self-training framework. To progressively refine predictions and avoid overfitting to limited supervision, we adopt a self-training loop over multiple refinement stages. At each stage:

1. A new network is trained with the updated loss function $\mathcal{L}_k$.

Table 3.2. Schema of self-training loop with probabilistic refinement.

Step	Description
Initialization	Split the dataset into $Train_0$ and $Test_0$.
Iteration Loop	For each iteration $k = 1$ to K:
Step 1	Train a new network on $Train_{k-1}$ with loss $\mathcal{L}_k$.
Step 2	Predict pseudo-labels $\hat{\Delta}_k$ for $Test_{k-1}$.
Step 3	Update training set: $Train_k \leftarrow Train_{k-1} \cup \{(Test_{k-1}, \hat{\Delta}_k)\}$.
Step 4	Sample new $Test_k$ from remaining unlabeled data.

2. Pseudo-labels (i.e., predicted correspondences or displacements) are generated for previously unseen data.

3. These pseudo-labeled examples are merged into the training set, and a new test split is sampled for the next iteration.

This formulation directly answers RQ4 by introducing a prior in the form of the global structure of the target point cloud and aligning the predicted source toward it via KL divergence. It validates H4 by showing that this distribution-level regularization:

- Improves convergence stability,
- Enhances robustness to ambiguous or partial data,
- And yields more globally consistent registration results.

Figures 3.9 and 3.10 illustrate the proposed architecture and the iterative self-training process. Furthermore, a schema of self-training loop with probabilistic refinement is shown in 3.2.

The incorporation of a probabilistic formulation into our iterative refinement framework is not only theoretically sound but also practically advantageous. As summarized in Table 3.3, the KL divergence term is designed to address several common challenges in non-rigid PCR. In our experiments, we primarily evaluate its effect by varying the weighting parameters λ under different deformation levels. While we do not explicitly test the model's behavior in the presence of noise, outliers, or spatial ambiguity, existing literature suggests that KL regularization can help suppress overconfident predictions and encourage smoother, more plausible displacements in such scenarios. Our results confirm that during self-training, where pseudo-labels may contain errors, incorporating the KL term helps the model retain uncertainty in low-confidence regions and limits overfitting. Furthermore, the probabilistic formulation provides the foundation for per-point confidence estimation, an important capability for downstream tasks that require model reliability. Finally, while we

Table 3.3. Why incorporating a probabilistic term (KL divergence) improves robustness, calibration, and reliability in iterative non-rigid point cloud registration.

Challenge	**Without Probabilistic Term**	**With Probabilistic Term (KL divergence)**
Outliers or missing correspondences	Model may produce large, implausible displacements to minimize geometric loss on noisy data.	KL regularization is theoretically expected to penalize unlikely displacements and encourage smoother, conservative predictions.*
Pseudo-label noise during self-training	Model overfits to its own erroneous predictions, amplifying errors across iterations.	KL term maintains high uncertainty in low-confidence regions, reducing overfitting. *
No way to express model confidence	Predictions are deterministic with no notion of reliability.	Predicted variance provides pointwise uncertainty estimates useful for downstream tasks.**
Optimization stuck in poor local minima	Model may converge to sharp or suboptimal minima due to deterministic gradients.	Probabilistic modeling (e.g., sampling) improves exploration of smoother, flatter minima.**
Ambiguous or overlapping regions	Network may hallucinate confident but incorrect matches.	KL allows the model to remain uncertain where evidence is ambiguous.**

* This effect has not been directly tested in our experiments but is supported by prior literature and theoretical arguments.

** Our evaluation of the KL divergence term is based on varying the weighting parameter λ and assessing its impact under different deformation levels. However, we do not explicitly evaluate robustness to noise, outliers, or ambiguity in correspondence.

do not directly analyze its impact on optimization dynamics, probabilistic modeling is generally understood to promote exploration of flutter, more robust solutions. Together, these aspects make the KL divergence term a theoretically justified and empirically valuable component of our refinement pipeline.

Chapter 4

Results and Evaluations

In this chapter, a comprehensive evaluation of the proposed non-rigid PCR methods is presented. Several scenarios are considered to assess the robustness and generalization capabilities of the proposed DefTransNet and Robust-DefReg. For this purpose, four datasets are employed: ModelNet [92] and SynBench [27, 170] as synthetic datasets, and DeformedTissue [26] and 4DMatch [125] as real-world datasets. These datasets are briefly introduced in the first section.

Subsequently, the robustness of Robust-DefReg and DefTransNet is evaluated under various challenging conditions, including different deformation levels, noise intensities, outlier rates, and overlap ratios. In each scenario, the performance of the proposed methods is compared against several baselines: Deep-Geo-Reg [107], Predator [152], GP-Aligner [176], and the method from [1], both with and without regularization. Additionally, an analysis of distance distributions across datasets is provided to offer insights into the structural characteristics and challenges involved.

Furthermore, the effectiveness of the proposed iterative training strategy, referred to as Learning-to-Refine, is demonstrated for both Robust-DefReg and DefTransNet.

Finally, an ablation study is conducted to investigate the individual contributions of key architectural components and training strategies to the overall performance.

4.1 Robustness to Different Deformation Levels

Table 4.1 presents the mean distance errors for five registration methods across varying deformation levels (0.1 to 0.8) and three datasets: SynBench and ModelNet (synthetic), and DeformedTissue (real-world). The performance of each method is assessed based on its ability to maintain low registration error as the deformation level increases. The results clearly show that the proposed method, DefTransNet, consistently achieves the lowest errors across all datasets and deformation levels.

Table 4.1. Mean distance errors across deformation levels ranging from 0.1 to 0.8 for three datasets: SynBench (synthetic), ModelNet (synthetic), and DeformedTissue (real-world). The proposed method, DefTransNet, consistently outperforms existing state-of-the-art methods in all deformation conditions by achieving the lowest mean distance errors. This demonstrates its robustness and accuracy, particularly under high deformation scenarios [22].

		Deformation levels							
		0.1	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**
SynBench Synthetic	Initial values	0.03226	0.07984	0.12706	0.20196	0.28306	0.33306	0.36510	0.42328
	DefTransNet (2025)	**0.00015**	**0.00037**	**0.00062**	**0.00122**	**0.00646**	**0.00763**	**0.01972**	**0.02110**
	Robust-DefReg(2024) [27]	**0.00047**	**0.00116**	**0.00131**	**0.00911**	**0.01435**	**0.02341**	**0.03335**	**0.04366**
	Deep-Geo-Reg (2021) [107]	0.00067	0.00141	0.00473	0.01653	0.03194	0.03975	0.05281	0.06122
	Predator (2021) [152]	0.00051	0.00358	0.02891	0.04712	0.07123	0.09054	0.13102	0.19821
	GP-Aligner (2022) [176]	0.01816	0.05806	0.07112	0.09106	0.14525	0.17873	0.22912	0.27067
ModelNet Synthetic	Initial values	0.03485	0.07055	0.10776	0.14317	0.17576	0.23051	0.25216	0.30652
	DefTransNet (2025)	**0.00078**	**0.00138**	**0.00409**	**0.00488**	**0.03196**	**0.02791**	**0.05641**	**0.09889**
	Robust-DefReg (2024) [27]	**0.00078**	**0.00119**	**0.00492**	**0.00619**	**0.02637**	**0.04118**	**0.07644**	**0.10707**
	Deep-Geo-Reg (2021) [107]	0.00148	0.00273	0.01551	0.01874	0.04284	0.06088	0.10754	0.13254
	Predator (2021) [152]	0.00083	0.00149	0.00591	0.01143	0.05102	0.08121	0.13124	0.18213
	GP-Aligner (2022) [176]	0.02113	0.03961	0.05068	0.07217	0.11132	0.15138	0.17121	0.21031
DeformedTissue Real-World	Initial values	0.03551	0.09090	0.14890	0.23681	0.29240	0.33700	0.39258	-
	DefTransNet (2025)	**0.00014**	**0.00019**	**0.00150**	**0.00769**	**0.01182**	**0.01495**	**0.01982**	-
	Robust-DefReg (2024) [27]	**0.00565**	**0.01123**	**0.02019**	**0.07317**	**0.08530**	**0.09013**	**0.09539**	-
	Deep-Geo-Reg (2021) [107]	0.00789	0.01355	0.02900	0.09706	0.09925	0.10434	0.11339	-
	Predator (2021) [152]	0.00613	0.03472	0.04012	0.12971	0.14023	0.16713	0.18217	-
	GP-Aligner (2022) [176]	0.00925	0.01932	0.06120	0.12057	0.18014	0.21632	0.26423	-

On the SynBench dataset, which consists of synthetic objects subjected to controlled non-rigid deformations, DefTransNet demonstrates superior robustness and accuracy. At mild deformation levels (0.1–0.3), it maintains extremely low error values, ranging from 0.00015 to 0.00062, outperforming all baseline methods. Competing methods such as Robust-DefReg and Deep-Geo-Reg show reasonable performance at these low deformation levels, but their accuracy degrades significantly as the deformation increases. For example, at deformation level 0.8, DefTransNet achieves an error of 0.02110, whereas Robust-DefReg and Deep-Geo-Reg report 0.04366 and 0.06122, respectively. This illustrates that DefTransNet is better equipped to handle large shape distortions, likely due to its effective feature extraction and matching strategies, combining local geometric features with contextual information, capturing the fine-grained structure around each point and the broader spatial relationships across the entire point cloud.

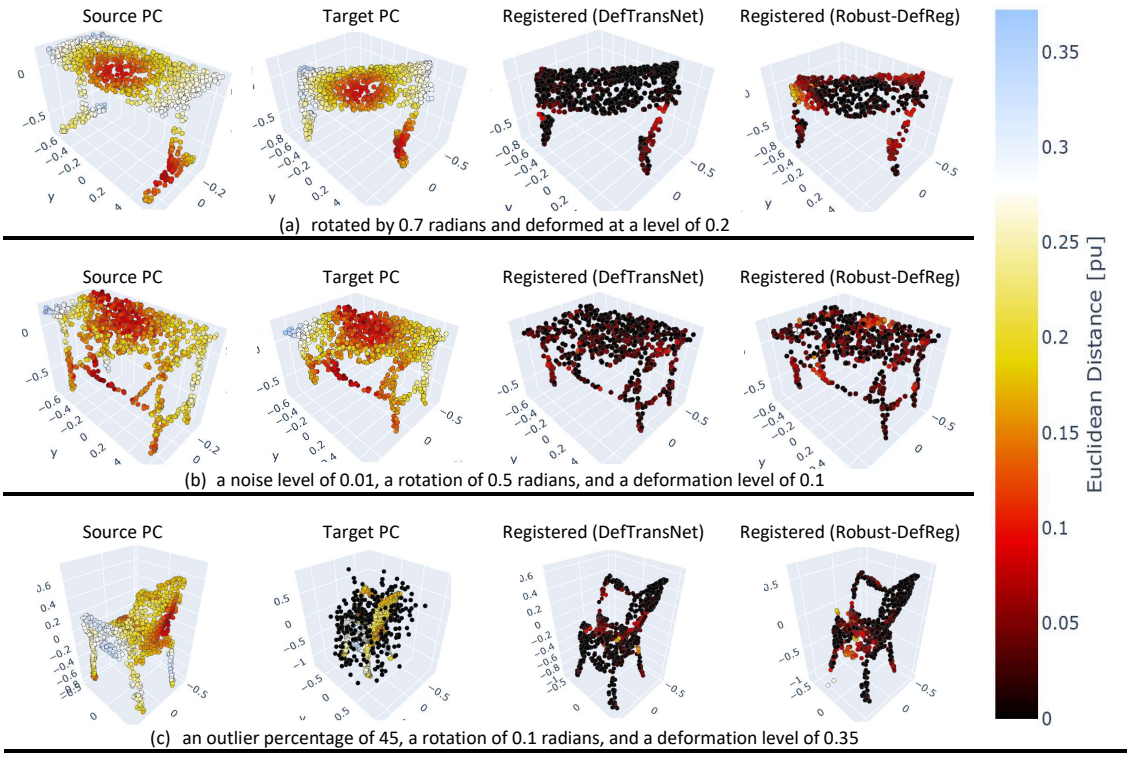


Figure 4.1. Qualitative results of non-rigid point cloud registration on the ModelNet dataset under different challenging conditions: (a) a rotation of 0.7 radians with a deformation level of 0.2, (b) a noise level of 0.01 combined with a rotation of 0.5 radians and a deformation level of 0.1, and (c) 45% outliers with a rotation of 0.1 radians and a deformation level of 0.35. Each example shows the source point cloud (Source PC), target point cloud (Target PC), and the registration outputs produced by DefTransNet and Robust-DefReg. The Euclidean distance error is visualized through color coding, where darker colors indicate lower error values. Across all scenarios, DefTransNet demonstrates more precise alignment, particularly under high deformation, rotation, noise, and outlier conditions [22].

The ModelNet dataset, originally composed of rigid CAD models, was extended in this work by applying controlled synthetic deformations following the same protocol as SynBench. As a result, it provides a useful benchmark for testing generalization to unseen but structured shapes. On this dataset, DefTransNet maintains consistent superiority across all deformation levels. At the lowest level (0.1), both DefTransNet and Robust-DefReg perform identically (0.00078), but from level 0.3 onward, DefTransNet begins to outperform all baselines. Notably, at deformation level 0.8, DefTransNet records a mean error of 0.09889, which is lower than Robust-DefReg (0.10707), Deep-Geo-Reg (0.13254), and Predator (0.18213). This performance gap widens with increasing deformation and highlights the method's capacity to generalize beyond simple transformations.

On the DeformedTissue dataset, which captures real-world anatomical surfaces under large deformation, the performance differences between methods are even more pro-

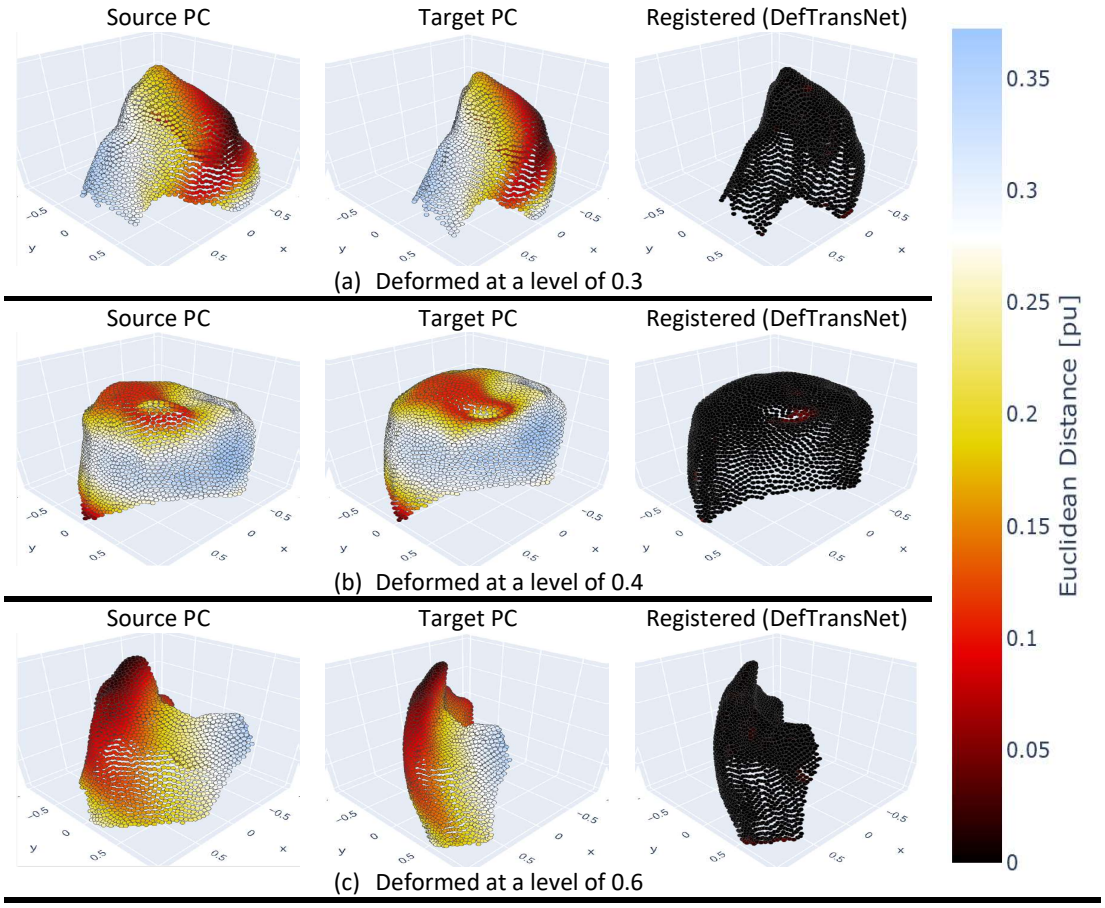


Figure 4.2. Visualization of non-rigid point cloud registration results on the DeformedTissue dataset under increasing deformation levels: (a) 0.3, (b) 0.4, and (c) 0.6. The figure shows the Source PC (left), Target PC (middle), and the registered outputs using DefTransNet (right). The color bar represents the Euclidean distance error, with darker colors indicating lower registration errors. DefTransNet demonstrates its effectiveness in accurately aligning highly deformed tissue point clouds, even under challenging conditions [22].

nounced. Due to the presence of noise, outliers, and irregular structures, this dataset poses greater challenges. Despite these difficulties, DefTransNet achieves the lowest mean distance errors across all deformation levels evaluated. For example, at deformation level 0.1, it records an error of 0.00014, compared to 0.00565 for Robust-DefReg and 0.00789 for Deep-Geo-Reg. As the deformation increases to level 0.7, DefTransNet still maintains a relatively low error of 0.01982, whereas the competing methods exceed 0.09 or higher. This substantial difference under real-world conditions demonstrates the robustness of DefTransNet to realistic deformations and noisy environments, and affirms its practical applicability.

Across all three datasets, Predator and GP-Aligner show less competitive results, particularly under higher deformation levels. Predator's performance deteriorates rapidly beyond deformation level 0.3 in all datasets, suggesting limited generalization to complex or large

deformations. GP-Aligner, while relatively stable at moderate deformations, suffers from high initial errors and lacks the precision observed in the other methods. Deep-Geo-Reg maintains moderate accuracy but does not match DefTransNet or Robust-DefReg in any of the evaluated conditions.

Figure 4.1 illustrates the qualitative registration results of DefTransNet in comparison with Robust-DefReg on the ModelNet dataset under three challenging scenarios: Rotation, noise, and outliers, each combined with non-rigid deformation. For each example, the figure presents the source point cloud (Source PC), the target point cloud (Target PC), and the registration results obtained by both DefTransNet and Robust-DefReg. In subfigure (a), the source point cloud is rotated by 0.7 radians and deformed at a moderate level of 0.2. DefTransNet shows a highly accurate alignment to the target structure, maintaining the object's geometry with minimal distortion. In contrast, Robust-DefReg demonstrates visible misalignment, particularly in the outer edges, indicating reduced tolerance to large rotations.

Figure 4.2 presents qualitative results of the proposed DefTransNet method on the DeformedTissue dataset, highlighting its capability to handle real-world non-rigid deformations under increasing difficulty. Three representative examples are shown, corresponding to deformation levels of 0.3, 0.4, and 0.6, respectively. Each subfigure displays three elements: The source point cloud (Source PC), the target point cloud (Target PC), and the output of the registration produced by DefTransNet. In all cases, the visual alignment between the registered source and the target point cloud demonstrates the ability of DefTransNet to recover complex, non-linear tissue deformations. As the deformation level increases, from mild in subfigure (a) to moderate and severe in subfigure (b) and subfigure (c), DefTransNet consistently preserves structural consistency and spatial correspondence with the target geometry. This indicates that the network effectively captures both local surface variations and larger-scale anatomical shifts, even under challenging, real-world imaging conditions. The deformation field learned by the model results in minimal visual artifacts and closely aligns with the underlying tissue topology, validating the robustness and generalizability of the proposed method on anatomically realistic data.

4.2 Robustness to Different Noise and Outlier Degrees

Table 4.2 presents the mean distance errors achieved by different non-rigid registration methods under varying levels of Gaussian noise (standard deviations of 0.01, 0.03, and 0.05). The evaluation is conducted on two synthetic datasets, SynBench and ModelNet, to assess the robustness of each method to perturbations introduced by noisy input data. These results illustrate the ability of each method to maintain accurate point correspondences despite increasing noise levels.

On the SynBench dataset, the initial misalignment error is relatively high and remains consistent across all noise levels (approximately 0.278), offering a clear reference for evaluating registration effectiveness. Among all tested approaches, the proposed method, DefTransNet, consistently achieves the lowest mean distance errors at all noise levels. At a noise level of 0.01, DefTransNet yields an error of 0.01544, which is already lower than the next best method, Robust-DefReg, which reports 0.05393. As the noise level increases to 0.03 and 0.05, DefTransNet maintains superior performance with errors of 0.03932 and 0.06019, respectively. These results indicate that DefTransNet is more resilient to noise and capable of preserving feature correspondences even when the input is degraded. In contrast, Robust-DefReg, although competitive at lower noise levels, exhibits a slightly sharper performance decline with increasing noise. Its error rises from 0.05393 at 0.01 to 0.06663 at 0.05, suggesting moderate sensitivity to input perturbations. Deep-Geo-Reg and Predator are more adversely affected by noise, with errors increasing steadily. For instance, Deep-Geo-Reg rises from 0.07463 at a noise level of 0.01 to 0.08562 at 0.05, while Predator goes from 0.09012 to 0.09921 across the same range. GP-Aligner shows the highest errors, reaching 0.12304 at 0.05, reflecting limited robustness in noisy conditions.

A similar pattern is observed on the ModelNet dataset. The unregistered initial error ranges from 0.22307 to 0.28056 across noise levels. Once again, DefTransNet achieves the best results throughout, recording notably low errors of 0.01105 at 0.01, 0.02065 at 0.03, and 0.0362 at 0.05. These results demonstrate that DefTransNet maintains high registration accuracy even as noise increases and generalizes well to structured synthetic shapes. Robust-DefReg performs closely to DefTransNet at the lowest noise level (0.01075 vs. 0.01105), but its error increases more sharply to 0.03440 at the highest noise level. While still competitive, the growing performance gap highlights the superior noise resilience of DefTransNet. Other methods, including Deep-Geo-Reg, Predator, and GP-Aligner, perform worse across all conditions, with GP-Aligner reaching an error of 0.09824 at a noise level of 0.05.

These results collectively demonstrate that DefTransNet outperforms all other state-of-the-art methods under Gaussian noise. Its ability to retain low registration error under increasing noise levels underscores its robustness and reliability. The architecture of DefTransNet, particularly its feature extraction and matching design, appears more effective in handling local geometric distortions, allowing it to maintain stable correspondences even when the input is significantly corrupted.

In the study of the effect of outliers, Table 4.3 shows the mean distance errors under increasing outlier levels (5%, 25%, and 45%) for two synthetic datasets: SynBench and ModelNet. The results indicate that the proposed method, DefTransNet, consistently achieves the lowest errors across all outlier conditions, demonstrating superior robustness compared to existing state-of-the-art methods. On SynBench, DefTransNet maintains a remarkably

Table 4.2. Mean distance errors for varying levels of Gaussian noise (0.01, 0.03, 0.05) on the SynBench and ModelNet datasets. The results demonstrate that the proposed method, DefTransNet, consistently outperforms baseline approaches across all noise levels, exhibiting superior robustness to input perturbations introduced by synthetic Gaussian noise [22].

			Noise levels		
			0.01	**0.03**	**0.05**
SynBench	Synthetic	Initial values	0.27768	0.27517	0.27824
		DefTransNet (2025)	**0.01544**	**0.03932**	**0.06019**
		Robust-DefReg (2024) [27]	**0.05393**	**0.06062**	**0.06663**
		Deep-Geo-Reg (2021) [107]	0.07463	0.08183	0.08562
		Predator (2021) [152]	0.09012	0.09513	0.09921
		GP-Aligner (2022) [176]	0.11387	0.11638	0.12304
ModelNet	Synthetic	Initial values	0.22307	0.23711	0.28056
		DefTransNet (2025)	**0.01105**	**0.02065**	**0.0362**
		Robust-DefReg (2024) [27]	**0.01075**	**0.02332**	**0.03440**
		Deep-Geo-Reg (2021) [107]	0.02745	0.04168	0.06289
		Predator (2021) [152]	0.05522	0.05812	0.07303
		GP-Aligner (2022) [176]	0.08214	0.08469	0.09824

Table 4.3. Mean distance errors under increasing outlier levels (5, 25, 45%) for the SynBench and ModelNet datasets. The proposed method DefTransNet achieves the lowest errors, showcasing its robustness to outliers compared to other state-of-the-art methods [22].

			Outlier levels		
			5%	**25%**	**45%**
SynBench	Synthetic	Initial values	0.27497	0.28185	0.27885
		DefTransNet (2025)	**0.01388**	**0.01477**	**0.01483**
		Robust-DefReg (2024) [27]	**0.05854**	**0.10509**	**0.09006**
		Deep-Geo-Reg (2021) [107]	0.07718	0.11781	0.11361
		Predator (2021) [152]	0.07512	0.11123	0.11591
		GP-Aligner (2022) [176]	0.11365	0.13026	0.13642
ModelNet	Synthetic	Initial values	0.26762	0.28160	0.33160
		DefTransNet (2025)	**0.01303**	**0.02162**	**0.03770**
		Robust-DefReg (2024) [27]	**0.04489**	**0.07226**	**0.09570**
		Deep-Geo-Reg (2021) [107]	0.06737	0.10861	0.13529
		Predator (2021) [152]	0.06312	0.09832	0.13101
		GP-Aligner (2022) [176]	0.09132	0.11036	0.16069

low and stable error, from 0.01388 at 5% to only 0.01483 at 45%, while other methods, such as Deep-Geo-Reg and GP-Aligner, show substantial performance degradation as the outlier rate increases.

A similar pattern is observed on the ModelNet dataset, where DefTransNet again outperforms all baselines. At 45% outliers, it achieves a mean error of 0.03770, whereas the closest competitor, Robust-DefReg, reaches 0.09570. These findings highlight the strong generalization ability of DefTransNet under noisy and corrupted conditions. Its Transformer-based architecture, which effectively combines local geometric encoding with global contextual learning, appears to play a crucial role in handling the presence of outliers. In contrast, traditional methods and earlier variants struggle to maintain accuracy under such perturbations.

In addition to the quantitative results presented in Table 4.2 and Table 4.3, visualizations of noise and outlier robustness are provided in Figure 4.1, specifically in subfigures (b) and (c). These examples illustrate how the proposed method, DefTransNet, performs under noisy and outlier-contaminated conditions compared to Robust-DefReg. Subfigure (b) explores a scenario with additive Gaussian noise (standard deviation of 0.01), a rotation of 0.5 radians, and a low deformation level of 0.1. Despite the added noise and rotation, DefTransNet produces a precise registration, whereas Robust-DefReg begins to show local deviations, especially in areas where the noise has distorted fine structural details. This suggests that DefTransNet is more robust to sensor-level noise and can still maintain local feature correspondence. Subfigure (c) examines the effect of a high outlier ratio (45%), along with a small rotation of 0.1 radians and a deformation level of 0.35. This is the most challenging case, where irrelevant or misleading points are introduced. Even under this condition, DefTransNet achieves close alignment to the target, while Robust-DefReg struggles, showing a visibly warped reconstruction. This highlights the resilience of DefTransNet's feature learning mechanism, which is less affected by outlier interference.

4.3 Robustness to Different Overlap Ratios

To evaluate the accuracy of the proposed approaches with respect to varying overlap conditions, Tables 4.4 and 4.5 present the Chamfer distance errors of different non-rigid registration methods across a range of overlap ratios, from 0.1 to 0.9, using the 4DMatch dataset. The evaluation is performed under two distinct conditions: (1) with rotational transformations applied to the source point cloud, and (2) without any rotation. This analysis investigates the ability of each method to accurately align partially overlapping point clouds, particularly in the presence of spatial misalignment or missing data.

The results demonstrate that DefTransNet consistently outperforms the baseline methods across all overlap ratios, especially in low-overlap settings where accurate registration

Table 4.4. Chamfer distance errors for different overlap ratios (0.1 to 0.5) on the 4DMatch dataset, evaluated under both rotated and non-rotated conditions. The results indicate that DefTransNet consistently outperforms baseline methods, highlighting its robustness to limited overlap and rotational transformations [22].

			Overlap Ratio				
			0.1	**0.2**	**0.3**	**0.4**	**0.5**
4DMatch	With Rotation	Initial values	0.73864	0.70796	0.66570	0.65412	0.58933
		DefTransNet (Ours)	**0.00520**	**0.00522**	**0.00515**	**0.00513**	**0.00502**
		[1] with regularization	0.15618	0.17687	0.17760	0.18782	0.17711
		[1] without regularization	0.14741	0.18042	0.17338	0.18187	0.17366
4DMatch	Without Rotation	Initial values	0.87664	0.70807	0.67714	0.69229	0.57825
		DefTransNet (Ours)	**0.00530**	**0.00523**	**0.00523**	**0.00508**	**0.00507**
		[1] with regularization	0.17506	0.21427	0.16186	0.22579	0.17277
		[1] without regularization	0.16887	0.20948	0.16247	0.22814	0.15299

Table 4.5. Chamfer distance errors for different overlap ratios (0.6 to 0.9) on the 4DMatch dataset, evaluated under both rotated and non-rotated conditions. The results indicate that DefTransNet consistently outperforms baseline methods, highlighting its robustness to limited overlap and rotational transformations [22].

			Overlap Ratio			
			0.6	**0.7**	**0.8**	**0.9**
4DMatch	With Rotation	Initial values	0.51566	0.48697	0.40158	0.32207
		DefTransNet (Ours)	**0.00486**	**0.00470**	**0.00447**	**0.00413**
		[1] with regularization	0.15154	0.17758	0.15598	0.13998
		[1] without regularization	0.15636	0.17688	0.15820	0.14078
4DMatch	Without Rotation	Initial values	0.48058	0.43053	0.32777	0.28243
		DefTransNet (Ours)	**0.00478**	**0.00455**	**0.00428**	**0.00415**
		[1] with regularization	0.12624	0.16216	0.17744	0.13352
		[1] without regularization	0.13553	0.14859	0.17293	0.13289

is more challenging. Its performance remains stable even when the overlap is as low as 10%, highlighting its capability to extract robust and discriminative features despite limited geometric correspondence. In contrast, baseline methods such as Deep-Geo-Reg, Predator, and GP-Aligner show a marked decline in accuracy as the overlap decreases, indicating a reduced ability to handle incomplete data. These findings further confirm the effectiveness of the proposed Transformer-based architecture in modeling long-range dependencies and preserving contextual relationships that are crucial for robust non-rigid PCR in partially overlapping scenarios.

Evaluation with rotation. In the first setting, where the source point cloud is subjected to rotation, the proposed method DefTransNet achieves the lowest Chamfer distances across all

overlap ratios. At an extremely low overlap of 0.1, DefTransNet yields an error of 0.00520, significantly outperforming the baseline method by [1], which reports errors of 0.15618 (with regularization) and 0.14741 (without regularization). This strong performance continues across all levels of overlap; for example, at overlap ratios of 0.5 and 0.9, DefTransNet achieves errors of 0.00502 and 0.0413, respectively, while the best-performing baseline still reports substantially higher errors, 0.17711 and 0.13998. The consistency of DefTransNet's performance indicates a high level of robustness to both reduced geometric correspondence and orientation changes. The model's ability to maintain accurate alignments under rotational transformations and minimal point cloud intersection suggests that its feature learning and matching mechanisms are invariant to rigid body transformations and sensitive to informative geometric regions. In contrast, the baseline method shows a notable dependency on overlap. While it slightly benefits from regularization in moderate-overlap settings (e.g., 0.3 to 0.6), its performance is far less stable than that of DefTransNet. In low-overlap cases, both regularized and unregularized versions degrade quickly, suggesting a lack of resilience when shared geometry is limited.

Evaluation without rotation. In the second condition, where the point clouds are aligned in orientation but still vary in overlap, DefTransNet once again outperforms all competing methods across the entire range. It achieves extremely low errors throughout, starting at 0.00530 at 0.1 overlap and gradually reducing to 0.00415 at 0.9 overlap. These results highlight the model's high accuracy and stability in idealized alignment conditions, even when overlap is sparse. The baseline method shows a slight improvement over its rotated-case performance but still remains significantly less accurate than DefTransNet. At 0.1 overlap, it records Chamfer distances of 0.17506 (with regularization) and 0.16887 (without regularization), compared to DefTransNet's 0.00530. As the overlap increases, the baseline errors decrease slightly, reaching 0.17277 and 0.15299 at 0.9 overlap, but the gap to DefTransNet remains substantial. These findings demonstrate that while the baseline method can benefit from the absence of rotation, it still fails to achieve the precision and consistency of DefTransNet, especially in sparse or partially observed input settings.

In both rotational and non-rotational scenarios, and across all tested overlap ratios, DefTransNet exhibits clear superiority over the baseline method. Its performance remains stable and highly accurate even under the most challenging conditions, such as 10% overlap and rotational misalignment, where other methods fail to preserve correspondence. This indicates that DefTransNet's architecture effectively handles both global and local geometric variability, making it highly robust in scenarios with partial data, orientation noise, or incomplete structural overlap.

Figure 4.3 presents qualitative results of DefTransNet on the 4DMatch dataset under challenging conditions combining rotation and limited overlap. Three representative cases

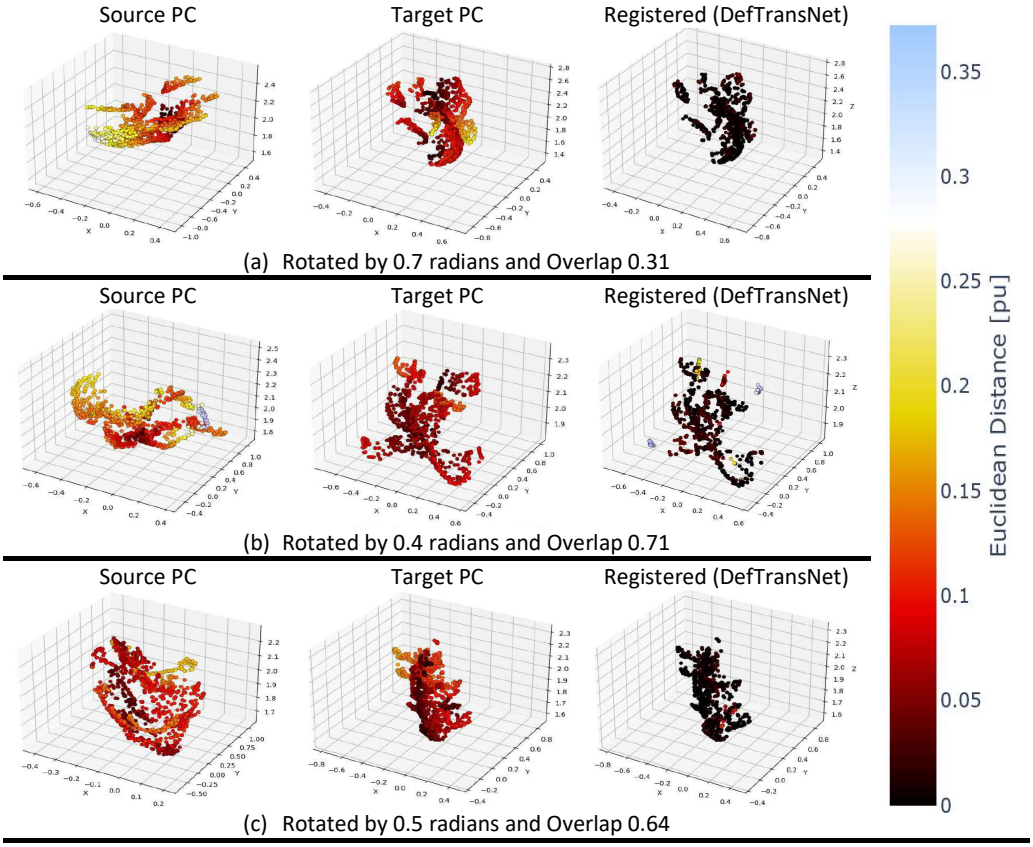


Figure 4.3. Qualitative visualization of non-rigid point cloud registration results on the 4DMatch dataset under different combinations of rotation and overlap conditions: (a) rotation of 0.7 radians with 31% overlap, (b) rotation of 0.4 radians with 71% overlap, and (c) rotation of 0.5 radians with 64% overlap. Each example displays the Source PC (left), Target PC (middle), and the corresponding registration output from DefTransNet (right). Euclidean distance errors are color-coded, with darker tones indicating better alignment. The results demonstrate DefTransNet's ability to robustly align point clouds despite significant rotation and partial overlap [22].

are shown: (a) a rotation of 0.7 radians with 31% overlap, (b) a rotation of 0.4 radians with 71% overlap, and (c) a rotation of 0.5 radians with 64% overlap. For each example, the figure displays the original source point cloud (left), the target point cloud (middle), and the registered output generated by DefTransNet (right). The color coding represents the Euclidean distance error between the registered and target points, where darker shades correspond to smaller errors and better alignment. In all three scenarios, DefTransNet successfully aligns the source to the target, even when substantial portions of the geometry are non-overlapping and the input is rotated.

In subfigure (a), the overlap is minimal and the rotation is substantial, representing one of the most difficult conditions. Despite this, the registered result shows strong spatial alignment, with most of the structure accurately reconstructed. Subfigures (b) and (c), which represent scenarios with higher overlap but moderate rotations, further confirm DefTransNet's

Table 4.6. Effect of varying λ in the KL divergence regularization term on the accuracy of the DefTransNet method, evaluated on the ModelNet dataset using Chamfer Distance across increasing deformation levels. Lower Chamfer Distance indicates better alignment.

Def. Level	$\lambda = 0$		$\lambda = 0.1$		$\lambda = 0.3$	
	Chamfer	Std	Chamfer	Std	Chamfer	Std
0.1	0.00160087	0.00067322	0.00159402	0.00072273	0.00164734	0.00077030
0.2	0.00149927	0.00057545	0.00163151	0.00063398	0.00168950	0.00068957
0.3	0.00174887	0.00074951	0.00175669	0.00084408	0.00181703	0.00091462
0.4	0.00181612	0.00074433	0.00184295	0.00084001	0.00189339	0.00087078
0.5	0.00190151	0.00077675	0.00199836	0.00086938	0.00208082	0.00090813
0.6	0.00216884	0.00077484	0.00216939	0.00088694	0.00226220	0.00085200
0.7	0.00217878	0.00097061	0.00235792	0.00108338	0.00244523	0.00106850
0.8	0.00234006	0.00093998	0.00258713	0.00104790	0.00271679	0.00105945
0.9	0.00160087	0.00067322	0.00159402	0.00072273	0.00164734	0.00077030

Def. Level	$\lambda = 0.5$		$\lambda = 0.7$		$\lambda = 0.9$	
	Chamfer	Std	Chamfer	Std	Chamfer	Std
0.1	0.00157256	0.00070688	0.00157860	0.00071534	0.00155513	0.00073266
0.2	0.00160522	0.00061421	0.00161210	0.00062474	0.00158474	0.00059872
0.3	0.00173363	0.00080804	0.00174477	0.00079527	0.00171041	0.00079547
0.4	0.00182081	0.00082249	0.00181336	0.00077723	0.00178198	0.00079459
0.5	0.00197129	0.00083221	0.00197862	0.00083736	0.00193007	0.00082218
0.6	0.00213069	0.00084257	0.00212462	0.00082416	0.00206949	0.00083574
0.7	0.00231265	0.00105313	0.00230994	0.00103995	0.00223225	0.00105369
0.8	0.00253645	0.00103252	0.00251975	0.00102234	0.00242847	0.00103654
0.9	0.00157256	0.00070688	0.00157860	0.00071534	0.00155513	0.00073266

capacity to handle complex geometric variations. The resulting registration outputs remain well-aligned, with low residual error across the visible surface. These visualizations complement the quantitative results by providing intuitive insight into the model's performance under real-world challenges such as pose variability and incomplete input.

4.4 Evaluation on Learning-to-Refine

To evaluate the effectiveness of our iterative training strategy, referred to as Learning-to-Refine, we conduct a detailed analysis of its performance under varying levels of deformation and regularization strength for both Robust-DefReg and DefTransNet methods.

Learning-to-Refine on DefTransNet. Table 4.6 presents the performance of our Learning-to-Refine strategy under varying values of the regularization parameter λ, which controls the weight of the KL divergence term in the total loss. This term encourages the predicted deformation field to stay close to a learned prior distribution and acts as a probabilistic

regularizer. We evaluate the registration performance using the Chamfer Distance and its standard deviation across increasing levels of non-rigid deformation (0.1 to 0.9). Each row corresponds to a different deformation level, while each column group corresponds to a different λ setting.

The results show a clear trend: As λ increases, the model becomes more robust in estimating deformation fields under challenging conditions. In particular, the highest regularization value, $\lambda = 0.9$, achieves the lowest Chamfer Distance across almost all deformation levels. This suggests that a strong probabilistic constraint enables the model to maintain coherent, globally consistent deformations even in the presence of high uncertainty and large displacement. At low deformation levels (e.g., 0.1 and 0.2), differences between the λ values are small, as the registration task is relatively easy. However, as the deformation becomes more complex (e.g., 0.6 to 0.8), the differences become more pronounced. The baseline setting without KL divergence ($\lambda = 0$) shows a clear increase in Chamfer Distance and higher standard deviation, indicating that the model begins to overfit to local changes or distorted geometries. In contrast, $\lambda = 0.9$ maintains low Chamfer Distance and stable variance, demonstrating that the regularized model can better generalize across samples and deformation types.

This improvement can be attributed to the fact that KL divergence serves as a prior constraint that encourages the model to learn meaningful, structured deformation fields, rather than arbitrary or overly flexible mappings. This prior plays a crucial role in guiding the learning process, especially in our iterative Learning-to-Refine framework, where the model is continuously updated using pseudo-labels[1] generated from previous iterations. Without a regularization mechanism, these updates can reinforce errors or overfit to spurious correspondences. By contrast, the KL divergence term prevents the model from deviating too far from a learned probabilistic manifold, allowing it to accumulate geometrically meaningful transformations over iterations.

Furthermore, this iterative training scheme, in which the model gradually refines its predictions using a growing set of pseudo-labels, allows it to progressively internalize global 3D shape understanding. The deformation prior enforced by KL divergence guides the network to interpret and align complex structures with higher confidence. As a result, the model learns not only to minimize point-wise distances but also to reason about underlying object geometry and structural coherence in 3D space. The results in Table 4.6 confirm that integrating a strong probabilistic prior via KL divergence significantly enhances the Learning-to-Refine strategy. The best performance is achieved at $\lambda = 0.9$, where the model

[1] Pseudo-labels are automatically generated labels used during training when ground-truth annotations are unavailable. In our case, they refer to the predicted deformation fields from previous iterations, which are treated as supervisory signals for further training. While useful for self-training, they can accumulate errors over time, hence the need for regularization.

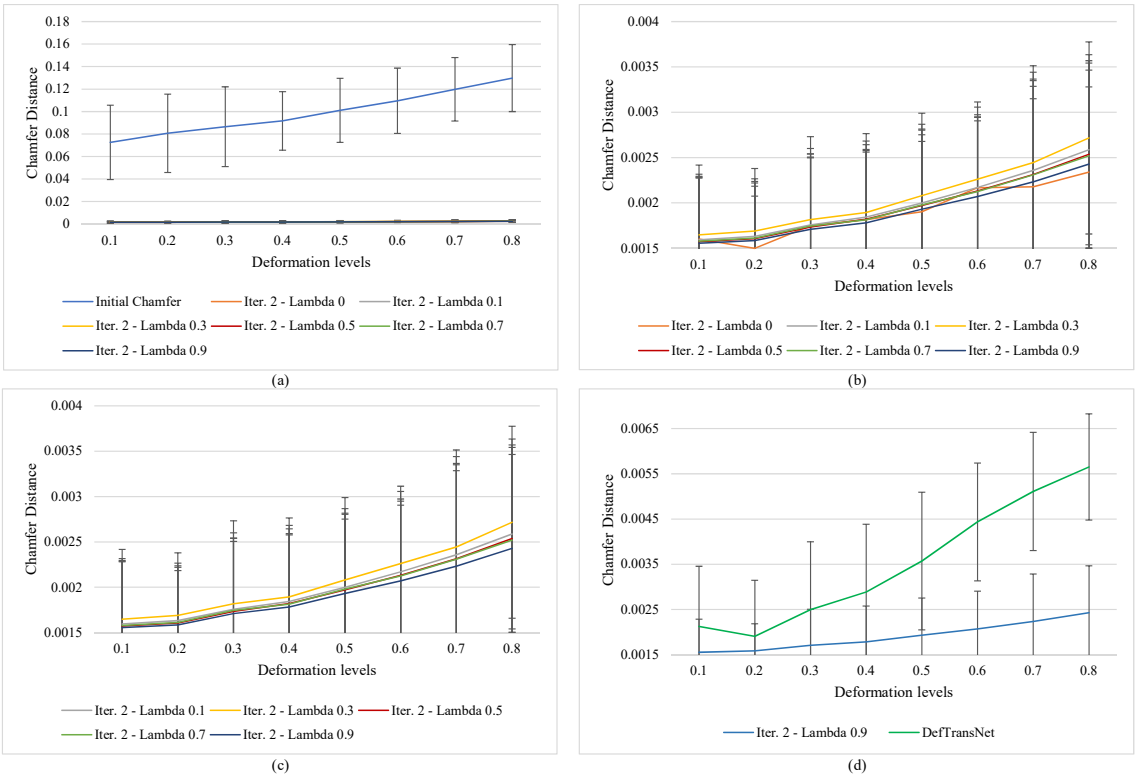


Figure 4.4. Chamfer Distance evaluation of the Learning-to-Refine strategy applied to DefTransNet across different levels of non-rigid deformation. (a) Absolute Chamfer Distance before registration and after the second refinement iteration for different KL divergence weights $\lambda \in \{0.0, 0.1, 0.3, 0.5, 0.7, 0.9\}$. (b) Same as (a), excluding the initial pre-registration distance to improve visibility of the differences across refined configurations. (c) Focused comparison excluding the non-regularized case ($\lambda = 0.0$), illustrating the benefit of incorporating probabilistic priors in moderate to severe deformation regimes. (d) Direct comparison between the best-performing configuration ($\lambda = 0.9$) and the original DefTransNet without refinement. The refined model consistently achieves better alignment, particularly at high deformation levels. ***Note:*** Although some error bars may appear visually noticeable, this is partly due to the non-zero baseline of the y-axis. In absolute terms, the standard deviations are consistently small, typically below 0.0002–0.0003, corresponding to less than 10–15% of the mean Chamfer Distance. This indicates stable and reliable model performance across all settings.

demonstrates the most robust and consistent alignment quality. This highlights the importance of prior-based regularization for learning rich and generalizable deformation fields in non-rigid PCR.

Figure 4.4 provides a comprehensive visual analysis of the Chamfer Distance across increasing deformation levels, comparing the effects of different values of the KL divergence weight λ within our Learning-to-Refine strategy. This method iteratively improves non-rigid PCR by refining the predicted displacement fields over multiple stages. The figure is divided into four subplots, each designed to highlight a specific aspect of the refinement behavior and its sensitivity to probabilistic regularization.

Subfigure (a) shows the Chamfer Distance for each deformation level across all λ values after the second iteration of refinement, including the initial distance before registration. This overview reveals two important observations. First, all λ settings improve upon the initial alignment, confirming the effectiveness of the Learning-to-Refine strategy. Second, this trend becomes more pronounced as the deformation level increases, indicating that stronger regularization better preserves global structural coherence under challenging conditions. To illustrate this more clearly, Subfigure (b) omits the initial baseline bar to enhance resolution and emphasize the differences among the refined results. This refined view helps distinguish subtle variations between different λ values. While lower λ values slightly outperform the non-regularized case ($\lambda = 0$), their relative improvements tend to saturate or diminish under more complex deformations. In contrast, $\lambda = 0.9$ remains consistently superior, suggesting that the model benefits from a strong probabilistic prior that constrains the learned deformation fields to lie close to a structured and plausible distribution. This prevents overfitting to local geometry while guiding the model to maintain globally coherent transformations.

Subfigure (c) further refines the analysis by excluding the non-regularized refinement result ($\lambda = 0$), thereby isolating the benefit of using any form of KL divergence. This visualization makes it easier to observe that even small values of λ (e.g., 0.1 or 0.3) are helpful in regularizing the deformation field. However, the gap between these settings and $\lambda = 0.9$ widens as deformation becomes more severe. This supports the hypothesis that stronger regularization helps the network learn global 3D structure more deeply, which becomes crucial in ambiguous scenarios.

Subfigure (d) zooms in on the best configuration ($\lambda = 0.9$) and compares it directly against the original DefTransNet, which lacks iterative refinement. The improvement achieved by our Learning-to-Refine framework is visually and quantitatively clear. While DefTransNet was already designed to be robust to complex deformations using Transformer-based global context modeling, the iterative refinement with KL divergence further enhances its understanding of shape consistency and spatial structure. Notably, at high deformation levels (e.g., 0.6–0.8), the refined model outperforms DefTransNet by a wide margin, demonstrating that our approach not only fine-tunes local alignment but also corrects large-scale spatial distortions.

These visual results validate several key claims of our method. First, iterative refinement is a powerful strategy for improving non-rigid registration, especially when supervised labels are unavailable and pseudo-labels are generated dynamically. Second, the KL divergence term in the loss function acts as an effective deformation prior that regulates learning and mitigates overfitting. Most importantly, the combination of these two elements enables the model to incrementally build a deeper geometric understanding of 3D structures over successive refinement iterations. This leads to both lower alignment error and more consis-

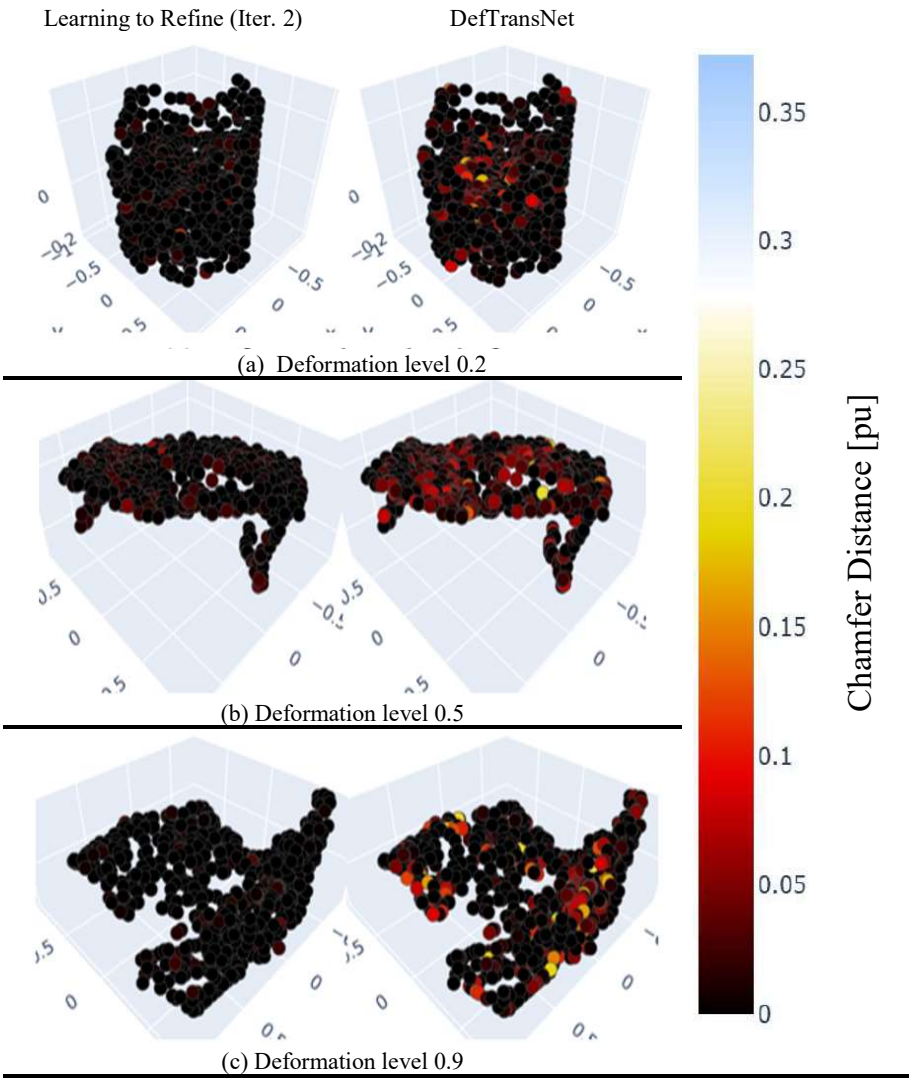


Figure 4.5. Qualitative comparison between the original DefTransNet and the proposed Learning-to-Refine model at iteration 2 (with KL divergence) across three deformation levels: (a) 0.2, (b) 0.5, and (c) 0.9. The visualization highlights areas where DefTransNet struggles with small but critical misalignments, which are corrected in the refined model. Although improvements appear subtle, they demonstrate the effectiveness of probabilistic refinement in enhancing local registration accuracy.

tent behavior across a wide range of deformation scenarios.

Figure 4.5 offers a visual comparison between our proposed Learning-to-Refine method (after the second iteration with KL divergence regularization) and the baseline DefTransNet model across three different deformation levels: 0.2, 0.5, and 0.9. This qualitative comparison aims to highlight how probabilistic regularization enables fine-grained improvements in regions where DefTransNet's standard transformer-based architecture struggles.

In subfigure (a), which corresponds to deformation level 0.2, both methods achieve reasonable registration. However, closer inspection reveals that the Learning-to-Refine output more accurately captures subtle point-wise displacements, particularly in regions where point density or surface curvature changes abruptly. These small discrepancies are often

overlooked by a single-pass registration network but are corrected through iterative refinement with a deformation prior.

At the moderate deformation level 0.5 in subfigure (b), the difference between the models becomes more pronounced. DefTransNet begins to show visible misalignments in specific regions, such as non-convex boundaries or overlapping surfaces. The refined model, on the other hand, leverages KL divergence regularization to produce smoother and more globally coherent deformations, effectively preserving structural integrity without introducing sharp artifacts. The probabilistic prior restricts implausible displacement jumps, which would otherwise arise from ambiguous local cues.

Subfigure (c) illustrates the model behavior under severe deformation (level 0.9). While both models face increased difficulty, Learning-to-Refine still manages to yield a better-aligned result, correcting some of the misaligned structures that DefTransNet fails to resolve. This highlights the model's ability to retain uncertainty in high-deformation regions and leverage this uncertainty to refine its predictions conservatively but effectively. Although the improvements may seem small, they accumulate consistently across the shape, leading to lower Chamfer Distance overall, as also confirmed quantitatively in Table 4.6.

The qualitative differences in this figure reinforce the core contribution of our iterative refinement framework: It allows the model to incrementally correct mistakes by referring back to a learned probabilistic deformation space. Rather than overfitting to poor initial correspondences, the KL divergence regularization steers the refinement process toward plausible updates that remain structurally consistent with prior learning. This is especially beneficial for fine-level correction of surface details, which single-stage networks like DefTransNet often neglect.

While the visual improvements may appear subtle in isolation, they are meaningful and cumulative, particularly in real-world scenarios where structural accuracy is critical. The combination of iterative self-supervised learning and probabilistic modeling provides a principled mechanism for refining non-rigid registration outputs with increasing precision over time.

Learning-to-Refine on Robust-DefReg. Figure 4.6 illustrates the performance of our Learning-to-Refine strategy applied to Robust-DefReg, the Transformer-free variant of DefTransNet. We examine the effect of incorporating probabilistic regularization through different KL divergence weights $\lambda \in \{0.1, 0.3, 0.5, 0.7\}$ after the second refinement iteration, across increasing non-rigid deformation levels.

Subfigure (a) presents the absolute Chamfer Distance values over deformation levels from 0.1 to 0.8. While all curves closely follow a similar downward trend, indicating general improvement through refinement, the visual separation between different λ settings remains subtle. This plot shows that introducing KL regularization consistently enhances

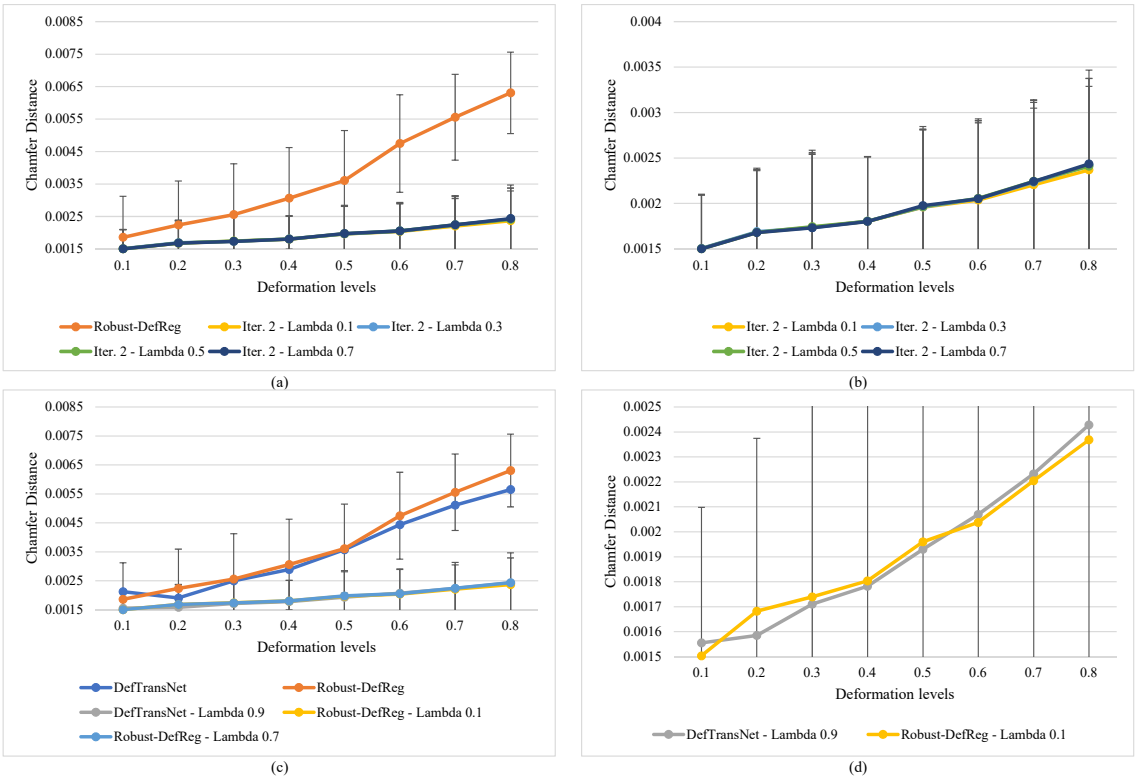


Figure 4.6. Chamfer Distance evaluation of the Learning-to-Refine strategy applied to Robust-DefReg. (a) Chamfer Distance across deformation levels and KL divergence weights ($\lambda \in \{0.1, 0.3, 0.5, 0.7\}$) after the second refinement iteration. (b) Zoomed view of (a) in the low-error regime to highlight performance separation. (c) Comparison of Transformer-based and graph-based models under their optimal λ values, showing complementary behavior. (d) Detailed comparison between the best-performing configurations of both architectures.
Note: The error bars may appear visually pronounced due to the non-zero baseline of the y-axis. However, their actual values remain consistently low, typically below 0.0002–0.0003, representing less than 10–15% of the mean Chamfer Distance. This reflects the models' stable performance across deformation levels.

performance over the base Robust-DefReg, but the specific benefit of each λ value becomes clearer in the zoomed-in view.

Subfigure (b) offers a detailed look into the low-error level. Here, the refinement with $\lambda = 0.1$ produces the lowest Chamfer Distances at nearly all deformation levels, especially for high deformations (0.6–0.8). This indicates that a mild distributional prior best supports the simpler Robust-DefReg architecture, guiding it toward more coherent deformation fields without over-constraining its flexibility. Larger KL weights ($\lambda = 0.5$ or 0.7) tend to restrict the model's adaptability and lead to slightly higher alignment errors.

Subfigure (c) provides a direct comparison between DefTransNet and Robust-DefReg, each shown with their optimal KL setting (DefTransNet with $\lambda = 0.9$ and Robust-DefReg with $\lambda = 0.1$). The results reveal a complementary pattern: DefTransNet consistently out-

Table 4.7. Numerical Chamfer Distance results across deformation levels for DefTransNet and Robust-DefReg with different KL divergence weights after the second refinement iteration. This table complements Figure 4.6 by providing exact values for clearer comparison.

Def. Level	**DefTransNet**	**Robust-DefReg**	**DefTransNet** $\lambda = 0.9$	**Robust-DefReg** $\lambda = 0.1$	**Robust-DefReg** $\lambda = 0.7$
0.1	0.002125653	0.001857724	0.001555125	0.001503293	0.001501096
0.2	0.001907051	0.002234026	**0.001584741**	0.001682059	0.001678533
0.3	0.002496228	0.002559829	**0.001710409**	0.001739061	0.001731922
0.4	0.002887124	0.003061104	**0.001781984**	0.001803182	0.001801097
0.5	0.003571596	0.003608849	**0.00193007**	0.001960186	0.001957374
0.6	0.004434763	0.004746935	0.002069486	**0.002037409**	0.002050843
0.7	0.005109824	0.005557513	0.002232254	**0.002204411**	0.002242111
0.8	0.005650648	0.006306534	0.002428467	**0.002368519**	0.002434513

performs in low to moderate deformation (0.2–0.5), while Robust-DefReg achieves better results in high deformation levels (0.6–0.8). This suggests that stronger priors benefit expressive architectures that can maintain global structure, whereas lighter models perform better with weaker regularization that preserves flexibility.

Subfigure (d) zooms into the two best configurations, DefTransNet ($\lambda = 0.9$) and Robust-DefReg ($\lambda = 0.1$), to highlight this contrast. DefTransNet excels in moderate deformation levels where both structure preservation and fine alignment are crucial, while Robust-DefReg becomes favorable in extreme deformation where flexibility dominates and strong constraints may hinder effective adaptation.

These findings uncover important relationships between model architecture, regularization strength, and deformation complexity:

- ***Model capacity determines regularization need.*** Robust-DefReg lacks the global modeling power of a Transformer, making it more sensitive to over-regularization. It benefits most from mild KL divergence, particularly in large deformation scenarios where preserving adaptability is crucial.

- ***Strong KL divergence complements transformer-based models.*** In DefTransNet, strong KL regularization (e.g., $\lambda = 0.9$) aligns well with the model's architectural capacity to encode global structure. This synergy supports better alignment under moderate deformation and provides robustness across broader conditions without retuning.

- ***Transformer-based design enhances generalization.*** Although Robust-DefReg ($\lambda = 0.1$) slightly outperforms under extreme deformation, its success relies on careful tun-

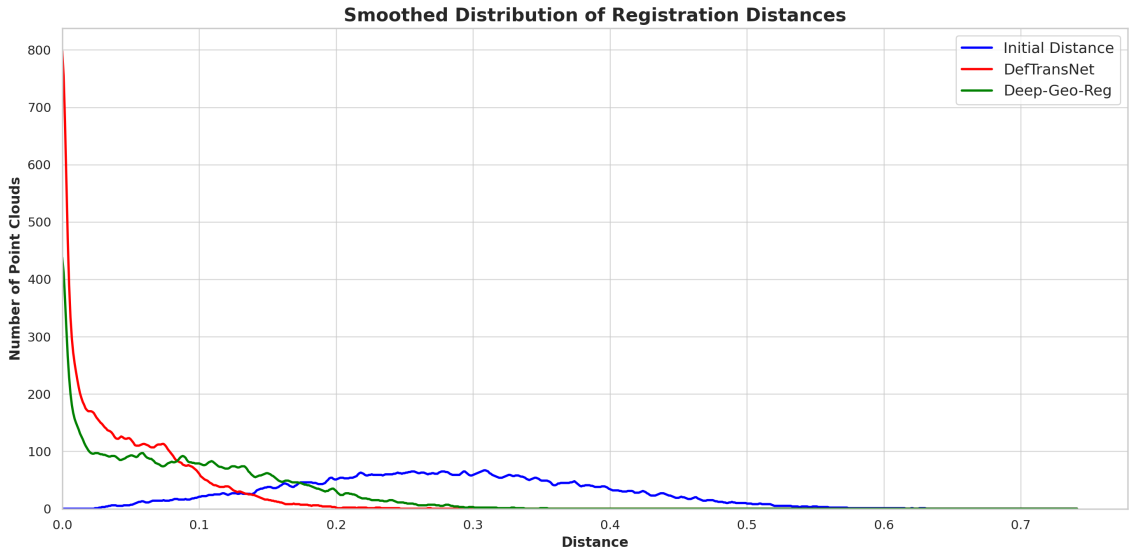


Figure 4.7. Histogram distribution of the mean Euclidean distance errors across all point clouds in the test set on the ModelNet dataset. The x-axis represents the average distance between source and target points for each point cloud, while the y-axis shows the number of point clouds per bin. Two methods are compared with initial misalignment (blue): DefTransNet (red) and Deep-Geo-Reg (green). The results demonstrate that **DefTransNet (ours)** yields the most accurate and consistent registrations, with the majority of point clouds concentrated in lower error bins.

ing of λ. By contrast, DefTransNet delivers competitive results across all deformation levels with a fixed λ, demonstrating better generalization and practical deployment potential.

Despite the visual prominence of error bars in Figures 4.4 and 4.6, their actual magnitudes are consistently small across all configurations. In both figures, the standard deviations typically remain below 0.0002–0.0003, corresponding to less than 10–15% of the mean Chamfer Distance values. This low variance confirms the stability and reliability of the registration performance across different deformation levels and regularization weights. The perceived size of some error bars is partially influenced by the non-zero baseline of the y-axis, which can exaggerate visual differences without reflecting significant statistical variation.

In conclusion, while both methods benefit from the Learning-to-Refine strategy, the Transformer-based DefTransNet offers greater reliability and consistency across varying conditions. The choice of KL divergence strength should be aligned with the model's architectural expressiveness and the deformation complexity of the task at hand. For completeness, Table 4.7 reports the exact Chamfer Distance values corresponding to the configurations shown in Figure 4.6, enabling more precise comparison across deformation levels.

4.5 Distance Distributions

This section analyzes the distance distribution of registration errors to provide insight beyond average accuracy. Rather than focusing solely on mean values, we examine how the errors are distributed across the dataset, providing a deeper understanding of each method's consistency and reliability. Figure 4.7 presents the smoothed histogram distribution of mean Euclidean distance errors for all point clouds in the test set of the ModelNet dataset. This visualization offers a statistical perspective on registration performance, highlighting not only accuracy but also distributional behavior.

The blue curve corresponds to the initial misalignment prior to any registration. This distribution is broadly spread across the error range, with a significant number of point clouds exhibiting high alignment errors. The lack of a distinct peak and the long tail toward the right illustrate the variability and inaccuracy of the raw, unaligned data, establishing a baseline for comparison.

The green curve represents the performance of the Deep-Geo-Reg baseline method. Compared to the initial state, it shows a clear shift toward lower errors, indicating that the registration has improved. However, the distribution remains relatively wide and less concentrated, suggesting that although Deep-Geo-Reg reduces error on average, it does not achieve high consistency across the entire test set.

The red curve, corresponding to our proposed method, DefTransNet, demonstrates the most favorable distribution. It is sharply peaked near the lowest error bins (typically below 0.05), with a steep decline toward higher values. This indicates that the vast majority of point clouds achieve low registration error. Moreover, the narrowness of the distribution reveals the model's robustness and consistency across diverse deformation levels and shape categories. The use of a global attention mechanism allows DefTransNet to effectively reason over long-range dependencies, resolving correspondences that traditional or geometric methods cannot reliably handle.

From a distributional perspective, DefTransNet offers:

- A clearly left-shifted peak, reflecting improved overall accuracy.
- A tight spread, suggesting low variance and high reliability across the dataset.
- Minimal high-error tailing, confirming robustness under difficult deformation scenarios.

While Deep-Geo-Reg improves upon the unregistered baseline, only DefTransNet consistently delivers low registration error across samples. This confirms that DefTransNet not only enhances average performance but also concentrates predictions within a low-error regime, offering both high accuracy and dependable outcomes across the dataset.

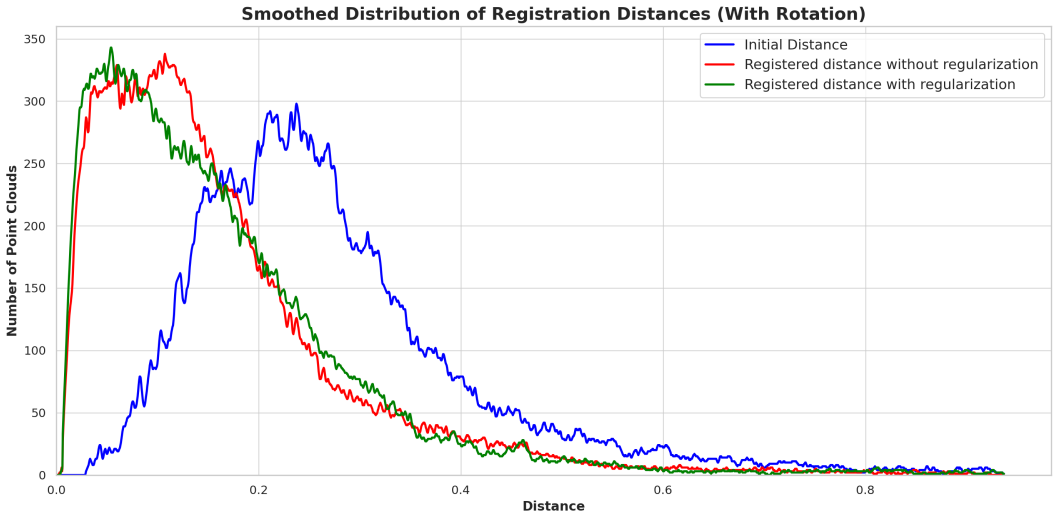


Figure 4.8. Histogram distributions of mean Euclidean registration errors across the 4DMatch test set, showing the performance of the unsupervised method by Croquet et al. [1] with and without regularization. The x-axis denotes the average registration error per point cloud, and the y-axis indicates the number of point clouds per bin. The blue curve corresponds to the initial misalignment, the red curve to the unregularized registration, and the green curve to the regularized variant.

This section analyzes the impact of regularization on registration accuracy, particularly for the unsupervised learning framework proposed by Croquet et al. [1]. Figure 4.8 presents smoothed histogram distributions of the mean Euclidean registration errors computed over the 4DMatch test set. Instead of reporting average errors alone, this visualization emphasizes the statistical distribution and consistency of performance across the dataset.

The blue curve represents the initial misalignment prior to any registration. Its broad spread and long tail toward high-error regions reflect substantial variability and the need for corrective alignment. The red curve illustrates the outcome of Croquet's method without regularization. Compared to the initial state, it shifts the distribution leftward, indicating improved accuracy. However, the distribution still remains relatively wide, with a noticeable presence of higher-error instances. This highlights the method's moderate effectiveness but also its susceptibility to noise, overlap variability, and rotational sensitivity.

The green curve represents the regularized variant of the same method. While it achieves a comparable leftward shift in the central mass of the distribution, the spread is slightly reduced, and the peak becomes more pronounced. This suggests that regularization improves not just the mean accuracy but also the robustness of the method, leading to more consistent results across samples. Nevertheless, the overlap with the unregularized curve shows that the benefit is incremental rather than transformative.

Overall, this distributional analysis confirms that the regularization proposed by Croquet et al. [1] offers modest improvements in registration accuracy and stability, but does not fully eliminate the variability inherent in unsupervised learning under challenging conditions like

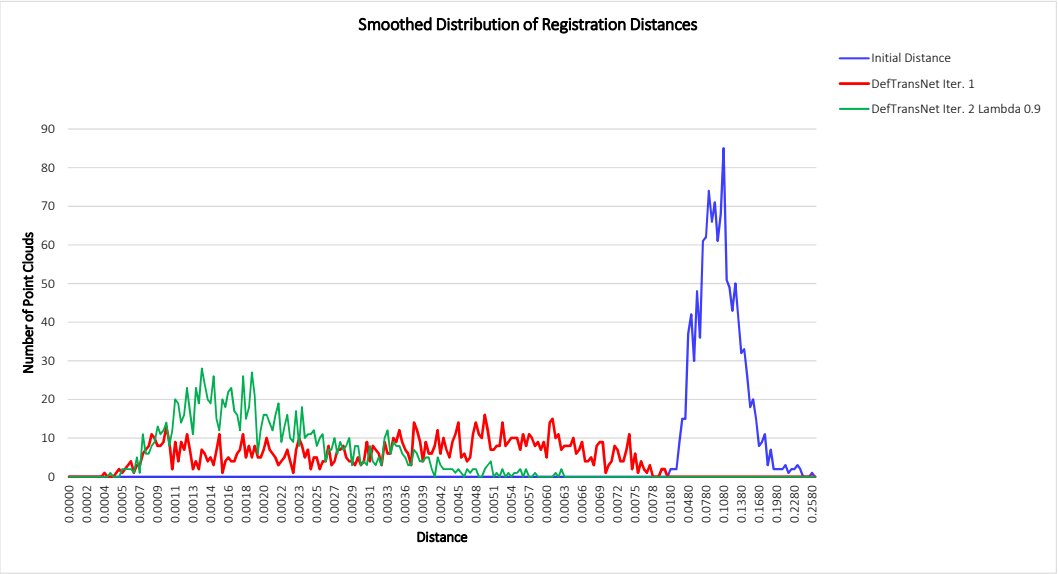


Figure 4.9. Smoothed distribution of mean Euclidean distance errors across all point clouds in the ModelNet test set. The x-axis represents the average registration error per point cloud, while the y-axis shows the number of point clouds per bin. Three stages are visualized: initial misalignment (blue), DefTransNet after the first iteration (red), and DefTransNet after the second iteration with KL divergence regularization ($\lambda = 0.9$, green). Note that the x-axis bins are non-uniform to better highlight performance differences across lower error ranges.

those in 4DMatch.

From a distributional perspective, the method of Croquet et al. [1] exhibits:

- A moderate left-shift in both regularized and unregularized variants, indicating some improvement over the initial misalignment.
- A wider spread, particularly in the unregularized version, suggesting variability in performance and sensitivity to input conditions.
- A slightly more concentrated peak in the regularized variant, pointing to increased consistency but still limited robustness under challenging registration scenarios.

Figure 4.9 presents the smoothed distributions of mean Euclidean registration errors for all point clouds in the ModelNet test set. Each curve represents a different stage in the registration pipeline: the blue line corresponds to the initial unaligned input; the red line shows results after the first application of DefTransNet; and the green line represents the refined results after a second iteration with KL-divergence regularization ($\lambda = 0.9$).

To better visualize the error behavior in low-error regions, where most improvements occur, the x-axis uses non-uniform bin spacing. This design choice enhances the visibility of differences in the critical low-error regime, which would otherwise appear compressed on a linear or uniform scale.

The initial distribution (blue) exhibits a broad and right-skewed shape, with a high number of samples having relatively large registration errors. After one iteration of DefTransNet (red), the curve shifts significantly to the left, indicating a substantial reduction in mean error. However, some spread remains, suggesting residual misalignments for a subset of samples. The second iteration (green) yields the most favorable distribution, narrow, sharply peaked near the origin, and with minimal presence in higher error bins. This reflects not only improved mean performance but also higher reliability and generalization across the dataset.

- Clear leftward shift: Each stage reduces the registration error, as seen by the progressive shift in the distribution's peak.
- Improved compactness: The second iteration concentrates predictions into a narrower, low-error regime.
- Non-uniform x-axis bins: This choice enhances visibility in low-error regions, where most of the differentiation occurs.

These findings demonstrate the strength of the Learning-to-Refine framework. Iterative refinement, especially with probabilistic regularization, not only lowers average registration errors but also enhances consistency, producing accurate and stable alignments across varied input shapes and deformation conditions.

4.6 Ablation Study

To evaluate the specific contribution of each architectural component in our registration framework, we conduct an ablation study using five model variants. Each variant incrementally removes a key module, T-Net, Transformer, or both, while keeping the rest of the architecture intact. All models are evaluated across eight deformation levels (0.1 to 0.8) using the average point-wise Euclidean distance. The results, illustrated in Figure 4.10, help quantify how each module contributes to robustness and accuracy under increasing geometric complexity.

Deep-Geo-Reg (baseline). This non-learning baseline does not include any of the proposed modules. It lacks feature learning, global alignment, or deformation modeling. As expected, its performance declines steeply with increasing deformation, from 0.00148 at level 0.1 to 0.1354 at level 0.8, highlighting its inability to adapt to non-rigid transformations. This serves as a lower bound for comparison.

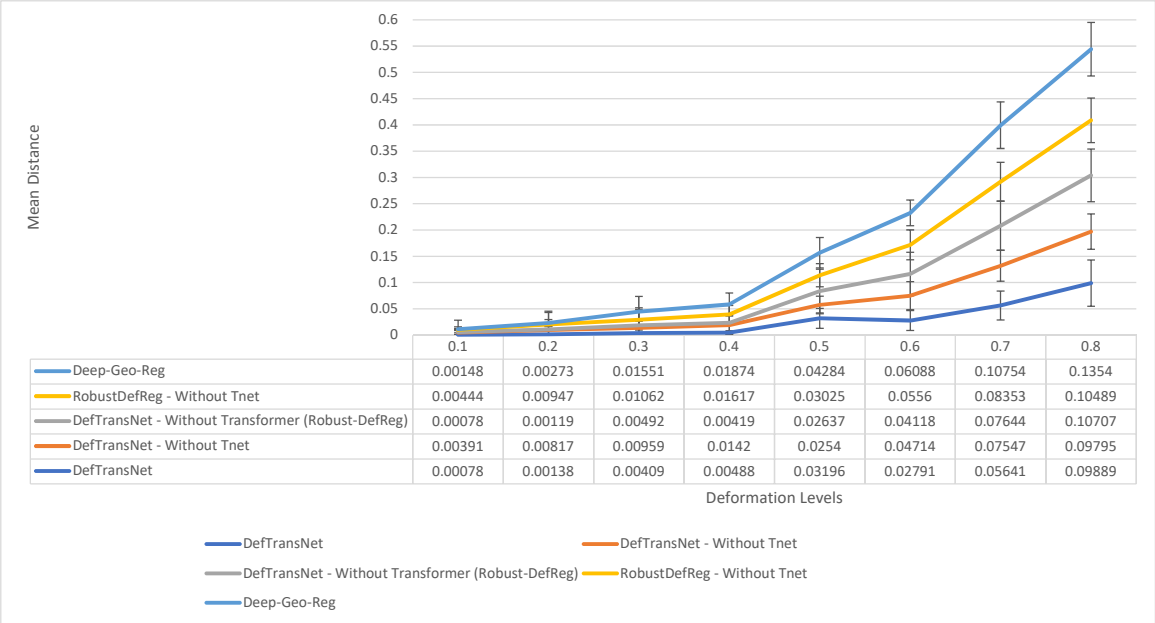
DefTransNet without T-Net vs. DefTransNet. The T-Net is designed to estimate a global alignment transformation that coarsely aligns the source and target point clouds before learning finer deformations. When the T-Net is removed, as in the "DefTransNet without T-Net" variant, we observe a clear performance drop across all deformation levels (e.g., 0.00391 vs. 0.00078 at level 0.1, and 0.09795 vs. 0.09889 at level 0.8). Although the Transformer can partially compensate at high deformation levels, the absence of an initial global alignment increases the learning burden on downstream modules. Without T-Net, the network starts training from a misaligned state, leading to slower convergence and suboptimal optimization, particularly at low-to-moderate deformation levels where coarse alignment is most effective.

RobustDefReg without T-Net vs. Robust-DefReg. A similar pattern is seen in the graph-based setting: Adding the T-Net to Robust-DefReg improves accuracy across the board. For example, at level 0.4, the error drops from 0.01617 (without T-Net) to 0.00419 (with T-Net). This confirms that even in architectures without global attention, a learnable alignment prior significantly helps reduce the initial displacement and allows the model to focus on learning local residuals.

Robust-DefReg vs. DefTransNet. The Transformer enables the model to capture long-range dependencies and global context across the point cloud. Removing this module reduces the model to relying only on local geometric learning and global alignment (T-Net + EdgeConv). As a result, while the model performs well at low deformation levels (0.00078 at 0.1), it deteriorates significantly at higher levels (0.10707 at 0.8), where non-local displacements dominate. The lack of self-attention makes it difficult to resolve feature ambiguity and correspondences that lie far apart spatially but are semantically similar, a challenge common in large deformations.

Deep-Geo-Reg vs. all learning-based models The baseline model performs the worst under all deformation levels, confirming the critical role of learned features. The EdgeConv-based GCNN enables each point to aggregate local neighborhood information, enhancing robustness to noise and partial deformation. All models using EdgeConv (RobustDefReg, DefTransNet variants) significantly outperform Deep-Geo-Reg, particularly at high deformation (e.g., 0.10489 vs. 0.1354 at level 0.8), validating the importance of local geometric learning in capturing fine-grained shape changes.

DefTransNet (full model) By integrating T-Net, GCNN (EdgeConv), and Transformer modules, DefTransNet combines global alignment, local structure learning, and global context modeling. This synergy results in the most stable and accurate registration performance across all deformation levels, demonstrated by minimal error fluctuations (e.g., 0.00078 to 0.09889) and no steep error escalation. The architecture benefits from the interplay of three key functional components:



	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8
Deep-Geo-Reg	0.00148	0.00273	0.01551	0.01874	0.04284	0.06088	0.10754	0.1354
RobustDefReg - Without Tnet	0.00444	0.00947	0.01062	0.01617	0.03025	0.0556	0.08353	0.10489
DefTransNet - Without Transformer (Robust-DefReg)	0.00078	0.00119	0.00492	0.00419	0.02637	0.04118	0.07644	0.10707
DefTransNet - Without Tnet	0.00391	0.00817	0.00959	0.0142	0.0254	0.04714	0.07547	0.09795
DefTransNet	0.00078	0.00138	0.00409	0.00488	0.03196	0.02791	0.05641	0.09889

Figure 4.10. Ablation study comparing the mean registration error across deformation levels. Our full model (DefTransNet) consistently outperforms all ablated variants and the baseline (Deep-Geo-Reg). Each curve corresponds to a model configuration, and lower values indicate better alignment.

- A ***global alignment initializer***, which provides a coarse pre-alignment between source and target point clouds. This helps reduce large initial displacements, enabling the network to converge more efficiently by narrowing the search space during optimization.

- A ***local neighborhood feature extractor***, which learns spatially-aware descriptors by aggregating geometric information from nearby points. This component enhances the model's sensitivity to fine-grained deformations and structural variations, especially in moderately distorted regions.

- A ***global context learning mechanism***, which models long-range dependencies across the point cloud. By capturing relationships between distant but semantically related regions, it helps resolve ambiguities in correspondence and improves robustness under severe, non-local deformation.

This ablation study reveals the functional necessity of each of these components:

- Removing the ***global alignment initializer*** results in poor initialization, which can lead to slower convergence and reduced accuracy, particularly under small to moderate deformations where coarse alignment is crucial.

- Disabling the ***global context learning*** module limits the network's ability to capture long-range structural relationships. As a result, the model fails to generalize under

high-deformation settings where local geometry alone is insufficient.

- Eliminating all learning components and relying purely on geometric heuristics, as in the classical baseline, leads to a dramatic performance drop across all conditions, underscoring the importance of feature learning in non-rigid registration.

The complete model demonstrates that these three components act in a complementary manner: The global initializer ensures a good starting point, the local feature extractor captures detailed deformations, and the global learning unit facilitates high-level structural understanding. When combined, they enable the model to effectively manage the hierarchical complexity of non-rigid PCR under a wide range of deformation scenarios.

Chapter 5

Discussion

This chapter provides a critical discussion of the three non-rigid PCR methods proposed in this thesis: Robust-DefReg, DefTransNet, and the iterative Learning-to-Refine strategy. Building upon the quantitative and qualitative findings reported in Chapter 4, we analyze the behavior of these methods under varying deformation conditions and discuss their respective advantages and limitations.

We begin with a high-level visual analysis that summarizes the overall performance of each method across deformation levels using accuracy–robustness diagrams. These compact representations offer an interpretable overview of how different methods compare in terms of both registration quality and stability, complementing the detailed level-wise evaluations previously presented.

The remainder of this chapter is structured around two perspectives, potential and limitations, for each of the three proposed techniques. Finally, we conclude with a section on further development, identifying promising directions for future research and refinement.

5.1 Comparative Analysis of Accuracy and Robustness

Before delving into the individual strengths and limitations of the proposed methods, we begin with a high-level visual analysis that synthesizes performance across deformation levels. To this end, we introduce a compact and interpretable representation of each method's accuracy and robustness using scatter plots for three benchmark datasets: SynBench, ModelNet, and DeformedTissue. These plots serve as an entry point into the discussion by highlighting the overall positioning of each method in terms of both registration quality and stability under deformation. By summarizing results along two key dimensions, mean error and robustness, this section complements the detailed level-wise evaluations in Chapter 4 and offers a holistic view of each method's behavior across diverse deformation scenarios.

Accuracy and robustness computation. To visualize the performance of each method

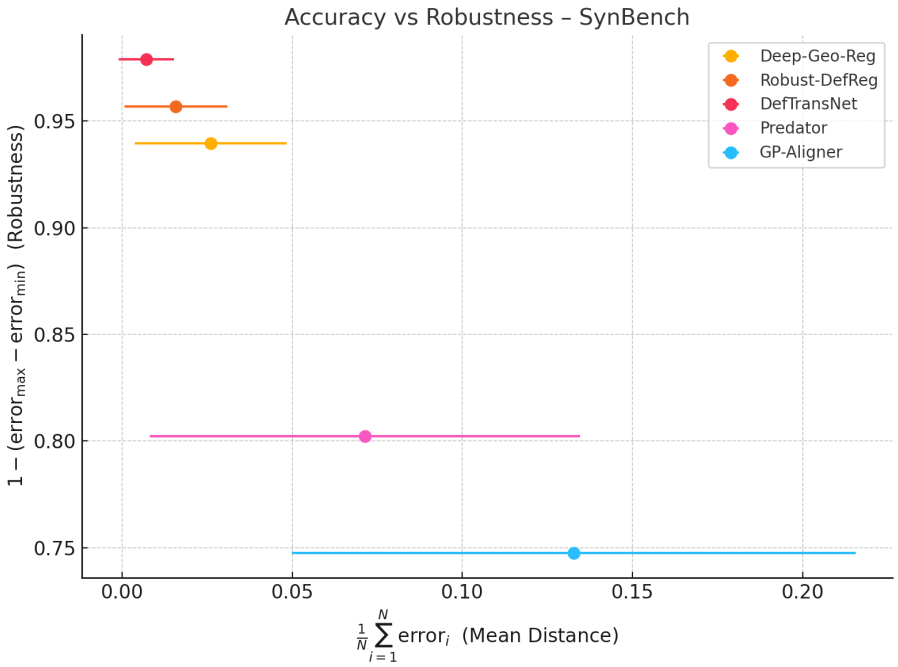


Figure 5.1. Accuracy–robustness summary for SynBench. DefTransNet shows the lowest average error and the highest robustness with minimal variation, outperforming all baseline methods. Robust-DefReg follows closely, while classical approaches show poor consistency and accuracy under deformation. For statistical comparisons, see Table 5.1.

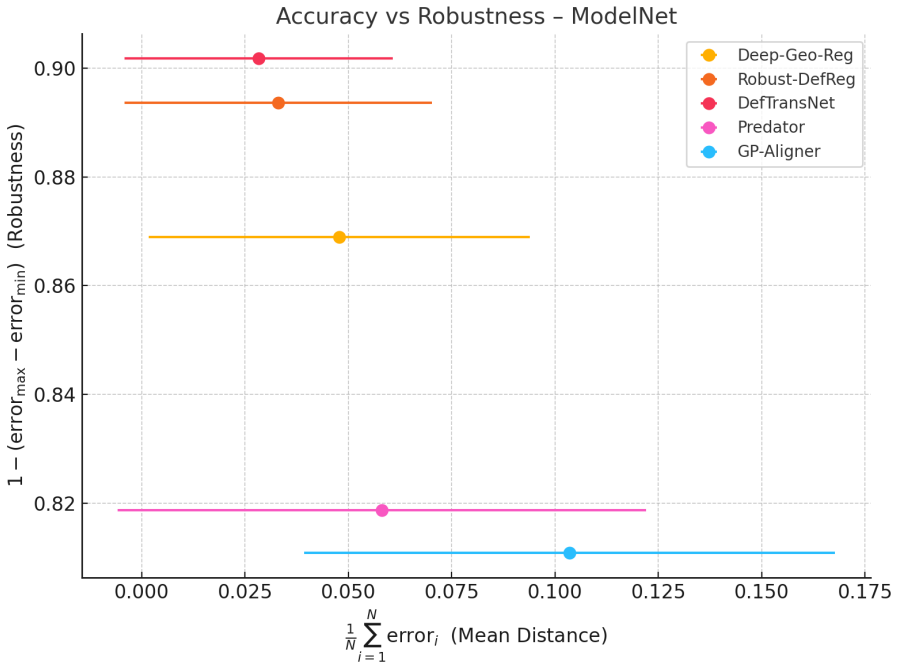


Figure 5.2. Accuracy–robustness summary for ModelNet. While DefTransNet maintains the best overall performance, the gap to Robust-DefReg is narrower than in SynBench. This reflects the simpler geometric structure of ModelNet objects, where both global and local methods perform comparably well. For statistical comparisons, see Table 5.1.

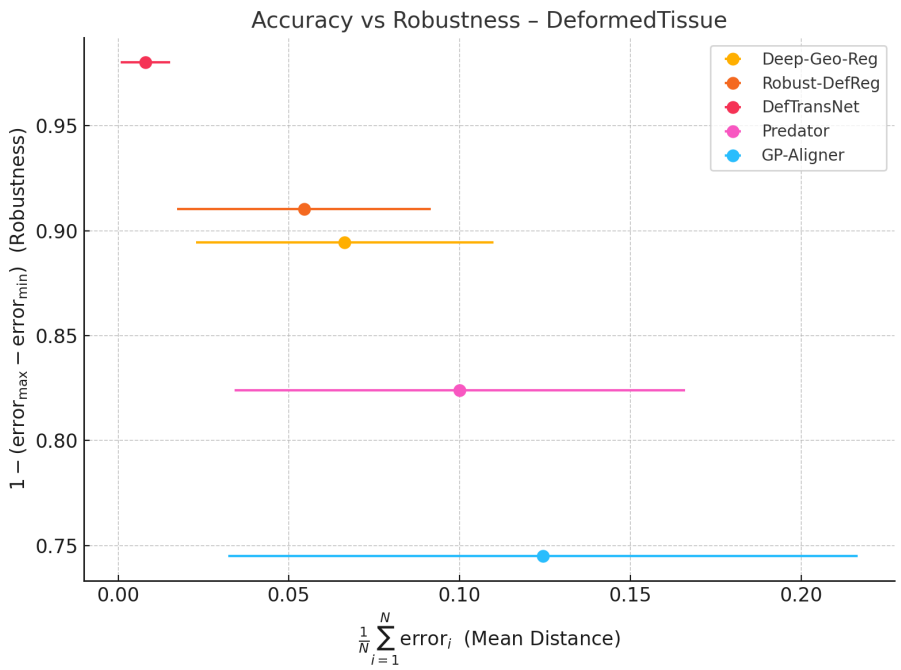


Figure 5.3. Accuracy–robustness summary for DeformedTissue. DefTransNet substantially outperforms other methods, especially under anatomically realistic deformation. The model achieves high robustness and low error, demonstrating strong generalization to real-world, irregular deformation. For statistical comparisons, see Table 5.1.

across varying deformation levels, we construct accuracy–robustness plots. The following metrics are used to evaluate and visualize performance:

Mean distance (X-axis). The average registration error across all deformation levels is computed as

$$\text{Mean Distance} = \frac{1}{N}\sum_{i=1}^{N}\text{error}_i \tag{5.1}$$

where N denotes the total number of deformation levels.

Robustness (Y-axis). Robustness is defined as the inverse of the absolute error spread between the lowest and highest deformation levels:

$$\text{Robustness} = 1 - (\text{error}_{\text{max}} - \text{error}_{\text{min}}) \tag{5.2}$$

This formulation avoids instability when the lowest error is near zero and better reflects consistency across the deformation spectrum.

Error variation (Error Bars). The horizontal error bars represent the standard deviation of the registration error across all deformation levels:

$$\text{Error Bar} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\text{error}_i - \bar{\text{error}})^2} \tag{5.3}$$

Table 5.1. Paired t-test results comparing DefTransNet with classical and learning-based baselines across all datasets. A result is considered statistically significant if $p < 0.05$.

Dataset	Compared Method	t-test p-value	Significant
SynBench	Deep-Geo-Reg	0.011658121	Yes
SynBench	Predator	0.018424738	Yes
SynBench	GP-Aligner	0.003064767	Yes
ModelNet	Deep-Geo-Reg	0.017692665	Yes
ModelNet	Predator	0.047145691	Yes
ModelNet	GP-Aligner	0.000911098	Yes
DeformedTissue	Deep-Geo-Reg	0.008273284	Yes
DeformedTissue	Predator	0.008743768	Yes
DeformedTissue	GP-Aligner	0.015281038	Yes

where $\bar{\text{error}}$ is the mean registration error across all levels.

Statistical significance tests. To quantify the reliability of the observed performance differences, we conducted a paired t-test, a parametric statistical test used to compare the means of two related groups under the assumption of normality. Table 5.1 presents the results, where significance is determined using $p < 0.05$.

All comparisons between DefTransNet and the baselines result in p-values below the 0.05 threshold, indicating statistically significant differences. This confirms that the superior performance observed visually in Figures 5.1–5.3 is not due to random variation but reflects meaningful and consistent improvements. Table 5.1 presents the results of paired t-tests comparing DefTransNet with each baseline method (Deep-Geo-Reg, Predator, and GP-Aligner) across all three benchmark datasets. For each pair, the test evaluates whether the observed performance improvements of DefTransNet are statistically significant, that is, unlikely to have occurred by chance. Specifically:

- On the SynBench dataset, DefTransNet significantly outperforms all classical and learning-based baselines, with particularly low p-values against GP-Aligner.
- On ModelNet, although the geometric structure is simpler, the differences remain statistically significant, reinforcing that even in less challenging scenarios, DefTransNet maintains a reliable edge.
- On DeformedTissue, which involves real-world anatomical deformation, the significance of all comparisons emphasizes DefTransNet's practical generalizability.

These findings further support the hypothesis that integrating global attention with localized geometric features not only improves average accuracy but also ensures consistently

better performance across deformation scenarios.

SynBench: Global learning improves robustness. On the SynBench dataset, as shown in Figure 5.1, the proposed DefTransNet clearly outperforms all baseline methods, exhibiting both the lowest average error and the highest robustness. This is reflected by its top-left position in the plot, with short error bars indicating minimal performance variation across deformation levels. Robust-DefReg shows relatively strong performance but greater sensitivity to increasing deformation. Deep-Geo-Reg, Predator, and GP-Aligner perform poorly in both accuracy and robustness, especially under large deformation. This visualization supports our hypothesis that the integration of local and global features, specifically through the Transformer-based design of DefTransNet, enhances the model's ability to handle both fine-grained and long-range correspondences under synthetic deformation.

ModelNet: Diminishing returns in structured geometry. In the ModelNet benchmark, as shown in Figure 5.2, DefTransNet maintains top performance, though the gap to Robust-DefReg narrows. This suggests that in structured geometric datasets, where object categories exhibit clear and repeatable shapes, the added capacity of global attention contributes less to the overall registration accuracy. Nonetheless, DefTransNet's consistency and reduced variability still demonstrate its advantage over classical and feature-matching baselines. These results confirm that while Transformer-based models may be over-parameterized for simple geometries, they remain at least as good, if not better, than simpler methods, even when their full capacity is not fully required.

DeformedTissue: Superior robustness in real-world deformation On the DeformedTissue dataset, as shown in Figure 5.3, which contains real-world, non-linear anatomical deformation, DefTransNet again stands out as the most accurate and most robust method. The wide gap between DefTransNet and all other baselines underlines its ability to generalize beyond synthetic benchmarks and into real-world scenarios. Its robustness remains high, and error bars remain short, showing that performance does not degrade substantially even under irregular and complex deformations. This validates the architecture's suitability for real applications such as soft tissue tracking or surgical navigation, where deformation patterns are unpredictable and spatial context is essential.

5.2 Robust-DefReg: GCNN-Based Method

Robust-DefReg, introduced in Chapter 3.2, was a graph-based non-rigid registration framework that combined local geometric learning with a coarse-to-fine alignment strategy. By leveraging EdgeConv for local feature extraction and T-Net for spatial normalization, the method was designed to handle moderate deformations while maintaining computational efficiency. Its modular architecture allowed for clear interpretation and adaptability, and

served as a strong baseline for further refinement through the Learning-to-Refine strategy. Below, we discuss the key strengths and limitations of Robust-DefReg based on the evaluation results.

5.2.1 Potential

The following discussion highlights the main strengths of Robust-DefReg, focusing on its coarse-to-fine registration, stability under low-to-medium noise and deformation, modularity and interpretability, and its capacity for effective refinement.

Coarse-to-fine registration. Robust-DefReg is designed around a hierarchical, coarse-to-fine strategy that allows it to handle complex deformation by progressively refining local geometric correspondences. The use of EdgeConv enables the model to encode neighborhood-level structure, capturing intricate shape deformations in a local region. This makes it well-suited for scenarios where fine-scale surface changes dominate. During training, this architecture allows Robust-DefReg to first identify global structure through coarse alignment and then correct local misalignments through finer updates, improving stability and convergence.

Stability under low-to-medium noise and deformation. In synthetic datasets like SynBench and ModelNet, as well as real-world scenarios such as DeformedTissue, Robust-DefReg consistently demonstrates strong performance in the presence of mild to moderate deformation (e.g., up to level 0.5). The method maintains relatively low registration errors even when exposed to Gaussian noise or partial outliers, indicating that its local feature aggregation and spatial learning strategy offers inherent robustness. This resilience makes it a viable choice in real-world environments where data may be imperfect but still structurally meaningful.

Modularity and interpretability. The architecture of Robust-DefReg is composed of interpretable and modular components, EdgeConv for local topology, T-Net for spatial normalization, and a graph-based correspondence mapping mechanism. The absence of a Transformer makes the model more transparent and allows for more straightforward debugging, tuning, and architectural ablation. Each module serves a distinct role, and this separation of concerns makes the system easier to analyze, maintain, and extend for specific applications or domain adaptations.

Effective refinement. Although the model lacks a global learning module, Robust-DefReg benefits significantly from iterative refinement when paired with the Learning-to-Refine strategy. Particularly with light KL divergence regularization ($\lambda = 0.1$), the model improves its alignment even at higher deformation levels (e.g., 0.7–0.8), where a single pass would struggle. The low-regularization setting allows the graph-based method to re-

tain flexibility while benefiting from probabilistic constraints, confirming that the method can be extended to more challenging tasks when embedded in a self-corrective loop.

5.2.2 Limitation

Despite its advantages, Robust-DefReg also exhibited several limitations, particularly related to its limited global learning, performance saturation under large deformation, and sensitivity to KL regularization during iterative refinement.

Limited global learning. Robust-DefReg, by design, does not incorporate a mechanism for capturing long-range dependencies across the point cloud. As a result, it struggles when deformation extends beyond the local neighborhood, such as when large object parts shift globally or rotate independently. This limitation becomes evident in scenarios with significant shape deformation, sparse overlap, or strong rotation, where purely local learning fails to establish meaningful global correspondences.

Performance saturates under large deformation. While the model performs well under moderate deformation, its accuracy deteriorates noticeably at higher levels. On both synthetic and real-world datasets, when deformation exceeds 0.6, error rates increase more sharply than for Transformer-based models. This suggests that Robust-DefReg lacks the capacity to generalize across samples with highly non-linear structural changes, and its coarse-to-fine mechanism alone is insufficient to handle large-scale shifts in object geometry.

KL sensitivity. When embedded in the Learning-to-Refine framework, the performance of Robust-DefReg becomes sensitive to the choice of regularization strength. Specifically, using high KL divergence weights ($\lambda > 0.3$) introduces over-regularization, constraining the model's flexibility and impairing its ability to explore plausible deformation spaces. This contrasts with DefTransNet, which benefits from stronger regularization. For graph-based methods, such constraints can stifle performance, especially in regions requiring adaptive deformation.

5.3 DefTransNet: Transformer-Based Method

DefTransNet, presented in Chapter 3.2, was a Transformer-based registration framework designed to capture both local geometry and global contextual relationships within non-rigidly deformed point clouds. It extended the architecture of Robust-DefReg by incorporating a self-attention mechanism, enabling long-range feature interaction and improved generalization under complex deformations. Combined with T-Net and EdgeConv modules, DefTransNet demonstrated strong robustness across various challenging conditions. In the following, we examine its main potential and limitations as observed through the experi-

mental results.

5.3.1 Potential

The strengths of DefTransNet are reflected in its superior accuracy and generalization, its ability for global context modeling, strong robustness to perturbations, consistent and reliable performance across diverse conditions, and its support for scalable refinement through iterative learning.

Superior accuracy and generalization. DefTransNet consistently outperforms all existing baselines and proposed alternatives across a broad spectrum of experimental settings. Whether under synthetic conditions (SynBench, ModelNet) or real-world deformations (DeformedTissue, 4DMatch), and whether subject to large rotation, noise, or sparse overlap, it maintains the lowest registration errors. Its high accuracy does not come at the cost of overfitting, as evidenced by the tight error distributions observed in multiple histogram analyses, indicating strong generalization to diverse, unseen data.

Global context modeling. The Transformer module is the core innovation of DefTransNet, enabling it to reason over the entire point cloud context. Unlike graph-based or correspondence-based methods, the attention mechanism allows each point to be informed by both its local neighborhood and distant regions of the shape. This proves especially powerful in settings with missing data, sparse overlap, or long-range dependencies, where local methods fail. Through its self-attention layers, DefTransNet learns not just point-wise features but also higher-order relationships, enabling precise alignment even when part correspondences are ambiguous.

Robustness to perturbations. DefTransNet exhibits exceptional resilience under various perturbations. In scenarios involving significant noise (up to $\sigma = 0.05$), high outlier ratios (up to 45%), or minimal overlap (as low as 10%), the method still produces stable and accurate results. These conditions simulate real-world constraints such as occlusion, measurement errors, or partial visibility. The robustness likely stems from the network's ability to filter out irrelevant input regions and concentrate attention on structurally meaningful areas during registration.

Consistent and reliable. Histogram analyses of registration errors show that DefTransNet not only lowers average error but also reduces variance across test samples. The model consistently clusters outputs in the low-error bins, with very few outliers or catastrophic failures. This consistency is critical for applications in clinical or industrial environments where reliability across different cases is essential. The structural regularity captured by the Transformer and stabilized by the T-Net ensures that the model performs predictably across varied geometric conditions.

Scalable refinement. When integrated with the Learning-to-Refine strategy, DefTransNet continues to improve in performance. With a high KL divergence weight ($\lambda = 0.9$), it maintains coherence across iterations and avoids overfitting to noisy pseudo-labels. This setup enables the model to incrementally correct registration errors, particularly under high deformation, by enforcing a probabilistic prior over plausible deformation fields. The synergy between global attention and probabilistic regularization makes the model both expressive and controlled.

5.3.2 Limitation

Despite its strong performance, DefTransNet also presents some limitations, including high model complexity, dependence on strong priors for effective refinement, and sensitivity to unnormalized or misaligned input data.

High model complexity. A significant drawback of DefTransNet is its computational cost. Transformer-based architectures inherently involve quadratic complexity with respect to the number of points due to the self-attention mechanism. As a result, the model requires more memory, longer training times, and higher computational resources compared to graph-based alternatives. This limits its practicality in real-time or embedded applications, such as surgical navigation or mobile robotics, where inference speed and resource efficiency are critical.

Dependent on strong priors for refinement. Although DefTransNet benefits greatly from iterative training, it is highly dependent on well-calibrated regularization. Without KL divergence during Learning-to-Refine, the model can overfit to inaccurate pseudo-labels, particularly at higher deformation levels where the initial registration may contain substantial errors. This suggests that while the architecture is expressive, it needs the additional guidance of a deformation prior to avoid drifting into implausible or unstable deformation spaces during training.

Sensitivity to unnormalized input. While the inclusion of T-Net mitigates input irregularities, the model still exhibits some sensitivity to misalignment or inconsistent point distribution. If the input is highly skewed or lacks a meaningful canonical orientation, early attention layers may struggle to converge to optimal alignment, especially when operating in low-overlap or noisy settings. Preprocessing, such as normalization or canonicalization, may still be required to achieve optimal performance.

5.4 Learning-to-Refine: Iterative Refinement Approach

Learning-to-Refine, introduced in Chapter 3.2, was developed as an iterative training strategy aimed at progressively improving registration accuracy without relying on ground truth correspondences. By reusing predicted deformation fields as pseudo-labels and incorporating a KL divergence-based probabilistic prior, the method guided both DefTransNet and Robust-DefReg toward more coherent and stable solutions over multiple refinement stages. The following discussion explores its key advantages and limitations observed during evaluation.

5.4.1 Potential

The key advantages of Learning-to-Refine lie in its ability to enable progressive improvement of registration results, its incorporation of uncertainty through probabilistic regularization, its flexibility to support different network architectures, and its scalability in scenarios where ground truth annotations are unavailable.

Progressive improvement. The Learning-to-Refine strategy introduces an iterative training loop that improves registration quality over successive passes. Each iteration uses the output deformation field as a pseudo-label for the next, enabling the model to refine its understanding of the underlying geometry. Across both DefTransNet and Robust-DefReg, this leads to a consistent reduction in Chamfer and Euclidean distances, with each iteration yielding a more accurate and coherent registration outcome. The framework helps mitigate initial misalignments and allows for correction over time, particularly beneficial in high-deformation regimes.

Incorporation of uncertainty. The addition of a KL divergence regularization term enables the model to learn deformation fields that are not only precise but also statistically plausible. This probabilistic prior constrains the learned displacement vectors to lie within a structured latent space, avoiding erratic or overconfident mappings. It also enables the network to encode geometric uncertainty, which is critical when operating under noisy, incomplete, or ambiguous input conditions. Higher λ values enforce stronger priors, providing global coherence especially in Transformer-based architectures.

Flexible framework. A key strength of Learning-to-Refine lies in its compatibility with both graph-based and transformer-based networks. While DefTransNet benefits from stronger regularization, Robust-DefReg requires lighter priors, showing that the refinement strategy can adapt to the underlying model's expressive capacity. This flexibility allows Learning-to-Refine to be viewed as a meta-algorithm for improving any base registration method, provided it is capable of producing pseudo-labels and supporting deformation regularization.

Scalable without ground truth. One of the most important practical advantages is that Learning-to-Refine operates effectively in semi-supervised or fully unsupervised regimes. By using predictions from previous iterations as pseudo-labels, it circumvents the need for dense ground truth correspondence, a major limitation in medical and real-world datasets. This allows the model to bootstrap its learning process and accumulate accuracy over time without relying on costly annotations.

5.4.2 Limitation

Despite its effectiveness, Learning-to-Refine also has several limitations, including sensitivity to parameter tuning (especially the KL divergence weight), the potential accumulation of errors from inaccurate pseudo-labels, and the fact that performance gains vary depending on the underlying model architecture.

Parameter sensitivity. The method is sensitive to the choice of the KL divergence weight λ, which controls the strength of the probabilistic prior. Incorrect tuning can either make the model too rigid (over-constrained) or too flexible (under-regularized), degrading performance. Furthermore, the optimal λ differs between model types and dataset complexity, requiring cross-validation or empirical testing for best results.

Accumulation of errors. The refinement framework is built upon pseudo-labels generated from previous iterations. If the initial predictions are of low quality, e.g., due to sparse overlap or large initial misalignment, these pseudo-labels may reinforce incorrect correspondences, creating a feedback loop of accumulated errors. While KL regularization mitigates this risk, it does not eliminate it entirely, especially if the early iterations are poorly conditioned.

Model-specific gains. The degree of improvement introduced by Learning-to-Refine is model-dependent. While DefTransNet consistently benefits from the framework across all conditions, Robust-DefReg shows improvements primarily at high deformation levels and with carefully tuned regularization. This variability indicates that the strategy is not universally optimal and that its benefits must be evaluated in the context of specific architectural and data constraints.

5.5 Further Developments

Building upon the findings of this thesis, several promising directions emerge for extending and improving non-rigid PCR methods. These directions span architectural innovations, probabilistic modeling, multi-modal integration, and practical deployment optimization.

Enhancing hybrid architectures beyond DefTransNet. DefTransNet already demon-

strates the potential of hybrid architectures by combining EdgeConv-based local feature extraction with global self-attention mechanisms. This design effectively leverages neighborhood-level detail and long-range spatial learning , offering strong generalization under challenging deformation conditions. However, future research could further enhance this paradigm by exploring tighter and more dynamic coupling between the graph and attention modules. For example, instead of using static EdgeConv followed by Transformer layers, future models could implement joint graph-attention blocks, where attention scores are modulated by local geometric affinities, or use graph-attentional message passing to unify the two operations in a shared representation space. Moreover, multi-scale graph hierarchies could be aligned with multi-head attention patterns to better encode both coarse and fine structures in an anatomically consistent way. Such improvements may increase both efficiency and alignment precision, especially in low-resolution or partially missing regions.

Learning richer deformation priors via generative models. The current use of KL divergence in Learning-to-Refine assumes a fixed, often isotropic Gaussian prior over deformation fields, which may be too simplistic for capturing the complex, non-linear patterns observed in real anatomical deformations. To address this, future work could leverage deep generative models such as Variational Autoencoders (VAEs), normalizing flows, or diffusion models to learn deformation priors directly from data. For instance, a VAE trained on a large collection of plausible deformation fields could encode typical motion patterns in soft tissue or mechanical surfaces, which can then be used to guide the registration model during refinement. Normalizing flows would allow for exact likelihood modeling and reversible sampling, offering tighter control over the prior distribution. These learned priors would enable the network to model multi-modal, structured deformation behaviors, leading to more realistic and constrained registration outcomes, especially in ambiguous regions.

Cross-Modal supervision and anatomically informed constraints. In medical contexts, relying solely on geometric correspondences may be insufficient, particularly in cases with partial views, occlusion, or tissue resection. Future research could explore the integration of cross-modal information, such as preoperative MR or CT imaging, intraoperative endoscopic video, or ultrasound scans, to provide auxiliary supervision during registration. These modalities can offer anatomical context or segmentation-based landmarks that supplement the point cloud data. Moreover, incorporating biomechanical simulation data or known anatomical constraints (e.g., joint limits, volume preservation, tissue elasticity) can serve as priors during both training and inference. Embedding such constraints into the loss function or architecture, through physics-informed neural networks or energy-based regularizers, may enhance the plausibility, stability, and interpretability of the predicted deformations.

Model compression and real-time optimization. For deployment in real-world applications such as image-guided surgery or mobile robotics, registration methods must operate under strict latency and hardware constraints. Transformer-based models, while powerful, are computationally expensive due to their quadratic scaling with respect to the number of points. Future work should explore methods such as sparse attention (e.g., Linformer, Performer, or windowed attention), low-rank projection layers, and point pruning techniques to reduce model size and computational complexity. Knowledge distillation, where a smaller ”student” model is trained to mimic a larger ”teacher” network, could further compress the model without significant loss in accuracy. Additionally, real-time online refinement mechanisms could be developed to incrementally update the deformation field as new data arrives, enabling interactive feedback in time-sensitive scenarios. Combining these strategies could make high-performing models like DefTransNet suitable for real-time clinical deployment or embedded systems.

Advanced representation learning for deformation modeling. Beyond the architectural improvements explored in this thesis, several cutting-edge techniques in representation learning present exciting opportunities for advancing non-rigid PCR. Recent advances, such as diffusion models and flow matching, offer powerful generative formulations for modeling complex deformation trajectories, which could replace or augment current refinement modules. Normalizing flows and invertible neural networks enable exact likelihood modeling and bidirectional mapping between deformation spaces, making them ideal for learning reversible and structure-preserving transformations. Furthermore, metric flow learning could be employed to directly learn geodesic-consistent deformation fields, which may improve robustness in topology-altering scenarios. As alternatives to attention-based architectures, state space models offer scalable sequence modeling and have shown promise in long-range learning tasks, potentially enabling lower-latency deformation tracking. Finally, recent efforts in adapting foundation models to 3D data suggest that pre-trained cross-task embeddings could serve as powerful priors or feature extractors for registration, especially in low-data or zero-shot settings. Integrating such trends could pave the way toward more generalizable, interpretable, and data-efficient registration frameworks.

Benchmark expansion and robustness evaluation. Finally, future work should consider expanding current evaluation protocols to cover a broader spectrum of real-world variability. This includes testing across more anatomical regions, deformation types (e.g., pathological swelling, surgical cutting), sensor modalities (e.g., Lidar, structured light, RGB-D), and failure modes such as data dropout or motion blur. The current SynBench and DeformedTissue datasets provide valuable testbeds, but further development of standardized benchmarks with annotated ground truth in real surgical or industrial settings will be essential for fair comparison and robust validation. Moreover, uncertainty quantification techniques, such

as Monte Carlo dropout or ensemble variance estimation, can be integrated to assess confidence in the predicted deformation, which is particularly important in critical applications like neurosurgery or prosthetic design.

Chapter 6

Summary and Conclusion

This thesis presented a comprehensive investigation into the problem of non-rigid PCR, with a particular emphasis on scenarios involving soft tissue deformation. Motivated by the limitations of rigid and classical deformable models in capturing the complex, local-to-global transformations encountered in surgical environments, we formulated and addressed key research questions regarding the capabilities of deep learning models to enhance registration accuracy and robustness under realistic deformation, noise, and partial overlap conditions.

The research progressed through the development of three novel methods: Robust-DefReg, DefTransNet, and Learning-to-Refine. Each method was designed to address specific limitations observed in existing approaches, leading to a stepwise improvement in accuracy, generalization, and stability. These contributions moved from local feature modeling using graph-based architectures to global attention mechanisms, and finally to iterative refinement guided by probabilistic regularization.

Robust-DefReg introduced a graph-based registration pipeline that leveraged a coarse-to-fine framework to capture local geometric variations. Using EdgeConv layers and T-Net modules, the method extracted rich local features and learned spatial transformations capable of handling moderate non-rigid deformations. The refinement of displacement fields using Loopy Belief Propagation further improved robustness, particularly under noisy and partially missing data.

DefTransNet addressed the limitations of purely local models by incorporating a transformer-based architecture for global feature embedding. Local descriptors generated via graph convolutions were passed through self-attention layers, enabling the network to model long-range dependencies across the point cloud. This significantly improved performance under severe deformations and challenging topological variations. Experimental results demonstrated that DefTransNet outperformed prior approaches in both synthetic and real-world datasets, particularly in scenarios with limited overlap.

Learning-to-Refine extended this framework by introducing an iterative strategy that

optimized deformation fields across multiple steps. At each stage, a KL divergence term was used to regularize the learned displacements against a predefined isotropic Gaussian prior. This formulation promoted more stable and plausible deformation learning while preventing overfitting to noisy or overly complex geometries. Results showed that stronger regularization improved performance, especially for high-deformation cases, and that the iterative nature of the model led to consistent error reduction over time.

In addition to method development, this thesis contributed two novel datasets designed to benchmark non-rigid PCR under realistic conditions. The SynBench dataset provided a synthetic yet controlled environment for studying soft tissue deformation, incorporating variations in deformation intensity, noise, and partiality. The DeformedTissue dataset complemented this with real-world scans collected from medical simulations, offering an important validation ground for the practical applicability of the proposed models. Together, these datasets enabled standardized comparisons across different methods and contributed to the advancement of benchmarking practices in this field.

Quantitative and qualitative results across all datasets revealed that the three proposed methods consistently outperformed state-of-the-art baselines, including Deep-Geo-Reg and Diffeomorphic models. The transformer-based and iterative frameworks achieved particularly strong results under difficult conditions, validating the initial research hypothesis that combining attention-based architectures with regularization can significantly enhance non-rigid registration performance. The ablation studies further confirmed that each architectural component contributed meaningfully to the final performance, and that the T-Net blocks improved robustness against random spatial transformations, supporting the second hypothesis stated in the introduction.

The scientific contributions of this work can be summarized as follows.

- First, a graph-based architecture (Robust-DefReg) was proposed for learning coarse-to-fine deformations using local feature aggregation and message passing.
- Second, a hybrid transformer-graph model (DefTransNet) was introduced to capture both local and global context for non-rigid alignment.
- Third, an iterative refinement method (Learning-to-Refine) was developed using KL divergence to constrain the deformation space, improving convergence and structural coherence.
- Finally, the SynBench and DeformedTissue datasets were developed and released to facilitate comprehensive benchmarking of deformable registration tasks.

The findings of this thesis demonstrate that deep learning models, when carefully designed with a balance of local structure, global context, and probabilistic regularization, can

offer significant improvements in non-rigid PCR. These models are well-suited to applications in surgical navigation, soft tissue modeling, and medical image-guided procedures, where robustness, accuracy, and adaptability are critical.

Future work may explore extending these models to temporal sequences for dynamic registration, incorporating anatomical priors to enforce domain-specific constraints, or integrating multimodal inputs such as segmentation masks or imaging data. Additionally, adapting these methods for deployment in real-time clinical environments will require further investigation into model compression, latency reduction, and hardware optimization.

Bibliography

[1] Balder Croquet, Daan Christiaens, Seth M Weinberg, Michael Bronstein, Dirk Vandermeulen, and Peter Claes. Unsupervised diffeomorphic surface registration and non-linear modelling. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part IV 24*, pages 118–128. Springer, 2021.

[2] Lena Maier-Hein, Swaroop S Vedula, Stefanie Speidel, Nassir Navab, Ron Kikinis, Adrian Park, Matthias Eisenmann, Hubertus Feussner, Germain Forestier, Stamatia Giannarou, et al. Surgical data science for next-generation interventions. *Nature Biomedical Engineering*, 1(9):691–696, 2017.

[3] Jong-yi Hong, Eui-ho Suh, and Sung-Jin Kim. Context-aware systems: A literature review and classification. *Expert Systems with applications*, 36(4):8509–8522, 2009.

[4] Herve Delingette. Toward realistic soft-tissue modeling in medical simulation. *Proceedings of the IEEE*, 86(3):512–523, 2002.

[5] Gary KL Tam, Zhi-Quan Cheng, Yu-Kun Lai, Frank C Langbein, Yonghuai Liu, David Marshall, Ralph R Martin, Xian-Fang Sun, and Paul L Rosin. Registration of 3d point clouds and meshes: A survey from rigid to nonrigid. *IEEE transactions on visualization and computer graphics*, 19(7):1199–1217, 2012.

[6] François Pomerleau, Francis Colas, Roland Siegwart, et al. A review of point cloud registration algorithms for mobile robotics. *Foundations and Trends® in Robotics*, 4(1):1–104, 2015.

[7] Xiaoshui Huang, Guofeng Mei, Jian Zhang, and Rana Abbas. A comprehensive survey on point cloud registration. *arXiv preprint arXiv:2103.02690*, 2021.

[8] Rogério Yugo Takimoto, Marcos de Sales Guerra Tsuzuki, Renato Vogelaar, Thiago de Castro Martins, André Kubagawa Sato, Yuma Iwao, Toshiyuki Gotoh, and Sei-

ichiro Kagei. 3d reconstruction and multiple point cloud registration using a low precision rgb-d sensor. *Mechatronics*, 35:11–22, 2016.

[9] Yuchao Zheng, Yujie Li, Shuo Yang, and Huimin Lu. Global-pbnet: A novel point cloud registration for autonomous driving. *IEEE Transactions on Intelligent Transportation Systems*, 23(11):22312–22319, 2022.

[10] Bilawal Mahmood and SangUk Han. 3d registration of indoor point clouds for augmented reality. In *Computing in Civil Engineering 2019: Visualization, Information Modeling, and Simulation*, pages 1–8. American Society of Civil Engineers Reston, VA, 2019.

[11] Ruisheng Wang, Jiju Peethambaran, and Dong Chen. Lidar point clouds to 3-d urban models : a review. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 11(2):606–627, 2018.

[12] Qian Wang and Min-Koo Kim. Applications of 3d point cloud data in the construction industry: A fifteen-year review from 2004 to 2018. *Advanced Engineering Informatics*, 39:306–319, 2019.

[13] Yirui Zhang, Yanni Zou, and Peter X Liu. Point cloud registration in laparoscopic liver surgery using keypoint correspondence registration network. *IEEE Transactions on Medical Imaging*, 2024.

[14] Zhihao Li and Manning Wang. Rigid point cloud registration based on correspondence cloud for image-to-patient registration in image-guided surgery. *Medical Physics*, 51(7):4554–4566, 2024.

[15] Martin Sinko, Patrik Kamencay, Robert Hudec, and Miroslav Benco. 3d registration of the point cloud data using icp algorithm in medical image analysis. In *2018 ELEKTRO*, pages 1–6. IEEE, 2018.

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

[17] Rainer Sprengel, Karl Rohr, and H Siegfried Stiehl. Thin-plate spline approximation for image registration. In *Proceedings of 18th annual international conference of the IEEE engineering in medicine and biology society*, volume 3, pages 1190–1191. IEEE, 1996.

[18] Guoli Song, Jianda Han, Yiwen Zhao, Zheng Wang, and Huibin Du. A review on medical image registration as an optimization problem. *Current Medical Imaging*, 13(3):274–283, 2017.

[19] Wen-Chung Chang and Van-Toan Pham. 3-d point cloud registration using convolutional neural networks. *Applied Sciences*, 9(16):3273, 2019.

[20] Zheng Qin, Hao Yu, Changjian Wang, Yuxing Peng, and Kai Xu. Deep graph-based spatial consistency for robust non-rigid point cloud registration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5394–5403, 2023.

[21] Zheng Qin, Hao Yu, Changjian Wang, Yulan Guo, Yuxing Peng, and Kai Xu. Geometric transformer for fast and robust point cloud registration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11143–11152, 2022.

[22] Sara Monji-Azad, Marvin Kinz, Siddharth Kothari, Robin Khanna, Amrei Carla Mihan, David Maennel, Claudia Scherl, and Juergen Hesser. Deftransnet: A transformer-based method for non-rigid point cloud registration in the simulation of soft tissue deformation. *arXiv preprint arXiv:2502.06336*, 2025.

[23] Sara Monji-Azad, Marvin Kinz, Jürgen Hesser, and Nikolas Löw. Simtool: A toolset for soft body simulation using flex and unreal engine. *Software Impacts*, 17:100521, 2023.

[24] Sara Monji-Azad, Marvin Kinz, Claudia Scherl, David Männle, Jürgen Hesser, and Nikolas Löw. Synbench: A synthetic benchmark for non-rigid 3d point cloud registration. *https://doi.org/10.48550/arXiv.2409.14474*, 2024.

[25] David Männle, Jan Pohlmann, Sara Monji-Azad, Jürgen Hesser, Nicole Rotter, Annette Affolter, Anne Lammert, Benedikt Kramer, Sonja Ludwig, Lena Huber, et al. Artificial intelligence directed development of a digital twin to measure soft tissue shift during head and neck surgery. *Plos one*, 18(8):e0287081, 2023.

[26] Sara Monji-Azad, David Männle, Jürgen Hesser, Jan Pohlmann, Nicole Rotter, Annette Affolter, Cleo Aron Weis, Sonja Ludwig, and Claudia Scherl. Point cloud registration for measuring shape dependence of soft tissue deformation by digital twins in head and neck surgery. *Biomedicine hub*, 9(1):9–15, 2024.

[27] Sara Monji-Azad, Marvin Kinz, David Männel, Claudia Scherl, and Jürgen Hesser. Robust-defreg: a robust coarse to fine non-rigid point cloud registration method based on graph convolutional neural networks. *Measurement Science and Technology*, 36(1):015426, 2024.

[28] Paul J Besl and Neil D McKay. Method for registration of 3-d shapes. In *Sensor fusion IV: control paradigms and data structures*, volume 1611, pages 586–606. Spie, 1992.

[29] Zhengyou Zhang. Iterative point matching for registration of free-form curves and surfaces. *International journal of computer vision*, 13(2):119–152, 1994.

[30] Szymon Rusinkiewicz and Marc Levoy. Efficient variants of the icp algorithm. In *Proceedings third international conference on 3-D digital imaging and modeling*, pages 145–152. IEEE, 2001.

[31] Fred L. Bookstein. Principal warps: Thin-plate splines and the decomposition of deformations. *IEEE Transactions on pattern analysis and machine intelligence*, 11(6):567–585, 1989.

[32] Umberto Castellani and Adrien Bartoli. 3d shape registration. In *3D Imaging, Analysis and Applications*, pages 353–411. Springer, 2020.

[33] Isabelle Guyon and André Elisseeff. An introduction to feature extraction. In *Feature extraction: foundations and applications*, pages 1–25. Springer, 2006.

[34] Thomas S Huang and Arun N Netravali. Motion and structure from feature correspondences: A review. *Proceedings of the IEEE*, 82(2):252–268, 1994.

[35] Sara Monji-Azad, Jürgen Hesser, and Nikolas Löw. A review of non-rigid transformations and learning-based 3d point cloud registration methods. *ISPRS Journal of Photogrammetry and Remote Sensing*, 196:58–72, 2023.

[36] ShaoCong Liu, Tao Wang, Yan Zhang, Ruqin Zhou, Chenguang Dai, Yongsheng Zhang, Haozhen Lei, and Hanyun Wang. Rethinking of learning-based 3d keypoints detection for large-scale point clouds registration. *International Journal of Applied Earth Observation and Geoinformation*, 112:102944, 2022.

[37] Kari Pulli. Multiview registration for large data sets. In *Second international conference on 3-d digital imaging and modeling (cat. no. pr00062)*, pages 160–168. IEEE, 1999.

[38] Radu Bogdan Rusu, Nico Blodow, and Michael Beetz. Fast point feature histograms (fpfh) for 3d registration. In *2009 IEEE international conference on robotics and automation*, pages 3212–3217. IEEE, 2009.

[39] Ivan Sipiran and Benjamin Bustos. Harris 3d: a robust extension of the harris operator for interest point detection on 3d meshes. *The Visual Computer*, 27:963–976, 2011.

[40] Xiaoshui Huang, Jian Zhang, Qiang Wu, Lixin Fan, and Chun Yuan. A coarse-to-fine algorithm for matching and registration in 3d cross-source point clouds. *IEEE Transactions on Circuits and Systems for Video Technology*, 28(10):2965–2977, 2017.

[41] Warren Cheung and Ghassan Hamarneh. n-sift: n-dimensional scale invariant feature transform. *IEEE Transactions on Image Processing*, 18(9):2012–2021, 2009.

[42] Linjia Hu and Saeid Nooshabadi. High-dimensional image descriptor matching using highly parallel kd-tree construction and approximate nearest neighbor search. *Journal of Parallel and Distributed Computing*, 132:127–140, 2019.

[43] K Somani Arun, Thomas S Huang, and Steven D Blostein. Least-squares fitting of two 3-d point sets. *IEEE Transactions on pattern analysis and machine intelligence*, (5):698–700, 1987.

[44] Martin A Fischler and Robert C Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981.

[45] Mehran Fotouhi, Hamid Hekmatian, Mohammad Amin Kashani-Nezhad, and Shohreh Kasaei. Sc-ransac: spatial consistency on ransac. *Multimedia Tools and Applications*, 78(7):9429–9461, 2019.

[46] Jiayuan Li, Qingwu Hu, and Mingyao Ai. Gesac: Robust graph enhanced sample consensus for point cloud registration. *ISPRS Journal of Photogrammetry and Remote Sensing*, 167:363–374, 2020.

[47] Jiayuan Li, Qingwu Hu, and Mingyao Ai. Point cloud registration based on one-point ransac and scale-annealing biweight estimation. *IEEE Transactions on Geoscience and Remote Sensing*, 59(11):9716–9729, 2021.

[48] Siwen Quan and Jiaqi Yang. Compatibility-guided sampling consensus for 3-d point cloud registration. *IEEE Transactions on Geoscience and Remote Sensing*, 58(10):7380–7392, 2020.

[49] Liang Cheng, Song Chen, Xiaoqiang Liu, Hao Xu, Yang Wu, Manchun Li, and Yanming Chen. Registration of laser scanning point clouds: A review. *Sensors*, 18(5):1641, 2018.

[50] Tal Darom and Yosi Keller. Scale-invariant features for 3-d mesh models. *IEEE Transactions on Image Processing*, 21(5):2758–2769, 2012.

[51] Yu Zhong. Intrinsic shape signatures: A shape descriptor for 3d object recognition. In *2009 IEEE 12th international conference on computer vision workshops, ICCV Workshops*, pages 689–696. IEEE, 2009.

[52] Marius Muja and David G. Lowe. Fast approximate nearest neighbors with automatic algorithm configuration. In *VISAPP (1)*, pages –, 2009.

[53] Kok-Lim Low. Linear least-squares optimization for point-to-plane icp surface registration. Technical Report TR04-004, University of North Carolina, 2004.

[54] Qin Zou, Qin Sun, Long Chen, Bu Nie, and Qingquan Li. A comparative analysis of lidar slam-based indoor navigation for autonomous vehicles. *IEEE Transactions on Intelligent Transportation Systems*, 23(7):6907–6921, 2021.

[55] Yulan Guo, Mohammed Bennamoun, Ferdous Sohel, Min Lu, and Jianwei Wan. 3d object recognition in cluttered scenes with local surface features: A survey. *IEEE transactions on pattern analysis and machine intelligence*, 36(11):2270–2287, 2014.

[56] Evdokia Saiti and Theoharis Theoharis. Multimodal registration across 3d point clouds and ct-volumes. *Computers & graphics*, 106:259–266, 2022.

[57] Sara Monji-Azad, Shohreh Kasaei, and Amir-Masoud Eftekhari-Moghadam. An efficient augmented reality method for sports scene visualization from single moving camera. In *2014 22nd Iranian Conference on Electrical Engineering (ICEE)*, pages 1064–1069. IEEE, 2014.

[58] Yifei Tian, Wei Song, Su Sun, Simon Fong, and Shuanghui Zou. 3d object recognition method with multiple feature extraction from lidar point clouds. *The Journal of Supercomputing*, 75:4430–4442, 2019.

[59] Blaine Rister, Mark A Horowitz, and Daniel L Rubin. Volumetric image registration from invariant keypoints. *IEEE Transactions on Image Processing*, 26(10):4900–4910, 2017.

[60] Jiaolong Yang, Hongdong Li, Dylan Campbell, and Yunde Jia. Go-icp: A globally optimal solution to 3d icp point-set registration. *IEEE transactions on pattern analysis and machine intelligence*, 38(11):2241–2254, 2015.

[61] Nicolas Mellado, Dror Aiger, and Niloy J Mitra. Super 4pcs fast global pointcloud registration via smart indexing. In *Computer graphics forum*, volume 33, pages 205–215. Wiley Online Library, 2014.

[62] S. Rusinkiewicz and M. Levoy. Efficient variants of the icp algorithm. In *3D Digital Imaging and Modeling*, pages –, 2001.

[63] Varuna De Silva, Jamie Roche, and Ahmet Kondoz. Robust fusion of lidar and wide-angle camera data for autonomous mobile robots. *Sensors*, 18(8):2730, 2018.

[64] Peter Meer. Robust techniques for computer vision. *Emerging topics in computer vision*, pages 107–190, 2004.

[65] Peter Lancaster and Kes Salkauskas. Surfaces generated by moving least squares methods. *Mathematics of computation*, 37(155):141–158, 1981.

[66] Shishir Pagad, Divya Agarwal, Sathya Narayanan, Kasturi Rangan, Hyungjin Kim, and Ganesh Yalla. Robust method for removing dynamic objects from point clouds. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pages 10765–10771. IEEE, 2020.

[67] Peter J. Huber. *Robust Statistics*. Wiley, 1981.

[68] Haili Chui and Anand Rangarajan. A new point matching algorithm for non-rigid registration. *Computer Vision and Image Understanding*, 89(2-3):114–141, 2003.

[69] Yuehua Zhao, Jiguang Zhang, Shibiao Xu, and Jie Ma. Deep learning-based low overlap point cloud registration for complex scenario: The review. *Information Fusion*, 107:102305, 2024.

[70] Andrew W Fitzgibbon. Robust registration of 2d and 3d point sets. *Image and vision computing*, 21(13-14):1145–1153, 2003.

[71] Timo Hackel, Jan D Wegner, and Konrad Schindler. Fast semantic segmentation of 3d point clouds with strongly varying density. *ISPRS annals of the photogrammetry, remote sensing and spatial information sciences*, 3:177–184, 2016.

[72] Can Qin, Haoxuan You, Lichen Wang, C-C Jay Kuo, and Yun Fu. Pointdan: A multi-scale 3d domain adaption network for point cloud representation. *Advances in Neural Information Processing Systems*, 32, 2019.

[73] Zhenghua Zhang, Guoliang Chen, Xuan Wang, and Mingcong Shu. Ddrnet: Fast point cloud registration network for large-scale scenes. *ISPRS Journal of Photogrammetry and Remote Sensing*, 175:184–198, 2021. doi:10.1016/j.isprsjprs.2021.03.003.

[74] Song Ge, Guoliang Fan, and Meng Ding. Non-rigid point set registration with global-local topology preservation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 245–251, 2014.

[75] Hao Xu, Shuaicheng Liu, Guangfu Wang, Guanghui Liu, and Bing Zeng. Omnet: Learning overlapping mask for partial-to-partial point cloud registration. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 3132–3141, 2021.

[76] Yang Chen and Gérard Medioni. Object modelling by registration of multiple range images. *Image and vision computing*, 10(3):145–155, 1992.

[77] Jayakorn Vongkulbhisal, Benat Irastorza Ugalde, Fernando De la Torre, and Joao P Costeira. Inverse composition discriminative optimization for point cloud registration. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2993–3001, 2018.

[78] Yuyan Liu, Wei He, and Hongyan Zhang. Glocnet: Robust feature matching with global–local consistency network for remote sensing image registration. *IEEE Transactions on Geoscience and Remote Sensing*, 61:1–13, 2023.

[79] Xuan He, Shuguo Pan, Wang Gao, and Xinyu Lu. Lidar-inertial-gnss fusion positioning system in urban environment: Local accurate registration and global drift-free. *Remote Sensing*, 14(9):2104, 2022.

[80] Ruikai Cui, Xibin Song, Weixuan Sun, Senbo Wang, Weizhe Liu, Shenzhou Chen, Taizhang Shang, Yang Li, Nick Barnes, Hongdong Li, et al. Lam3d: Large image-point clouds alignment model for 3d reconstruction from single image. *Advances in Neural Information Processing Systems*, 37:4454–4480, 2024.

[81] Bing Jian and Baba C Vemuri. Robust point set registration using gaussian mixture models. *IEEE transactions on pattern analysis and machine intelligence*, 33(8):1633–1645, 2010.

[82] Jonathan T Barron. A general and adaptive robust loss function. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4331–4339, 2019.

[83] Zi Jian Yew and Gim Hee Lee. Rpm-net: Robust point matching using learned features. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11824–11833, 2020.

[84] Yue Wang and Justin M Solomon. Deep closest point: Learning representations for point cloud registration. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 3523–3532, 2019.

[85] Ananth Ranganathan. The levenberg-marquardt algorithm. *Tutoral on LM algorithm*, 11(1):101–110, 2004.

[86] Serge Gratton, Amos S Lawless, and Nancy K Nichols. Approximate gauss–newton methods for nonlinear least squares problems. *SIAM Journal on Optimization*, 18(1):106–132, 2007.

[87] Yong Wang. Gauss–newton method. *Wiley Interdisciplinary Reviews: Computational Statistics*, 4(4):415–420, 2012.

[88] Dong C Liu and Jorge Nocedal. On the limited memory bfgs method for large scale optimization. *Mathematical programming*, 45(1):503–528, 1989.

[89] Bailin Deng, Yuxin Yao, Roberto M Dyke, and Juyong Zhang. A survey of non-rigid 3d registration. *Computer Graphics Forum*, 41(2):559–589, 2022.

[90] 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.

[91] 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.

[92] 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. URL: https://modelnet.cs.princeton.edu/.

[93] Hang Su, Subhransu Maji, Evangelos Kalogerakis, and Erik Learned-Miller. Multi-view convolutional neural networks for 3d shape recognition. In *Proceedings of the IEEE international conference on computer vision*, pages 945–953, 2015.

[94] Matthias Schaufelberger, Christian Kaiser, Reinald Kühle, Andreas Wachter, Frederic Weichel, Niclas Hagen, Friedemann Ringwald, Urs Eisenmann, Jürgen Hoffmann, Michael Engel, et al. 3d-2d distance maps conversion enhances classification of craniosynostosis. *IEEE Transactions on Biomedical Engineering*, 70(11):3156–3165, 2023.

[95] Hugues Thomas, Charles R Qi, Jean-Emmanuel Deschaud, Beatriz Marcotegui, François Goulette, and Leonidas J Guibas. Kpconv: Flexible and deformable convolution for point clouds. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 6411–6420, 2019.

[96] Yabo Fu, Yang Lei, Tonghe Wang, Pretesh Patel, Ashesh B. Jani, Hui Mao, Walter J. Curran, Tian Liu, and Xiaofeng Yang. Biomechanically constrained non-rigid mr-trus prostate registration using deep learning based 3d point cloud matching. *Medical image analysis*, 67:101845, 2021. doi:10.1016/j.media.2020.101845.

[97] Micha Pfeiffer, Carina Riediger, Stefan Leger, Jens-Peter Kühn, Danilo Seppelt, Ralf-Thorsten Hoffmann, Jürgen Weitz, and Stefanie Speidel. Non-rigid volume to surface registration using a data-driven biomechanical model. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 724–734. Springer, 2020.

[98] Lingjing Wang, Jianchun Chen, Xiang Li, and Yi Fang. Non-rigid point set registration networks. *arXiv preprint arXiv:1904.01428*, 2019.

[99] Lingyu Wei, Qixing Huang, Duygu Ceylan, Etienne Vouga, and Hao Li. Dense human body correspondences using convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1544–1553, 2016.

[100] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25, 2012.

[101] Soshi Shimada, Vladislav Golyanik, Edgar Tretschk, Didier Stricker, and Christian Theobalt. Dispvoxnets: Non-rigid point set alignment with supervised learning proxies. In *2019 International Conference on 3D Vision (3DV)*, pages 27–36. IEEE, 2019.

[102] Yuqi Yang, Shilin Liu, Hao Pan, Yang Liu, and Xin Tong. Pfcnn: Convolutional neural networks on 3d surfaces using parallel frames. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13578–13587, 2020.

[103] Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*, 2016.

[104] 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.

[105] Jiarong Xu, Junru Chen, Siqi You, Zhiqing Xiao, Yang Yang, and Jiangang Lu. Robustness of deep learning models on graphs: A survey. *AI Open*, 2:69–78, 2021.

[106] Si Zhang, Hanghang Tong, Jiejun Xu, and Ross Maciejewski. Graph convolutional networks: a comprehensive review. *Computational Social Networks*, 6(1):1–23, 2019.

[107] Lasse Hansen and Mattias P Heinrich. Deep learning based geometric registration for medical images: How accurate can we get without visual features? In *Information Processing in Medical Imaging: 27th International Conference, IPMI 2021, Virtual Event, June 28–June 30, 2021, Proceedings 27*, pages 18–30. Springer, 2021.

[108] Lasse Hansen, Doris Dittmer, and Mattias P Heinrich. Learning deformable point set registration with regularized dynamic graph cnns for large lung motion in copd patients. In *International Workshop on Graph Learning in Medical Imaging*, pages 53–61. Springer, 2019.

[109] Xiaobo Hu, Dejun Zhang, Jinzhi Chen, Yiqi Wu, and Yilin Chen. Nrtnet: An unsupervised method for 3d non-rigid point cloud registration based on transformer. *Sensors*, 22(14):5128, 2022.

[110] Puhua Jiang, Mingze Sun, and Ruqi Huang. Neural intrinsic embedding for non-rigid point cloud matching. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 21835–21845, 2023.

[111] Jiahui Huang, Tolga Birdal, Zan Gojcic, Leonidas J Guibas, and Shi-Min Hu. Multiway non-rigid point cloud registration via learned functional map synchronization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022.

[112] 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.

[113] 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.

[114] Yasuhiro Aoki, Hunter Goforth, Rangaprasad Arun Srivatsan, and Simon Lucey. Pointnetlk: Robust & efficient point cloud registration using pointnet. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 7163–7172, 2019.

[115] Xueqian Li, Jhony Kaesemodel Pontes, and Simon Lucey. Pointnetlk revisited. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12763–12772, 2021.

[116] Lingjing Wang, Xiang Li, and Yi Fang. Gp-aligner: Unsupervised non-rigid group-wise point set registration based on optimized group latent descriptor. *arXiv preprint arXiv:2007.12979*, 2020.

[117] Yang Li and Tatsuya Harada. Non-rigid point cloud registration with neural deformation pyramid. *Advances in Neural Information Processing Systems*, 35:27757–27768, 2022.

[118] Lingjing Wang, Xiang Li, Jianchun Chen, and Yi Fang. Coherent point drift networks: Unsupervised learning of non-rigid point set registration. *arXiv preprint arXiv:1906.03039*, 2019.

[119] 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.

[120] Yulan Guo, Hanyun Wang, Qingyong Hu, Hao Liu, Li Liu, and Mohammed Bennamoun. Deep learning for 3d point clouds: A survey. *IEEE transactions on pattern analysis and machine intelligence*, 43(12):4338–4364, 2020.

[121] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

[122] Hengshuang Zhao, Li Jiang, Jiaya Jia, Philip HS Torr, and Vladlen Koltun. Point transformer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 16259–16268, 2021.

[123] Meng-Hao Guo, Jun-Xiong Cai, Zheng-Ning Liu, Tai-Jiang Mu, Ralph R Martin, and Shi-Min Hu. Pct: Point cloud transformer. *Computational Visual Media*, 7(2):187–199, 2021.

[124] Hao Yu, Fu Li, Mahdi Saleh, Benjamin Busam, and Slobodan Ilic. Cofinet: Reliable coarse-to-fine correspondences for robust pointcloud registration. *Advances in Neural Information Processing Systems*, 34:23872–23884, 2021.

[125] Yang Li and Tatsuya Harada. Lepard: Learning partial point cloud matching in rigid and deformable scenes. In *Proceedings of the IEEE/CVF conference on computer*

vision and pattern recognition, pages 5554–5564, 2022. URL: https://github.com/rabbityl/lepard.

[126] Fan Yang, Lin Guo, Zhi Chen, and Wenbing Tao. One-inlier is first: Towards efficient position encoding for point cloud registration. *Advances in Neural Information Processing Systems*, 35:6982–6995, 2022.

[127] Zi Jian Yew and Gim Hee Lee. Regtr: End-to-end point cloud correspondences with transformers. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 6677–6686, 2022.

[128] Mehdi Mirza and Simon Osindero. Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*, 2014.

[129] Haolin Tang and Yanxiao Zhao. A conditional generative adversarial network for non-rigid point set registration. In *2021 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE)*, pages 1–6. IEEE, 2021.

[130] Wanquan Feng, Juyong Zhang, Hongrui Cai, Haofei Xu, Junhui Hou, and Hujun Bao. Recurrent multi-view alignment network for unsupervised surface registration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10297–10307, 2021.

[131] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[132] Boulbaba Ben Amor, Sylvain Arguillère, and Ling Shao. Resnet-lddmm: advancing the lddmm framework using deep residual networks. *arXiv preprint arXiv:2102.07951*, 2021.

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

[134] Haowen Deng, Tolga Birdal, and Slobodan Ilic. Ppfnet: Global context aware local features for robust 3d point matching. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 195–205, 2018.

[135] Haowen Deng, Tolga Birdal, and Slobodan Ilic. Ppf-foldnet: Unsupervised learning of rotation invariant 3d local descriptors. In *Proceedings of the European conference on computer vision (ECCV)*, pages 602–618, 2018.

[136] Andy Zeng, Shuran Song, Matthias Nießner, Matthew Fisher, Jianxiong Xiao, and Thomas Funkhouser. 3dmatch: Learning local geometric descriptors from rgb-d reconstructions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1802–1811, 2017.

[137] Julien Valentin, Angela Dai, Matthias Nießner, Pushmeet Kohli, Philip Torr, Shahram Izadi, and Cem Keskin. Learning to navigate the energy landscape. In *2016 Fourth International Conference on 3D Vision (3DV)*, pages 323–332. IEEE, 2016.

[138] Jamie Shotton, Ben Glocker, Christopher Zach, Shahram Izadi, Antonio Criminisi, and Andrew Fitzgibbon. Scene coordinate regression forests for camera relocalization in rgb-d images. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2930–2937, 2013.

[139] Jianxiong Xiao, Andrew Owens, and Antonio Torralba. Sun3d: A database of big spaces reconstructed using sfm and object labels. In *Proceedings of the IEEE international conference on computer vision*, pages 1625–1632, 2013.

[140] Kevin Lai, Liefeng Bo, and Dieter Fox. Unsupervised feature learning for 3d scene labeling. In *2014 IEEE International Conference on Robotics and Automation (ICRA)*, pages 3050–3057. IEEE, 2014.

[141] Andrew E Johnson and Martial Hebert. Using spin images for efficient object recognition in cluttered 3d scenes. *IEEE Transactions on pattern analysis and machine intelligence*, 21(5):433–449, 1999.

[142] Samuele Salti, Federico Tombari, and Luigi Di Stefano. Shot: Unique signatures of histograms for surface and texture description. *Computer Vision and Image Understanding*, 125:251–264, 2014.

[143] Federico Tombari, Samuele Salti, and Luigi Di Stefano. Unique shape context for 3d data description. In *Proceedings of the ACM workshop on 3D object retrieval*, pages 57–62, 2010.

[144] Yue Wang and Justin M Solomon. Prnet: Self-supervised learning for partial-to-partial registration. *Advances in neural information processing systems*, 32, 2019.

[145] Angel X Chang, Thomas Funkhouser, Leonidas Guibas, Pat Hanrahan, Qixing Huang, Zimo Li, Silvio Savarese, Manolis Savva, Shuran Song, Hao Su, et al. Shapenet: An information-rich 3d model repository. *arXiv preprint arXiv:1512.03012*, 2015.

[146] Zan Gojcic, Caifa Zhou, Jan D Wegner, and Andreas Wieser. The perfect match: 3d point cloud matching with smoothed densities. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5545–5554, 2019.

[147] François Pomerleau, Ming Liu, Francis Colas, and Roland Siegwart. Challenging data sets for point cloud registration algorithms. *The International Journal of Robotics Research*, 31(14):1705–1711, 2012.

[148] Tianye Li, Timo Bolkart, Michael J Black, Hao Li, and Javier Romero. Learning a model of facial shape and expression from 4d scans. *ACM Trans. Graph.*, 36(6):194–1, 2017.

[149] Federica Bogo, Javier Romero, Gerard Pons-Moll, and Michael J Black. Dynamic faust: Registering human bodies in motion. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 6233–6242, 2017.

[150] Jan Bednarik, Pascal Fua, and Mathieu Salzmann. Learning to reconstruct texture-less deformable surfaces from a single view. In *2018 international conference on 3d vision (3DV)*, pages 606–615. IEEE, 2018.

[151] Andreas Geiger, Philip Lenz, Christoph Stiller, and Raquel Urtasun. Vision meets robotics: The kitti dataset. *The International Journal of Robotics Research*, 32(11):1231–1237, 2013.

[152] Shengyu Huang, Zan Gojcic, Mikhail Usvyatsov, Andreas Wieser, and Konrad Schindler. Predator: Registration of 3d point clouds with low overlap. In *Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition*, pages 4267–4276, 2021.

[153] Gul Varol, Javier Romero, Xavier Martin, Naureen Mahmood, Michael J Black, Ivan Laptev, and Cordelia Schmid. Learning from synthetic humans. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 109–117, 2017.

[154] Simone Melzi, Riccardo Marin, Emanuele Rodolà, Umberto Castellani, Jing Ren, Adrien Poulenard, Peter Wonka, and Maks Ovsjanikov. Shrec 2019: Matching humans with different connectivity. In *Eurographics Workshop on 3D Object Retrieval*, volume 7, page 3. The Eurographics Association, 2019.

[155] Roger Grosse, Micah K Johnson, Edward H Adelson, and William T Freeman. Ground truth dataset and baseline evaluations for intrinsic image algorithms. In *2009 IEEE 12th International Conference on Computer Vision*, pages 2335–2342. IEEE, 2009.

[156] Gustavo Marques Netto and Manuel M Oliveira. Robust point-cloud registration based on dense point matching and probabilistic modeling. *The Visual Computer*, 38(9):3217–3230, 2022.

[157] Alexander M Bronstein, Michael M Bronstein, and Ron Kimmel. *Numerical geometry of non-rigid shapes*. Springer Science & Business Media, 2008.

[158] Daniel Vlasic, Ilya Baran, Wojciech Matusik, and Jovan Popović. Articulated mesh animation from multi-view silhouettes. *Acm Siggraph 2008 papers*, pages 1–9, 2008.

[159] Qianli Ma, Jinlong Yang, Anurag Ranjan, Sergi Pujades, Gerard Pons-Moll, Siyu Tang, and Michael J Black. Learning to dress 3d people in generative clothing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6469–6478, 2020.

[160] Gerard Pons-Moll, Sergi Pujades, Sonny Hu, and Michael J Black. Clothcap: Seamless 4d clothing capture and retargeting. *ACM Transactions on Graphics (ToG)*, 36(4):1–15, 2017.

[161] Yang Li, Hikari Takehara, Takafumi Taketomi, Bo Zheng, and Matthias Nießner. 4dcomplete: Non-rigid motion estimation beyond the observable surface. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 12706–12716, 2021.

[162] Aljaz Bozic, Michael Zollhofer, Christian Theobalt, and Matthias Nießner. Deepdeform: Learning non-rigid rgb-d reconstruction with semi-supervised data. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7002–7012, 2020.

[163] Fanbo Xiang, Yuzhe Qin, Kaichun Mo, Yikuan Xia, Hao Zhu, Fangchen Liu, Minghua Liu, Hanxiao Jiang, Yifu Yuan, He Wang, et al. Sapien: A simulated part-based interactive environment. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11097–11107, 2020.

[164] Federica Bogo, Javier Romero, Matthew Loper, and Michael J Black. Faust: Dataset and evaluation for 3d mesh registration. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3794–3801, 2014.

[165] Dragomir Anguelov, Praveen Srinivasan, Daphne Koller, Sebastian Thrun, Jim Rodgers, and James Davis. Scape: shape completion and animation of people. *ACM SIGGRAPH 2005 Papers*, pages 408–416, 2005.

[166] Xinyu Chen, Jiahui Luo, Yan Ren, Tong Cui, and Meng Zhang. Mafnet: a two-stage multiple attention fusion network for partial-to-partial point cloud registration. *Measurement Science and Technology*, 35(12):125113, 2024.

[167] Jan Bender, Matthias Müller, Miguel A Otaduy, Matthias Teschner, and Miles Macklin. A survey on position-based simulation methods in computer graphics. *Computer graphics forum*, 33(6):228–251, 2014.

[168] Andrew Nealen, Matthias Müller, Richard Keiser, Eddy Boxerman, and Mark Carlson. Physically based deformable models in computer graphics. *Computer graphics forum*, 25(4):809–836, 2006.

[169] Nadia Magnenat-Thalmann and Pascal Volino. From early draping to haute couture models: 20 years of research. *The Visual Computer*, 21(8):506–519, 2005.

[170] Marvin Kinz. SimTool-SynBench, 2023. doi:10.11588/data/R9IKCF.

[171] Xueqian Li, Jhony Kaesemodel Pontes, and Simon Lucey. Pointnetlk revisited. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12763–12772, 2021.

[172] Xu Cheng, Zhongwei Li, Kai Zhong, and Yusheng Shi. An automatic and robust point cloud registration framework based on view-invariant local feature descriptors and transformation consistency verification. *Optics and Lasers in Engineering*, 98:37–45, 2017.

[173] Seth D Billings, Emad M Boctor, and Russell H Taylor. Iterative most-likely point registration (imlp): A robust algorithm for computing optimal shape alignment. *PloS one*, 10(3):e0117688, 2015.

[174] Yasushi Miyagi, Fumio Shima, and Tomio Sasaki. Brain shift: an error factor during implantation of deep brain stimulation electrodes. *Journal of neurosurgery*, 107(5):989–997, 2007.

[175] Sara Monji Azad, Claudia Scherl, and David Männle. DeformedTissue Dataset, 2025. doi:10.11588/DATA/OAUXWS.

[176] Lingjing Wang, Nan Zhou, Hao Huang, Jifei Wang, Xiang Li, and Yi Fang. Gp-aligner: Unsupervised groupwise nonrigid point set registration based on optimizable group latent descriptor. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–13, 2022.

Appendix A

List of Publications

1. **Sara Monji-Azad**, Marvin Kinz, Siddharth Kothari, Robin Khanna, Amrei Carla Mihan, David Maennel, Claudia Scherl, Juergen Hesser, *DefTransNet: A Transformer-based Method for Non-Rigid Point Cloud Registration in the Simulation of Soft Tissue Deformation*, arXiv preprint arXiv:2502.06336, **2025**.

2. Claudia Scherl, Sonja Ludwig, Jürgen Hesser, **Sara Monji-Azad**, Jan Stallkamp, Frederic Jungbauer, Frederik Enders, Cleo-Aron Weis, Nicole Rotter, *Augmented Reality in head and neck surgery*, Laryngo-Rhino-Otologie, no. 12, vol. 103, **2024**.

3. **Sara Monji-Azad**, Marvin Kinz, David Männle, Claudia Scherl, Jürgen Hesser, *Robust-DefReg: a robust coarse to fine non-rigid point cloud registration method based on graph convolutional neural networks*, Measurement Science and Technology, no. 1, vol. 36, **2024**.

4. **Sara Monji-Azad**, Marvin Kinz, Claudia Scherl, David Männle, Jürgen Hesser, Nikolas Löw, *SynBench: A Synthetic Benchmark for Non-rigid 3D Point Cloud Registration*, arXiv preprint arXiv:2409.14474, **2024**.

5. **Sara Monji-Azad**, David Männle, Jürgen Hesser, Jan Pohlmann, Nicole Rotter, Annette Affolter, Cleo Aron Weis, Sonja Ludwig, Claudia Scherl, *Point cloud registration for measuring shape dependence of soft tissue deformation by digital twins in head and neck surgery*, Biomedicine Hub, no. 1, vol. 9, **2024**.

6. **Sara Monji-Azad**, Marvin Kinz, Jay Makadiya, David Männle, Claudia Scherl, Jürgen Hesser, *An Assessment of Non-Rigid Point Cloud Registration Methods on Two Novel Soft Tissue Deformation Datasets*, **2024**.

7. **Sara Monji-Azad**, Marvin Kinz, Jürgen Hesser, Nikolas Löw, *SimTool: A toolset for soft body simulation using Flex and Unreal Engine*, Software Impacts, no. 100521, vol. 17, **2023**.

8. David Männle, Jan Pohlmann, **Sara Monji-Azad**, Jürgen Hesser, Nicole Rotter, Annette Affolter, Anne Lammert, Benedikt Kramer, Sonja Ludwig, Lena Huber, Claudia Scherl, *Artificial intelligence directed development of a digital twin to measure soft tissue shift during head and neck surgery*, PLOS ONE, vol. 18, no. 8, **2023**.

9. Claudia Scherl, Jan Pohlmann, Jürgen Hesser, **Sara Monji-Azad**, Nicole Rotter, Annette Affolter, Anne Lammert, Lena Huber, Benedikt Kramer, David Männle, *Tissueshift: A shape dependent phenomenon with clinical relevance*, Laryngo-Rhino-Otologie, no. S02, vol. 102, **2023**.

10. **Sara Monji-Azad**, Jürgen Hesser, Nikolas Löw, *A Review of Non-Rigid Transformations and Learning-Based 3D Point Cloud Registration Methods*, ISPRS Journal of Photogrammetry and Remote Sensing, vol. 196, no. 4, pp. 58–72, **2023**.

11. David Männle, Jan Pohlmann, **Sara Monji-Azad**, Nikolas Löw, Nicole Rotter, Jürgen Hesser, Annette Affolter, Anne Lammert, Angela Schell, Benedikt Kramer, Claudia Scherl, *Development of AI based soft tissue shift tracking during surgery to optimize frozen section analysis*, Laryngo-Rhino-Otologie, no. S02, vol. 101, **2022**.