

UNIMORE
UNIVERSITÀ DEGLI STUDI DI
MODENA E REGGIO EMILIA

UNIVERSITY OF MODENA AND REGGIO EMILIA

"Enzo Ferrari" Department of Engineering

Master's Degree in Artificial Intelligence Engineering (LM-32)

Multimodal 3D Registration for Dental Imaging

Supervisor:
Prof. Federico Bolelli

Candidate:
Alex Veronese
Student ID 206814

ACADEMIC YEAR 2025/2026

Abstract

Multimodal 3D Registration for Dental Imaging

Digital dentistry increasingly relies on three-dimensional data acquired from different imaging technologies, each providing complementary information about dental anatomy. Intraoral Scans (IOS) offer detailed surface representations of the visible crowns and gingiva, while Cone-Beam Computed Tomography (CBCT) provides volumetric information about tooth roots, alveolar bone, and surrounding structures. Combining these sources can therefore support a more complete representation of dentoalveolar anatomy and enable advanced applications in orthodontics, treatment planning, and computer-assisted dental procedures.

This thesis addresses the registration of IOS and CBCT data, a challenging problem due to the limited anatomical overlap between the two modalities, their different geometric representations, and the presence of artifacts, missing structures, and large initial pose differences.

To overcome these limitations, a coarse-to-fine framework combining learning-based anatomical landmark detection with robust geometric registration is proposed. Corresponding dental landmarks are independently predicted from the IOS mesh and the CBCT volume and used to estimate an initial rigid transformation through a confidence-weighted RANSAC-Kabsch procedure. The alignment is then refined using point-to-plane Iterative Closest Point on the reconstructed dental surfaces. This design exploits anatomical information to obtain a reliable global initialization while using local surface geometry to improve registration accuracy.

The approach is evaluated on the MICCAI STS 2026 dataset. The results show that landmark-based initialization and surface refinement complement each other and improve robustness in the presence of uncertain correspondences and large initial misalignments.

Overall, this work contributes to the development of automated methods for integrating heterogeneous 3D dental data, combining deep learning and geometric techniques within a unified registration pipeline.

Keywords: Multimodal Registration, 3D Dental Imaging, Intraoral Scans, Cone-Beam Computed Tomography, Deep Learning, Orthodontics

AI usage and disclosure

In accordance with the Guidelines of the University of Modena and Reggio Emilia on the use of artificial intelligence tools (29 July 2026), I hereby declare that I have used the following generative artificial intelligence tools:

- Claude Code, during the development of the application, as programming support, in particular for writing, identifying errors, and commenting on the code. The code generated by the tool was checked and modified where necessary;
- ChatGPT, during the revision phase of the thesis, for checking punctuation and formal consistency, and for suggesting rewordings.

The contents of the thesis are the result of my own work. All content and suggestions obtained were critically reviewed and evaluated by me.

Contents

Abstract ii

AI usage and disclosure iii

1 Introduction 1

1.1 Context and Motivation . . . 1
1.2 Problem Statement . . . 1
1.3 Objectives . . . 3
1.4 Proposed Approach . . . 3
1.5 Main Contributions . . . 4

2 3D Dental Imaging and Multimodal Registration 6

2.1 Introduction . . . 6
2.2 Three-Dimensional Medical Data . . . 6
2.2.1 Volumetric Data . . . 6
2.2.2 Surface-Based Data . . . 7
2.3 Three-Dimensional Dental Imaging . . . 7
2.3.1 Intraoral Scans . . . 7
2.3.2 Cone-Beam Computed Tomography . . . 8
2.4 Data Formats . . . 9
2.4.1 NIfTI Format . . . 9
2.4.2 STL Format . . . 10
2.5 Multimodal Medical Data . . . 10
2.6 IOS–CBCT Integration . . . 11
2.6.1 Shared Dental Anatomy . . . 11
2.6.2 Challenges of Multimodal Dental Data . . . 12
2.6.3 Clinical Applications . . . 12
2.7 Dental Anatomy and Anatomical References . . . 12

2.7.1 Tooth Identification . . . 13
2.7.2 Dental Landmarks . . . 13
2.8 Registration of Multimodal 3D Data . . . 14

3 Related Work **16**
3.1 Introduction . . . 16
3.2 Learning on 3D Dental Data . . . 16
3.2.1 Semantic Segmentation . . . 16
3.2.2 Anatomical Landmark Detection . . . 17
3.3 Deep Learning Architectures for 3D Data . . . 17
3.3.1 Point-Based Networks . . . 18
3.3.2 Transformer-Based Point Cloud Processing . . . 18
3.3.3 Volumetric Medical Image Networks . . . 18
3.4 Three-Dimensional Registration . . . 19
3.4.1 Surface-Based Registration . . . 19
3.4.2 Correspondence-Based Registration . . . 19
3.4.3 Robust Estimation . . . 20
3.5 IOS–CBCT Registration . . . 20
3.5.1 Conventional and Surface-Based Approaches . . . 20
3.5.2 Segmentation-Assisted Coarse-to-Fine Registration . . . 21
3.5.3 Tooth-Wise and Global-to-Local Registration . . . 21
3.5.4 Feature-Based and Learning-Based Registration . . . 21
3.6 Positioning of the Proposed Approach . . . 22

4 Datasets and Data Preparation **23**
4.1 Introduction . . . 23
4.2 MICCAI STS 2026 Challenge Dataset . . . 23
4.3 Intraoral Scan Data . . . 24
4.4 CBCT Data . . . 24
4.5 IOS Normalization Data . . . 25
4.6 Landmark Selection . . . 25

4.7 Pseudo-Label Generation 27
4.8 CBCT Landmark Annotation Representation 27
4.9 Data Preparation and Augmentation 28

5 Proposed Multimodal Registration Framework 30
5.1 Introduction 30
5.2 Framework Overview 30
5.3 IOS-Normalizer 31
5.3.1 PCA-Based Alignment 32
5.3.2 Learned Orientation Correction 33
5.4 IOS-Landmarks 33
5.4.1 Tooth Segmentation and Identification 34
5.4.2 Landmark Prediction and Selection 35
5.5 CBCT-Landmarks 35
5.5.1 Heatmap-Based Landmark Localization 36
5.5.2 Landmark Confidence and Inference 36
5.6 Landmark-Based Coarse Registration 37
5.6.1 Cross-Modal Landmark Correspondences 38
5.6.2 Confidence-Weighted Kabsch Estimation 38
5.6.3 Confidence-Weighted RANSAC-Kabsch 39
5.6.4 Adaptive Consensus and Transformation Re-estimation 40
5.7 Surface-Based Fine Registration 41
5.7.1 Landmark-Guided CBCT Cropping 42
5.7.2 Surface Reconstruction and Preprocessing 42
5.7.3 Point-to-Plane ICP 43
5.7.4 Transformation Composition 43

6 Experiments and Results 45
6.1 Introduction 45
6.2 Experimental Setup 45
6.2.1 Training Protocols 45

6.2.2 Registration Parameters 47
6.3 Evaluation Metrics 47
6.3.1 Mean Translation Error 48
6.3.2 Mean Rotation Error 49
6.4 Ablation Study 49
6.5 Systematic Transformation Labeling Error 50
6.6 Quantitative Results on Validation Set 51
6.6.1 Validation Set 51
6.6.2 Test Set 51
6.7 Qualitative Results 52
6.8 Computational Performance 52

7 Discussion **54**
7.1 Discussion of Main Findings 54
7.2 Landmark Reliability and Spatial Distribution 54
7.3 Role of Surface Refinement 55
7.4 Robustness and Failure Modes 56
7.5 Limitations 56
7.6 Practical Relevance 57

8 Conclusions and Future Work **58**
8.1 Conclusions 58
8.2 Future Work 59
8.3 Final Remarks 59

A Challenge Paper Publication **61**

Acknowledgements **72**

References **73**

List of Figures

2.1 Intraoral Scans examples from the MICCAI STS 2026 Challenge dataset. 8
2.2 Cone-Beam CT examples from the MICCAI STS 2026 Challenge dataset. 9
2.3 IOS–CBCT Registration. 11
2.4 FDI tooth numbering. 13
2.5 Predicted Dental Landmarks Example for Patient case 027. 14

4.1 Example of IOS with impression material on the surface (Patient case 118). 26
4.2 Palatal and Lingual Landmarks distribution across the STS 2026 training unlabeled set. 26
4.3 Discarded pseudo-labeled Patient case 272 due to temporal discrepancy between the IOS and CBCT acquisitions. 28

5.1 Overview of the proposed coarse-to-fine IOS–CBCT registration framework. . . . 31
5.2 PointNet-style architecture. 34
5.3 Segmentation Example of the Lower Jaw for Patient case 027. 35
5.4 Predicted Landmarks for Patient case 027. 36
5.5 *L2L-Registration* module in detail. 37

6.1 Qualitative results, one patient per row (P027 and P214 from the top). 52

1. Introduction

1.1 Context and Motivation

The increasing digitalization of dentistry has led to the widespread use of three-dimensional imaging technologies for diagnosis, treatment planning, and computer-assisted procedures. Different acquisition modalities provide different representations of the same patient anatomy, each with specific advantages and limitations. The ability to relate these heterogeneous sources of information is therefore an important component of modern digital dental workflows.

Among the imaging technologies commonly used in dentistry, Intraoral Scans (IOS) and Cone-Beam Computed Tomography (CBCT) provide particularly complementary information [3, 6, 17]. IOS acquisitions generate detailed three-dimensional representations of the visible surfaces of the dental crowns and gingiva, making them suitable for applications that depend on accurate external morphology [4]. On the other hand, CBCT provides a volumetric representation of dentoalveolar anatomy and makes it possible to analyze structures that are not directly visible on the oral surface, such as tooth roots and alveolar bone.

The integration of the two modalities can therefore provide a more complete three-dimensional representation of dental anatomy than either source considered independently. Such an integrated representation can support digital workflows in orthodontics, implant planning, orthognathic procedures, and longitudinal assessment of dental changes. To combine the information, however, the two acquisitions must first be expressed in a common coordinate system. In this thesis, the problem is formulated as an automatic rigid registration task. The objective is to estimate the spatial transformation that maps an IOS mesh into the coordinate system of the corresponding CBCT volume without requiring a manually provided initial alignment.

1.2 Problem Statement

Automatic IOS–CBCT registration is challenging because the two modalities differ in representation, coverage, and image quality. An intraoral scan is an explicit surface model of visible anatomy,

whereas a CBCT acquisition is a volumetric image containing both dental and surrounding anatomical structures. The two inputs therefore cannot be compared directly through identical geometric or intensity representations.

A second difficulty is the limited anatomical overlap between the modalities. The IOS mainly describes the externally visible crown surfaces and gingiva, while the CBCT also contains tooth roots, alveolar bone, and surrounding tissues. Consequently, only a subset of the CBCT anatomy has a direct counterpart in the intraoral scan. Using the complete CBCT volume or reconstructed surface during registration can therefore introduce structures with no corresponding IOS geometry.

The shared anatomy is also not free from ambiguity. Teeth may exhibit similar local shapes, and approximately symmetric dental structures may be confused when correspondence is established using geometry alone. Missing teeth, restorations, orthodontic appliances, incomplete acquisitions, and imaging artifacts can further reduce the number and quality of usable correspondences.

Finally, the two acquisitions may exhibit substantial differences in their initial position and orientation. Local surface-registration methods such as Iterative Closest Point (ICP) [2] depend on the current spatial configuration to establish correspondences and can converge to incorrect solutions when initialization is poor. A reliable global registration stage is therefore required before dense local refinement can be applied.

The task can be expressed by considering an IOS point $\mathbf{x}_{\text{IOS}} \in \mathbb{R}^3$ and a rigid transformation composed of a rotation $\mathbf{R}$ and a translation $\mathbf{t}$ such that

$$\mathbf{x}_{\text{CBCT}} = \mathbf{R}\mathbf{x}_{\text{IOS}} + \mathbf{t}, \tag{1.1}$$

with

$$\mathbf{R} \in \text{SO}(3), \qquad \mathbf{t} \in \mathbb{R}^3. \tag{1.2}$$

The corresponding homogeneous transformation represents the mapping between the two coordinate systems. The objective of the proposed methodology is to estimate this transformation automatically from anatomical and geometric information extracted from the two acquisitions.

1.3 Objectives

The main objective of this thesis is to investigate an automatic and robust method for multimodal three-dimensional registration of IOS and CBCT data. Rather than directly matching the two heterogeneous representations, the proposed approach exploits anatomical structures that can be identified independently in both modalities. The specific objectives are:

- To identify corresponding anatomical information in IOS and CBCT data despite their different geometric representations;
- To estimate a reliable global registration in the presence of large initial position and orientation differences;
- To reduce the influence of inaccurate or inconsistent landmark predictions during transformation estimation;
- To exploit dense dental surface geometry to refine the alignment obtained from sparse anatomical correspondences; and
- To evaluate the individual and combined contribution of robust landmark registration and surface refinement in the semi-supervised setting of the MICCAI STS 2026 Challenge.

These objectives lead to a coarse-to-fine formulation in which semantic anatomical information is first used to recover the global pose, while local geometry is introduced only after a suitable initialization has been established.

1.4 Proposed Approach

The proposed methodology follows a landmark-guided coarse-to-fine strategy. The IOS and CBCT are first processed through separate modality-specific pipelines. The IOS is normalized into a canonical orientation and analyzed to identify individual teeth and corresponding anatomical landmarks on their surfaces. In parallel, a dedicated volumetric landmark detector localizes the same landmark classes directly within the CBCT acquisition. The semantic identities associated with the detected landmarks make it possible to establish cross-modal correspondences without requiring the

two inputs to be initially close in space. The resulting correspondences are filtered using confidence information and then processed by a confidence-weighted RANSAC–Kabsch procedure [8, 14, 13]. RANSAC identifies a geometrically consistent subset of landmark pairs, while the weighted Kabsch estimator computes the corresponding rigid transformation. This stage provides the global alignment between the modalities. The inlier landmarks selected during coarse registration are also used to restrict the CBCT volume to the region relevant for local surface matching. A surface is reconstructed from this crop, while the IOS is transformed according to the estimated coarse alignment. The two surfaces are then processed as point clouds and refined using point-to-plane ICP. The final transformation is obtained by composing the orientation correction, landmark-based transformation, and local ICP correction. The resulting framework deliberately separates semantic global alignment from geometric local refinement. Anatomical landmarks provide sparse but meaningful correspondences that can recover large pose differences, whereas the surface stage uses denser local information to improve the final alignment.

1.5 Main Contributions

The main contributions of this thesis are summarized as follows:

- A complete automatic framework for multimodal IOS–CBCT registration combining learning-based anatomical analysis with geometric registration;
- A modality-specific strategy for identifying corresponding dental landmarks independently from the IOS mesh and the CBCT volume;
- A confidence-aware landmark registration stage based on weighted Kabsch estimation embedded in an adaptive RANSAC procedure;
- A landmark-guided surface-refinement stage that uses the robust landmark consensus to restrict CBCT reconstruction to the anatomically relevant region;
- An experimental analysis of the complementary contributions of robust landmark selection and point-to-plane ICP through controlled ablation experiments.

The remainder of the thesis introduces the underlying medical and dental data, reviews the relevant literature, describes the datasets and preparation procedures, presents the proposed registration framework, and evaluates its performance experimentally.

2. 3D Dental Imaging and Multimodal Registration

2.1 Introduction

Three-dimensional medical information can be represented in different forms depending on the acquisition technology. Radiographic techniques generally produce volumetric images, whereas optical acquisition systems can reconstruct explicit surfaces. These representations encode different aspects of anatomy and are therefore complementary rather than interchangeable. This distinction is particularly relevant in digital dentistry, where Intraoral Scans (IOS) and Cone-Beam Computed Tomography (CBCT) are used to describe different portions of the same dentoalveolar anatomy.

This chapter introduces the characteristics of these data, explains the concept of multimodal integration, and describes the anatomical references that are relevant to the registration problem addressed in this thesis.

2.2 Three-Dimensional Medical Data

Three-dimensional medical data provide a digital representation of anatomy that can be visualized and processed computationally. The representation depends on the physical acquisition process and determines which anatomical properties can be directly observed.

2.2.1 Volumetric Data

A volumetric medical image represents a region of the body on a regular three-dimensional grid of voxels. Each voxel stores an intensity value related to the underlying acquisition. In addition to voxel intensities, the image contains spatial metadata such as voxel spacing, image origin, and orientation, which define the relationship between discrete image indices and physical coordinates.

Volumetric data have the important advantage of representing internal anatomy. Structures that are not externally visible can therefore be analyzed within the same three-dimensional representation. At the same time, the apparent boundaries of anatomical structures may be influenced by acquisition

resolution, noise, partial-volume effects, and imaging artifacts. These characteristics are particularly relevant when volumetric data are subsequently converted into explicit geometric surfaces.

2.2.2 Surface-Based Data

Surface-based data describe the geometry of an externally visible boundary rather than the complete volume surrounding it. A common representation is a triangular mesh, consisting of three-dimensional vertices connected by triangular faces. The mesh explicitly describes the geometry and connectivity of the reconstructed surface. For computational processing, a surface mesh can also be sampled as a point cloud. This representation is particularly suitable for point-based learning methods and three-dimensional geometric registration because the spatial coordinates of the visible anatomy can be processed directly.

The main limitation of a surface representation is that structures not directly observable by the acquisition process are absent. Surface and volumetric data therefore provide different and complementary descriptions of anatomy.

2.3 Three-Dimensional Dental Imaging

Digital dentistry relies on multiple three-dimensional acquisition technologies because no single modality provides an equally detailed description of all relevant anatomical structures. In the present work, IOS and CBCT are the two modalities considered for multimodal registration.

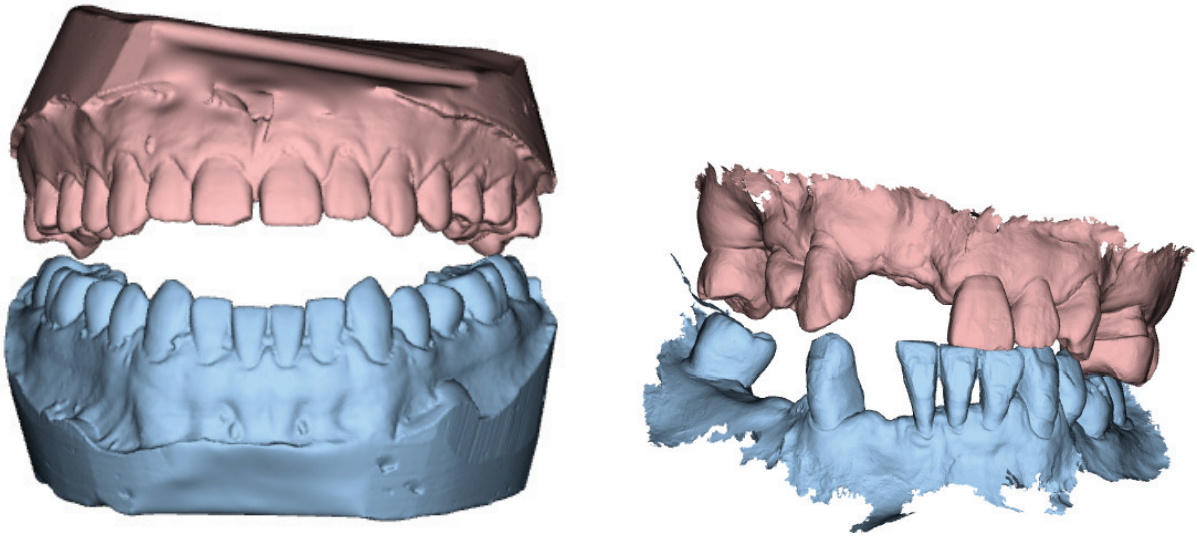
2.3.1 Intraoral Scans

Intraoral scans are acquired using optical scanning systems that progressively reconstruct the visible surfaces of the oral cavity from multiple viewpoints. The resulting data are generally represented as a three-dimensional surface mesh of the dental arches and surrounding gingiva.

The principal characteristic of IOS is the detailed description of visible crown morphology. Tooth surfaces, occlusal regions, arch geometry, and gingival areas can be directly visualized and processed in three dimensions. This makes IOS particularly suitable for digital workflows that depend on external dental geometry. For computational processing, the surface mesh can also be

sampled or converted into a point cloud, allowing the spatial coordinates of the observed anatomy to be processed directly by point-based methods.

The information provided by IOS is nevertheless limited to surfaces that can be optically observed. Tooth roots, alveolar bone, and deeper anatomical structures are not directly represented. Full-arch scanning can also introduce reconstruction inconsistencies, while missing teeth, restorations, orthodontic appliances, impression material, or incomplete acquisitions may modify the observed surface.



(a) Patient case 027.

(b) Patient case 215.

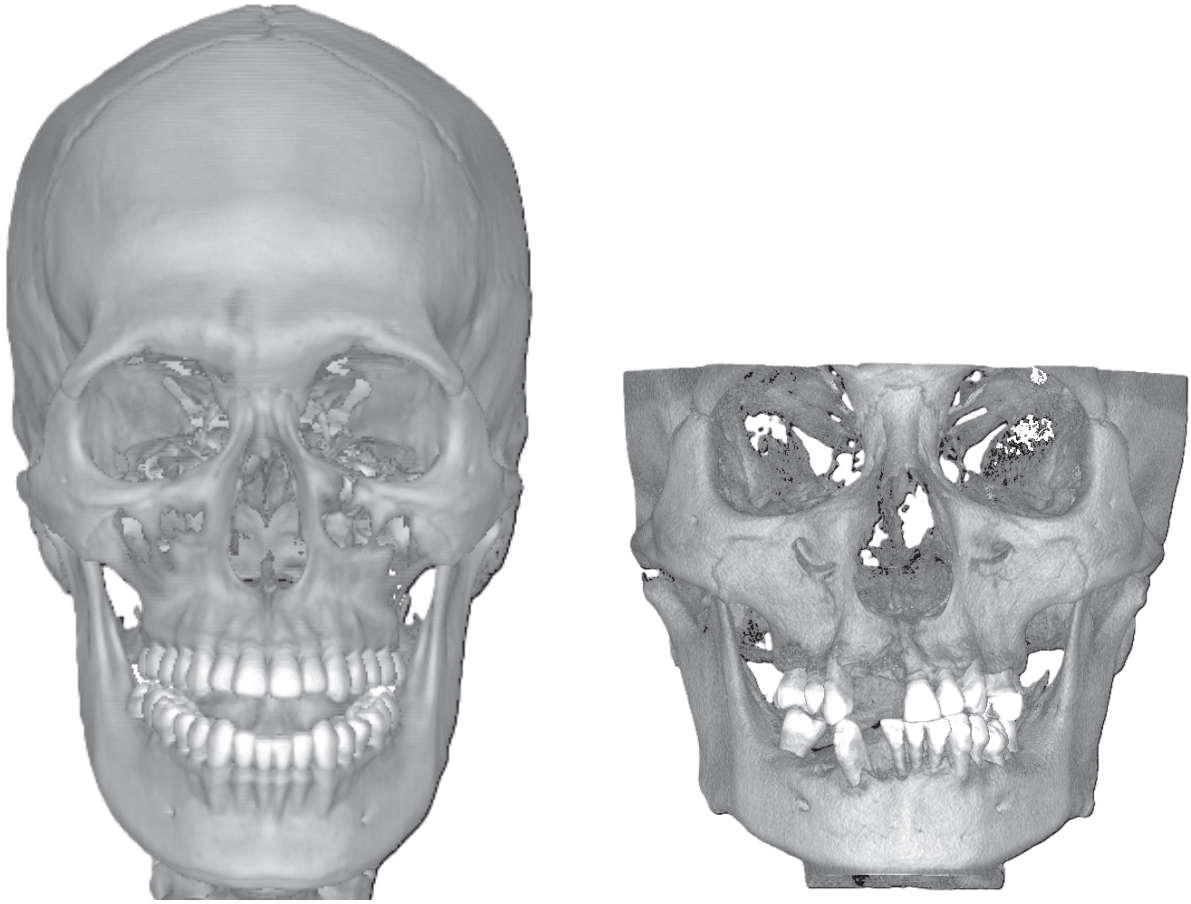
Figure 2.1: Intraoral Scans examples from the MICCAI STS 2026 Challenge dataset.

2.3.2 Cone-Beam Computed Tomography

Cone-Beam Computed Tomography provides a volumetric representation of the acquired dentoalveolar region. In comparison with IOS, CBCT extends the available information below the visible surface and can therefore represent tooth roots, alveolar bone, and surrounding anatomical structures. The volumetric nature of CBCT makes it suitable for analyses that require knowledge of the internal relationship between dental and skeletal structures.

However, the crown surfaces represented in CBCT generally do not provide the same geometric detail as an optical surface scan. Voxel spacing, effective spatial resolution, noise, partial-volume

effects, and metal artifacts can influence the apparent boundaries of dental structures. In particular, metallic restorations and orthodontic appliances can produce strong local artifacts. The resulting image quality may therefore vary considerably between anatomical regions and individual cases.



(a) Patient case 027.

(b) Patient case 215.

Figure 2.2: Cone-Beam CT examples from the MICCAI STS 2026 Challenge dataset.

2.4 Data Formats

2.4.1 NIfTI Format

The CBCT volumes used in this work are provided in the Neuroimaging Informatics Technology Initiative (NIfTI) format, with files stored using the `.nii.gz` extension. NIfTI represents the

CBCT as a three-dimensional volume composed of voxels, with each voxel associated with an intensity value. The format also stores spatial metadata, including information about voxel spacing, orientation, and the mapping between voxel indices and physical coordinates. This information is essential when relating the CBCT volume to geometric data acquired from other modalities.

2.4.2 STL Format

The intraoral scans are represented as three-dimensional surface meshes and stored in the Stereolithography (STL) format. An STL file represents a surface as a collection of triangular facets. Each facet is defined by three vertices and an associated surface normal. The vertices determine the geometric surface, while the normal specifies the orientation of the facet. For registration and point-based processing, the mesh can subsequently be sampled to obtain a point-cloud representation.

2.5 Multimodal Medical Data

Multimodal medical data arise when the same anatomical region is acquired using different technologies. The motivation for combining such data is that each modality can provide information that is unavailable or less precise in the other. The goal is therefore to express complementary information within a common spatial context rather than to replace one modality with another.

Different acquisition technologies may represent anatomy using different spatial resolutions, coordinate systems, fields of view, and data structures. A detailed surface in one modality may correspond only to an intensity transition or a partially represented structure in another.

In the registration setting considered in this thesis, multimodal integration requires an anatomical notion of correspondence together with a geometric transformation relating the coordinate systems. In the dental setting, IOS and CBCT provide a clear example of this complementarity. IOS contributes detailed external crown geometry, whereas CBCT provides volumetric information about internal and supporting structures. Their combination can therefore produce a more comprehensive three-dimensional representation than either acquisition considered independently, as shown in Fig. 2.3.

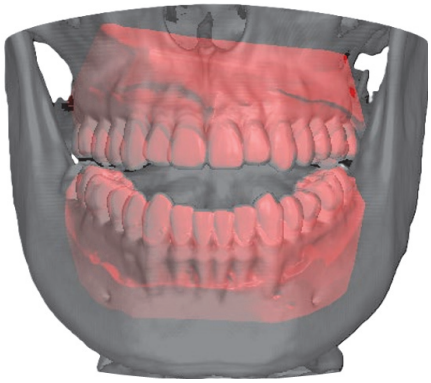


Figure 2.3: IOS–CBCT Registration.

2.6 IOS–CBCT Integration

The integration of IOS and CBCT is a specific multimodal registration problem in which an optical surface representation must be spatially related to a volumetric radiographic acquisition. Although the two modalities may be acquired for the same patient and anatomical region, they differ substantially in their fields of view and in the way anatomical structures are represented. Consequently, only a subset of the anatomy is directly useful for establishing correspondences between the two modalities.

2.6.1 Shared Dental Anatomy

For the registration problem considered in this thesis, the principal and most reliable overlapping anatomical structures are the dental crowns. IOS directly captures their external surfaces, while the corresponding crown anatomy is represented within the CBCT volume. In contrast, tooth roots, alveolar bone, and deeper internal anatomy have no direct surface counterpart in the IOS.

The regular organization of the dentition also provides anatomical context that can be exploited to establish correspondences between the modalities. Individual teeth have identifiable anatomical positions, allowing observations from IOS and CBCT to be associated according to their dental identity.

2.6.2 Challenges of Multimodal Dental Data

The different representations and spatial resolutions of IOS and CBCT make direct geometric comparison difficult. Detailed surface morphology visible in IOS may be represented less precisely by anatomical boundaries within the CBCT volume, while the complete CBCT contains substantial anatomy that is not present in the IOS.

Dental morphology introduces an additional source of ambiguity. Different teeth may have similar shapes, and approximately symmetric regions can produce geometrically plausible but anatomically incorrect matches when semantic information is not considered. Missing teeth, restorations, orthodontic appliances, scan incompleteness, and CBCT artifacts can further reduce the number and reliability of usable correspondences.

The acquisitions may also differ considerably in their initial position and orientation. Methods based primarily on local spatial proximity are therefore not sufficient when the two representations are initially far apart.

2.6.3 Clinical Applications

An accurate spatial relationship between IOS and CBCT can support workflows that require simultaneous access to external crown morphology and internal dentoalveolar anatomy. In orthodontics, such integration can relate visible tooth positions to root orientation and surrounding bone. In implant planning, it can connect detailed surface information with the internal structures visible in CBCT.

More generally, a common coordinate system enables measurements and analyses that combine information from both modalities. The purpose of multimodal registration is consequently not limited to visual superimposition. A reliable transformation provides the geometric basis for subsequent multimodal analysis.

2.7 Dental Anatomy and Anatomical References

The structured organization of the dentition provides semantic information that is particularly useful for automatic multimodal analysis. Individual teeth can be identified according to their anatomical

position, while characteristic locations on each tooth can serve as reproducible references.

2.7.1 Tooth Identification

Permanent teeth are identified using the FDI two-digit notation, as shown in Fig. 2.4. The first digit specifies the dental quadrant, while the second digit indicates the position of the tooth within that quadrant, progressing from the central incisor toward the posterior region.

FDI identifiers provide information that is not contained in geometry alone. Two teeth with similar local geometry can nevertheless be distinguished when their anatomical identities are known, because their positions within the dental arch differ. This semantic identity can therefore be used to establish or constrain cross-modal correspondences between IOS and CBCT.

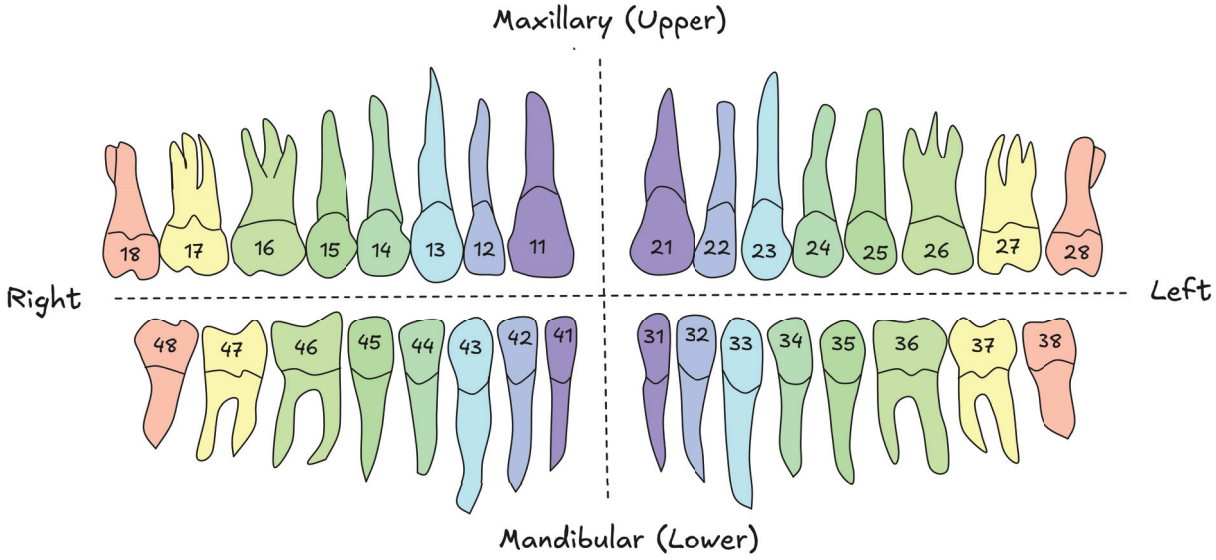


Figure 2.4: FDI tooth numbering.

2.7.2 Dental Landmarks

A dental landmark is a reproducible three-dimensional reference associated with a defined anatomical location. Depending on the landmark definition, such a point may correspond to a characteristic region, surface extremum, or other anatomically meaningful location on a tooth.

The main advantage of landmarks in multimodal registration is their semantic identity. A landmark defined for a specific tooth and anatomical position can be associated with the corresponding

anatomical reference in another modality even when the surrounding representations differ substantially.

Landmarks also reduce the correspondence problem from the full surface to a sparse set of meaningful points. However, their usefulness depends on availability and localization accuracy. Missing anatomy, acquisition artifacts, and modality-specific image quality can make some landmarks more difficult to detect reliably than others.

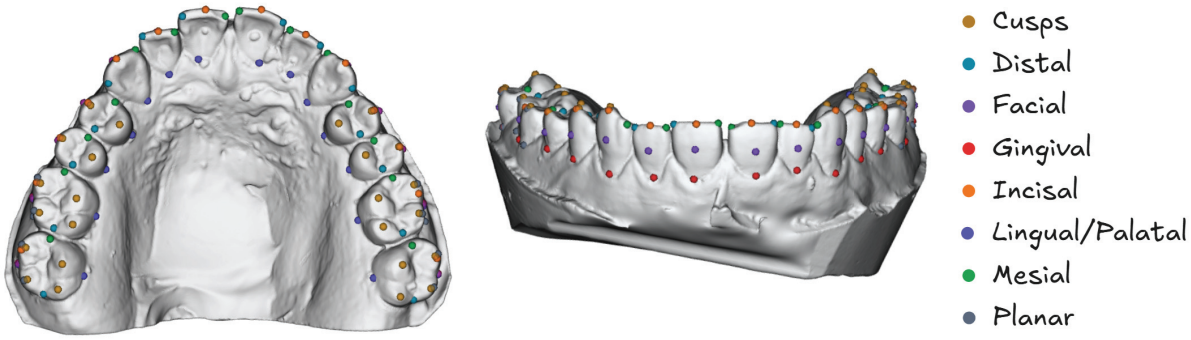


Figure 2.5: Predicted Dental Landmarks Example for Patient case 027.

2.8 Registration of Multimodal 3D Data

Three-dimensional registration establishes the spatial relationship between two acquisitions of the same subject by aligning corresponding anatomical structures within their overlapping fields of view. One acquisition is selected as the reference, while the source data are transformed into the reference coordinate system.

For the problem considered in this thesis, the IOS is treated as the source and the CBCT as the target. The registration is assumed to be rigid; therefore, the transformation is composed of a rotation and a translation and preserves distances and angles within the source geometry.

Multimodal registration differs from registration between repeated acquisitions of the same modality because corresponding structures may not have the same appearance or representation. In the IOS–CBCT setting, a surface point in the IOS does not directly correspond to a CBCT voxel through a simple identity relation. Instead, correspondence must be established between the IOS surface geometry and anatomical structures or boundaries represented within the CBCT volume. A

registration method must therefore exploit some form of common information, such as anatomical landmarks, segmented structures, geometric features, or learned representations.

A second distinction is between global and local alignment. In this context, global registration refers to methods intended to recover substantial differences in relative pose, whereas local refinement assumes an already approximate alignment and improves the transformation around that solution. Combining the two objectives is particularly useful for IOS–CBCT registration because the initial relative pose can be unknown and the shared anatomy is only partially overlapping.

The specific registration methods and previous IOS–CBCT approaches relevant to this thesis are reviewed in Chapter 3.

3. Related Work

3.1 Introduction

Automatic integration of three-dimensional dental data combines several established problems in medical image analysis. In the context of IOS–CBCT registration, useful anatomical information must first be identified within two heterogeneous representations, after which the resulting information can be used to estimate their relative pose and refine the alignment.

This chapter reviews the methods most relevant to the proposed framework. Learning approaches for three-dimensional dental data are introduced first, with emphasis on semantic segmentation and anatomical landmark localization. The principal network families used for point-based and volumetric data are then summarized. Classical three-dimensional registration methods are discussed next, followed by a review of approaches specifically developed for IOS–CBCT integration. The chapter concludes by positioning the proposed method with respect to these approaches.

3.2 Learning on 3D Dental Data

Three-dimensional dental analysis includes tasks operating at different levels of anatomical abstraction. Semantic segmentation assigns an anatomical class to each spatial element, whereas landmark detection identifies a smaller number of clinically or anatomically meaningful points. Both are relevant to multimodal registration because they transform an unstructured acquisition into information that can be associated across modalities.

3.2.1 Semantic Segmentation

Semantic segmentation assigns an anatomical label to each element of a three-dimensional representation. Depending on the modality, predictions may be associated with surface vertices, point-cloud samples, or voxels in a volumetric image.

In dental applications, segmentation can distinguish individual teeth, gingiva, bone, and other structures. Tooth-level segmentation is especially relevant to multimodal registration because it

provides anatomical identity in addition to geometry. This reduces ambiguity in dental arches, where different teeth may have similar shapes and left–right symmetry can make purely geometric matching unreliable.

Segmentation has consequently been used as a preprocessing step in several automatic IOS–CBCT registration methods. Instead of registering the complete raw acquisitions, these approaches first isolate the dental structures that are relevant to both modalities and then perform geometric alignment on the selected regions.

3.2.2 Anatomical Landmark Detection

Landmark detection provides a complementary representation in which a small set of anatomically meaningful points is localized instead of labeling every spatial element. When the same landmark definition is available in different modalities, the resulting points provide explicit semantic correspondences.

For surface data, landmark localization can be formulated as point-wise or vertex-wise probability prediction, with high-response regions indicating the target location. For volumetric medical data, a common formulation is heatmap regression, in which each landmark is associated with a three-dimensional response map and the predicted coordinate is obtained from its maximum response.

The principal advantage of landmarks for registration is that correspondence can be established from anatomical identity rather than spatial proximity. Their limitation is that automatic predictions are uncertain and may vary with anatomy, data quality, and acquisition artifacts. Robust transformation estimation is therefore important when sparse predicted landmarks are used as geometric constraints.

3.3 Deep Learning Architectures for 3D Data

The neural architecture used for a three-dimensional task depends strongly on the underlying representation. Point clouds are unordered sets of spatial samples and therefore require permutation-invariant processing, whereas volumetric medical images are defined on regular grids and can be processed using three-dimensional convolutions.

3.3.1 Point-Based Networks

PointNet introduced a general architecture for processing unordered point sets directly [5]. Shared point-wise transformations are followed by symmetric global aggregation, providing invariance to the ordering of the input samples.

PointNet++ extended this concept through hierarchical neighborhood grouping and local feature aggregation [18]. The resulting representation captures geometric information at multiple spatial scales and is particularly suitable for tasks in which both global shape and local morphology are relevant.

Point-based processing is well suited to IOS data because a surface mesh can be sampled directly into a point cloud without converting the geometry to a dense volumetric grid. This preserves the explicit three-dimensional structure of the visible dental surfaces.

3.3.2 Transformer-Based Point Cloud Processing

Transformer-based point-cloud models use attention mechanisms to represent relationships between spatially distributed features. Point Transformer architectures adapt this principle to irregular three-dimensional data by incorporating spatial relationships into the attention mechanism.

Point Transformer V3 provides a more efficient formulation for large point clouds while retaining hierarchical geometric modeling capabilities [20]. These properties are relevant to dental analysis, where local tooth morphology and broader arch-level context are both useful.

3.3.3 Volumetric Medical Image Networks

Volumetric medical images are naturally processed using three-dimensional convolutional architectures. Encoder–decoder networks based on the U-Net family can aggregate contextual information while recovering spatial resolution for dense prediction tasks [19].

The nnU-Net framework introduced a self-configuring strategy in which important preprocessing, architecture, and training choices are adapted to the target dataset [10]. This design has become a common basis for three-dimensional medical image analysis.

Landmark localization can be formulated within the same general framework by predicting one heatmap per anatomical landmark rather than a segmentation mask. nnLandmark extends the

self-configuring philosophy of nnU-Net to three-dimensional anatomical landmark detection [7].

3.4 Three-Dimensional Registration

Three-dimensional registration estimates the spatial transformation relating two geometric or medical acquisitions. Two broad categories are particularly relevant to this thesis: correspondence-based methods, which estimate a transformation from matched points or features, and surface-based methods, which exploit dense geometric information.

3.4.1 Surface-Based Registration

Iterative Closest Point (ICP) is one of the most established methods for rigid surface alignment [2]. The algorithm alternates between establishing spatial correspondences and estimating a transformation that reduces the alignment error. Point-to-point and point-to-plane formulations are commonly used.

ICP is attractive because it does not require explicitly defined semantic correspondences. However, it is local and depends on a sufficiently accurate initialization. Large pose differences, partial overlap, repetitive geometry, and unrelated structures can lead to incorrect correspondences and convergence toward a local solution. In multimodal IOS–CBCT dental registration, these limitations are amplified because only a subset of the CBCT anatomy corresponds to the IOS.

3.4.2 Correspondence-Based Registration

When corresponding points are known, the rigid transformation can be estimated directly from their coordinates. The Kabsch algorithm provides a closed-form least-squares solution for the optimal rotation between two point sets [14, 13]. The solution is obtained from the centered correspondences and the singular value decomposition of their cross-covariance matrix.

Correspondence-based estimation is particularly attractive when anatomical landmarks provide explicit semantic matches. Nevertheless, the solution is sensitive to incorrect or poorly localized correspondences. A small number of erroneous points can substantially affect the estimated transformation when the total number of available landmarks is limited.

3.4.3 Robust Estimation

RANSAC provides a general strategy for estimating geometric models when the correspondence set may contain outliers [8]. Candidate transformations are generated from small subsets of observations and evaluated according to the number of observations that are consistent with the candidate model.

For rigid registration, RANSAC can therefore be combined with Kabsch estimation: Kabsch provides the transformation for a selected subset, while RANSAC identifies a consensus of geometrically compatible correspondences. This combination is suitable for automatic landmark registration because the semantic identity of a point does not guarantee that its predicted location is accurate.

3.5 IOS–CBCT Registration

Automatic registration of IOS and CBCT has been investigated using a range of strategies, from manually defined correspondences and direct surface alignment to segmentation-assisted and learning-based methods. A recent systematic review summarizes the development of the field and the range of techniques adopted for CBCT–IOS registration [21]. The main approaches relevant to this thesis are summarized below.

3.5.1 Conventional and Surface-Based Approaches

Conventional workflows commonly rely on manually selected anatomical references or surfaces reconstructed from CBCT. A small number of corresponding landmarks can provide an initial transformation, after which ICP refines the surface alignment.

Manual references can be accurate but introduce user interaction and reduce reproducibility at scale. Surface-based strategies remove part of this interaction but remain dependent on the quality of the reconstructed CBCT surface and on initialization. Direct ICP is therefore difficult to apply when the relative pose is initially unknown or the overlap between the acquisitions is limited.

3.5.2 Segmentation-Assisted Coarse-to-Fine Registration

Segmentation-assisted approaches introduce an anatomical preprocessing stage before geometric registration. Kim et al. combined deep-learning-based tooth segmentation with RANSAC-based coarse alignment and ICP refinement [15]. By restricting the registration to identified dental structures, the method reduces the influence of unrelated CBCT anatomy.

The approach illustrates the value of combining learned semantic information with classical geometric registration. Deep learning is used to identify anatomically meaningful structures, while robust geometric methods estimate the transformation and refine it locally. The coarse-to-fine organization also reduces the sensitivity of ICP to large initial pose differences.

3.5.3 Tooth-Wise and Global-to-Local Registration

Another line of work explicitly uses individual teeth as anatomical units. Jang et al. proposed a fully automatic framework based on individual-tooth segmentation and identification, using a global-to-local strategy and addressing stitching errors in full-arch intraoral impressions [11].

Tooth-wise registration introduces semantic organization into the geometric problem and can restrict matching to smaller, interpretable anatomical regions. Its effectiveness consequently depends on the accuracy of tooth segmentation and identification. Errors in these preprocessing stages can propagate into the subsequent registration.

The use of individual teeth as semantic units motivates the use of even more localized anatomical references, such as reproducible landmarks on selected teeth.

3.5.4 Feature-Based and Learning-Based Registration

Feature-based methods attempt to establish correspondences from local geometric descriptors. Liu et al. investigated geometric feature matching within a multimodal tooth–bone reconstruction framework, including descriptors such as Fast Point Feature Histograms (FPFH) [16]. Such approaches can reduce the need for manually specified anatomical points, but descriptors derived from geometry alone may be affected by the different representations of IOS and CBCT and by imaging artifacts.

Learning-based registration methods attempt to learn more discriminative representations from data. Ji et al. investigated PointNet++ feature encoding combined with differentiable SVD-based transformation estimation for CBCT–IOS registration [12]. George et al. investigated a semi-supervised CBCT–IOS approach based on PointNetLK, combining point-cloud representations with iterative alignment [9]. These methods move toward learned or partially learned correspondence estimation. At the same time, the limited overlap between IOS and CBCT, the heterogeneous representations, large initial pose differences, and scarcity of paired data with accurate transformations remain important challenges.

3.6 Positioning of the Proposed Approach

The reviewed literature shows a progression from manual and purely geometric registration toward approaches that incorporate semantic and learned information. Surface methods provide dense local refinement but remain sensitive to initialization and partial overlap. Segmentation-assisted and tooth-wise methods improve the anatomical interpretation of the data, while feature- and learning-based approaches attempt to derive correspondences automatically.

The approach developed in this thesis follows the same general coarse-to-fine direction but uses sparse anatomical landmarks as the common representation between the modalities. Landmarks are predicted independently in IOS and CBCT and associated according to tooth identity and landmark type. This avoids requiring the complete surface or intensity representation to be directly comparable across modalities.

The predicted correspondences are then processed using confidence-aware robust geometric estimation. Confidence weighting accounts for differences in prediction reliability, while adaptive RANSAC identifies a geometrically consistent consensus before weighted Kabsch re-estimation. The resulting inliers are subsequently reused to define the CBCT region for local surface refinement with point-to-plane ICP.

The distinguishing aspect of the proposed framework is therefore the explicit separation of semantic global initialization and dense local refinement, with the same robust landmark consensus providing the transition between the two stages. The complete methodology is presented in Chapter 5.

4. Datasets and Data Preparation

4.1 Introduction

The development of multimodal registration methods requires paired acquisitions describing the same underlying anatomy in different representations. In this thesis, each case contains an Intraoral Scan (IOS), represented as a three-dimensional surface, and a corresponding Cone-Beam Computed Tomography (CBCT) acquisition, represented as a volumetric image.

The experiments use the data provided for Task 2 of the MICCAI STS 2026 Challenge. The challenge adopts a semi-supervised setting in which a limited labeled subset is accompanied by a larger unlabeled subset. The labeled and validation data were kept separate from model development, while the unlabeled training data were used to fine-tune the IOS orientation component and to construct the pseudo-labeled data required for CBCT landmark detection.

This chapter describes the organization of the challenge data, the characteristics relevant to the two branches of the pipeline, the landmark-selection procedure, the construction of CBCT pseudo-labels, and the final data preparation strategy.

4.2 MICCAI STS 2026 Challenge Dataset

Task 2 of the MICCAI STS 2026 Challenge addresses rigid registration between IOS and CBCT data in a semi-supervised setting. The available data consist of 30 labeled training cases, 300 unlabeled training cases, and 50 validation cases, as reported in Tab. 4.1. The 30 labeled training cases are associated with reference 4×4 transformation matrices. These matrices were not used to train or fine-tune the components of the proposed framework; they were reserved for controlled quantitative evaluation and ablation experiments. The 300 unlabeled training cases do not provide target registration transformations and were therefore used for a pseudo-labeling strategy. The official validation set and test set contain 50 and 99 cases, respectively, for which the reference transformations are hidden, and evaluation is performed through the challenge system.

Table 4.1: MICCAI STS 2026 Challenge dataset.

Set	Cases	Labels	Training	Evaluation
Train Labeled	30	✓	✗	✓
Train Unlabeled	300	✗	✓	✗
Validation	50	✗	✗	✓
Test	99	✗	✗	✓

4.3 Intraoral Scan Data

The IOS component of each case is an explicit three-dimensional surface representation of the visible dentition and surrounding gingiva for the lower and upper jaw. The geometry is used by the IOS branch at three successive stages: canonical orientation, tooth-level semantic analysis, and anatomical landmark prediction.

The challenge scans exhibit variability in completeness and dental condition. During development, some acquisitions contained residual orthodontic brackets or dental impression material, particularly in anterior regions. Such effects can reduce the reliability of surface landmarks and motivated the restricted landmark set adopted in this work.

The IOS landmark models used in the registration pipeline operate on the normalized surface geometry. The normalization procedure itself is described in Chapter 5, while its training configuration is reported in Chapter 6.

4.4 CBCT Data

The CBCT component is a three-dimensional volumetric image with associated spatial metadata. The volume includes the dental crowns as well as structures that are absent from the IOS, such as tooth roots, alveolar bone, and surrounding anatomy.

CBCT spatial metadata are required at several stages of the proposed pipeline. During landmark

detection, predicted voxel coordinates are converted into physical coordinates before cross-modal correspondence is established. During surface reconstruction, the same spatial information is used to map the extracted CBCT surface into the physical coordinate system of the volume.

The larger anatomical coverage of CBCT is relevant to both training and registration. Since only part of the volume has a corresponding IOS surface, the final refinement stage operates on a landmark-guided crop rather than on the complete CBCT acquisition.

4.5 IOS Normalization Data

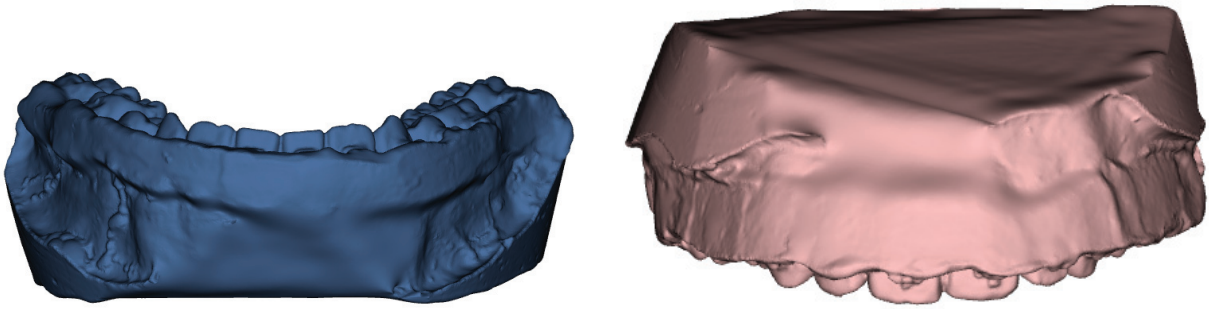
The IOS-Normalizer was initialized from a pretrained checkpoint obtained from the Teeth3DS dataset and subsequently adapted to the orientation characteristics of the challenge data [1]. A subset of 60 IOS cases was selected from the unlabeled training portion and manually oriented to define the canonical reference configuration. Of these 60 scans, 50 were used for fine-tuning and 10 for validation. Each scan was sampled using Farthest-Point Sampling, and synthetic orientation examples were generated by either keeping the canonical orientation or applying a 180° rotation around one of the three Cartesian axes. These four transformations form the target orientation classes used by the classifier described in Section 5.3.2.

The normalization subset provides supervision only for the orientation-correction problem. The reference transformations of the labeled registration cases were not used in this process.

4.6 Landmark Selection

The pretrained IOS landmark detector can predict multiple landmark categories for individual teeth. For multimodal registration, however, only landmarks that are sufficiently available and reproducible across the considered data are useful.

Landmark availability was therefore evaluated separately in the four dental hemi-arches. From visual inspection of the available data only the Lingual and Palatal landmark of each selected tooth is used, respectively for the lower and upper jaw. This choice also reduces the influence of landmarks located near regions that can be affected by orthodontic appliances or impression material as shown in Fig. 4.1



(a) Lower Jaw. (b) Upper Jaw.

Figure 4.1: Example of IOS with impression material on the surface (Patient case 118).

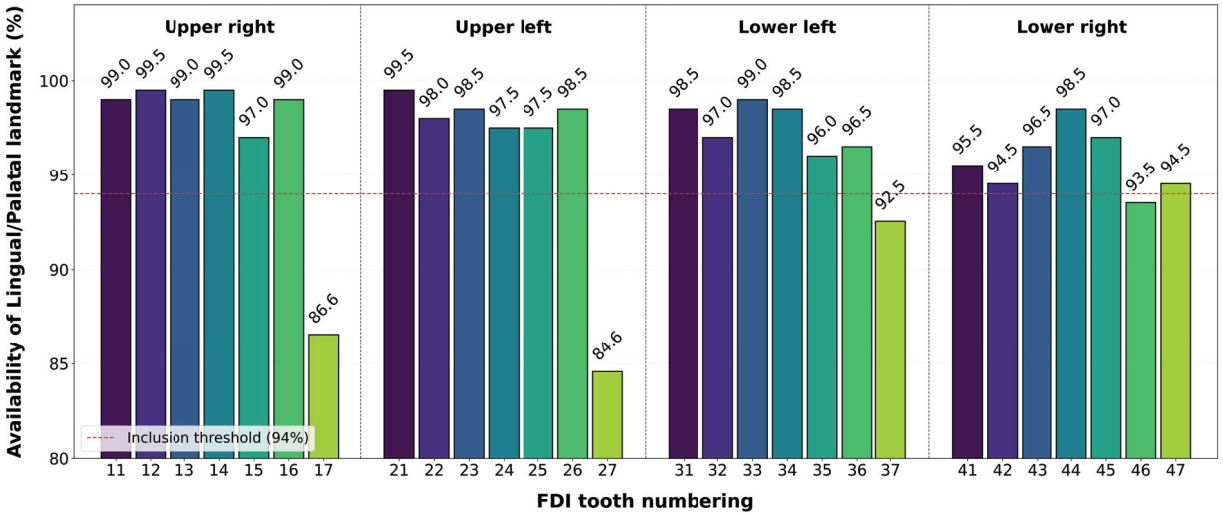


Figure 4.2: Palatal and Lingual Landmarks distribution across the STS 2026 training unlabeled set.

For each tooth position, the proportion of cases containing a valid prediction of the Lingual and Palatal landmark was measured. A minimum coverage criterion of 94% was applied independently to each hemi-arch.

Fig. 4.2 exploits that tooth positions 1–5 satisfied this criterion and were retained for subsequent registration. With four hemi-arches, this produces up to 20 semantic landmark classes. The final landmark set consequently favors availability and consistency over the use of all predictions generated by the underlying model.

4.7 Pseudo-Label Generation

The CBCT landmark detector requires annotations expressed in the coordinate system of the CBCT volumes. Such annotations were not provided for the 300 unlabeled training cases. A pseudo-labeling procedure was therefore used to exploit this additional paired data.

For each unlabeled case, the pretrained IOS landmark pipeline was first applied to the intraoral scan to obtain anatomically identified landmark coordinates. These IOS landmarks were then transferred into the corresponding CBCT coordinate system using an internal IOS–CBCT registration model. The transferred coordinates were used as pseudo-reference positions for training the CBCT landmark detector. Let $\mathbf{p}_k^{\text{IOS}}$ denote an IOS landmark and $\mathbf{T}_{\text{transfer}}$ the transformation used exclusively during pseudo-label generation. The transferred annotation can be written as

$$\begin{bmatrix} \mathbf{q}_k^{\text{CBCT}} \\ 1 \end{bmatrix} = \mathbf{T}_{\text{transfer}} \begin{bmatrix} \mathbf{p}_k^{\text{IOS}} \\ 1 \end{bmatrix}. \tag{4.1}$$

Because pseudo-label quality depends on both landmark localization and the transfer registration, all generated cases were qualitatively inspected. From the original 300 unlabeled cases, 201 were retained after this quality-control step. Also, some examples were excluded due to temporal discrepancy as shown in Fig. 4.3.

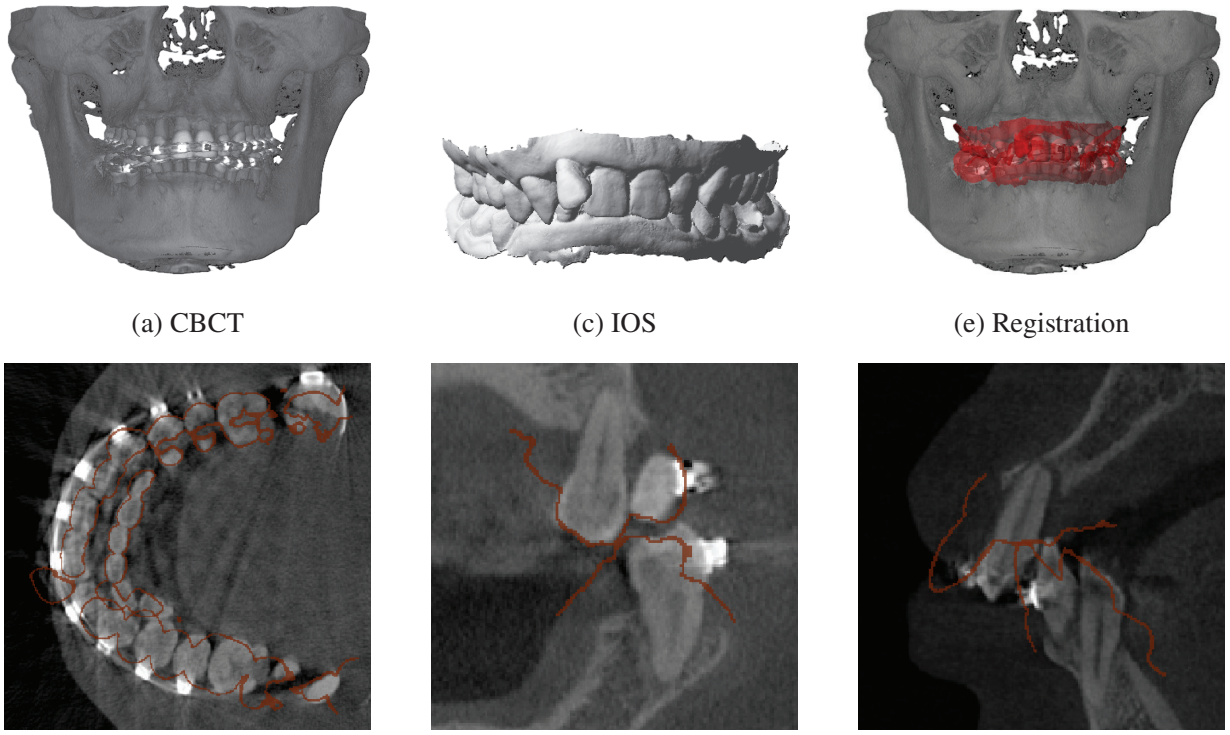
The 30 labeled training cases and the 50 official validation cases were excluded from pseudo-label generation. The internal transfer model was used only to construct the training annotations and is not part of the registration method evaluated in this thesis.

4.8 CBCT Landmark Annotation Representation

The retained pseudo-landmarks are represented as voxel-wise targets for volumetric heatmap prediction. Rather than directly regressing landmark coordinates, a separate output channel is defined for each of the 20 selected anatomical landmark classes.

For a landmark centered at voxel $\mathbf{v}_k$, the positive target region is represented by a $3 \times 3 \times 3$ cube around that location. Equivalently, the neighborhood can be written as

$$\mathcal{N}(\mathbf{v}_k) = \left\{ \mathbf{v} : |v_x - v_{k,x}| \leq 1,\ |v_y - v_{k,y}| \leq 1,\ |v_z - v_{k,z}| \leq 1 \right\}. \tag{4.2}$$



(a) CBCT (c) IOS (e) Registration

(b) Axial View (d) Coronal View (f) Sagittal View

Figure 4.3: Discarded pseudo-labeled Patient case 272 due to temporal discrepancy between the IOS and CBCT acquisitions.

All other voxels of the corresponding output channel are treated as background. This representation preserves landmark identity while allowing the volumetric network to learn local spatial responses around each target position.

4.9 Data Preparation and Augmentation

The two branches require different preparation procedures because of their different data representations. IOS meshes are centered, scaled to a unit sphere for orientation estimation, and sampled using Farthest-Point Sampling. The orientation classifier uses the four discrete transformations defined by the PCA axis-sign ambiguity. The pretrained IOS landmark detector was used without additional challenge-specific training.

For CBCT landmark localization, the 201 retained pseudo-labeled cases were processed using the volumetric training pipeline described in Chapter 5. Standard spatial and intensity augmentations

were used together with foreground oversampling to increase the frequency of landmark-containing regions during training. Left–right mirroring was disabled. Because the output classes encode anatomical identity, a left–right reflection would change the semantic meaning of the landmark classes unless labels were simultaneously remapped.

The exact optimization parameters, patch size, batch size, training schedule, and model size are reported in Chapter 6.

5. Proposed Multimodal Registration Framework

5.1 Introduction

The proposed method addresses IOS–CBCT registration through a coarse-to-fine strategy. The two modalities are first processed independently to obtain semantically corresponding anatomical landmarks. These sparse correspondences are then used to estimate a robust global transformation. Once the major pose difference has been removed, the alignment is refined using local surface geometry.

The separation between the two stages is central to the design. Anatomical landmarks provide modality-independent semantic information for global alignment, while surface registration exploits dense geometry only after the two acquisitions are sufficiently close.

5.2 Framework Overview

An overview of the proposed multimodal registration framework is shown in Fig. 5.1. The input consists of an IOS mesh and the corresponding CBCT volume. The output is a rigid transformation that maps the original IOS into the reference coordinate system of the CBCT.

The processing is divided into two modality-specific branches. In the IOS branch, the scan is first normalized to the canonical orientation expected by the pretrained landmark model. The normalized geometry is then segmented into individual teeth, and an anatomical landmark predictor estimates the selected landmarks on each tooth surface. In parallel, the CBCT branch uses a volumetric landmark detector to localize the corresponding landmarks directly in the image volume.

The predicted landmarks are associated according to their anatomical identity. Low-confidence CBCT predictions are rejected before the remaining correspondences are processed by a confidence-weighted RANSAC–Kabsch procedure. The resulting consensus provides the coarse global transformation and identifies the landmark subset that is geometrically consistent with it.

The inlier CBCT landmarks are then used to define a restricted region of interest. A dental surface is reconstructed from this crop, while the IOS is transformed according to the coarse

alignment. The two surfaces are sampled and downsampled, and point-to-plane ICP is used to estimate a local rigid correction. The final output combines the normalization, landmark-based alignment, and ICP correction.

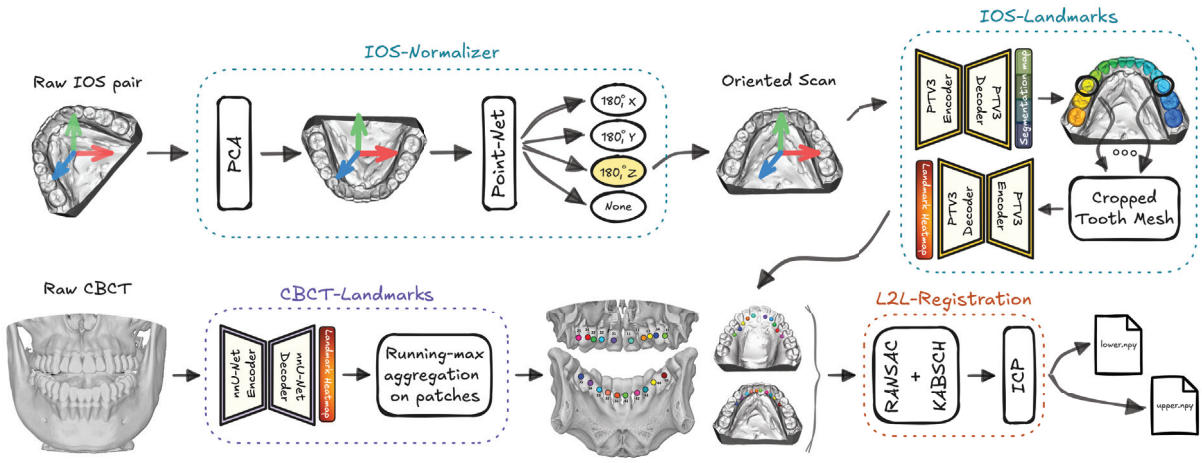


Figure 5.1: Overview of the proposed coarse-to-fine IOS–CBCT registration framework.

5.3 IOS-Normalizer

The *IOS-Normalizer* transforms the input scan into the reference orientation required by the subsequent *IOS-Landmarks* module. The procedure combines deterministic alignment through Principal Component Analysis (PCA) with a learned correction of the remaining orientation ambiguity.

The input geometry is first centered and a scale factor is computed from the maximum distance to the centroid. A unit-sphere version of the centered scan is used for Farthest-Point Sampling (FPS) and PCA.

The PCA basis provides a stable set of principal directions, but the sign of each eigenvector is not uniquely determined. A lightweight point-based classifier therefore selects among four possible corrective transformations: identity, or a 180° rotation around the x or y or z axis. The resulting orientation parameters are retained for subsequent transformation composition.

5.3.1 PCA-Based Alignment

Let the IOS contain N vertices represented by the set

$$\mathbf{X} = \{\mathbf{x}_i\}_{i=1}^{N}, \qquad \mathbf{x}_i \in \mathbb{R}^3. \tag{5.1}$$

The centroid is computed as

$$\mathbf{c} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i. \tag{5.2}$$

A scale factor is obtained from the maximum centered distance,

$$s = \max_i \|\mathbf{x}_i - \mathbf{c}\|_2. \tag{5.3}$$

The normalized points used for orientation estimation are

$$\hat{\mathbf{x}}_i = \frac{\mathbf{x}_i - \mathbf{c}}{s} \tag{5.4}$$

Thus, the orientation estimation operates on a centered point cloud contained in the unit sphere. A fixed number M of representative points is then selected with Farthest-Point Sampling. The mean of the sampled points is recomputed as

$$\bar{\mathbf{x}} = \frac{1}{M}\sum_{i=1}^{M}\hat{\mathbf{x}}_i. \tag{5.5}$$

The covariance matrix is computed using the unbiased normalization

$$\mathbf{C} = \frac{1}{M-1}\sum_{i=1}^{M}(\hat{\mathbf{x}}_i - \bar{\mathbf{x}})(\hat{\mathbf{x}}_i - \bar{\mathbf{x}})^T. \tag{5.6}$$

Because $\mathbf{C}$ is symmetric, its eigenvectors form an orthonormal basis. The eigenvectors are ordered by decreasing eigenvalue and assembled into

$$\mathbf{B} = \begin{bmatrix}\mathbf{b}_1 & \mathbf{b}_2 & \mathbf{b}_3\end{bmatrix}. \tag{5.7}$$

The determinant of the basis is checked to ensure that it defines a proper rotation. If $\det(\mathbf{B}) < 0$, the sign of the last eigenvector is reversed so that $\det(\mathbf{B}) = 1$. PCA determines the principal directions but does not uniquely determine their signs. An eigenvector and its negative describe the same principal axis. Consequently, equivalent scans can remain separated by discrete 180° rotations after PCA. This ambiguity is resolved by the learned orientation-correction module.

5.3.2 Learned Orientation Correction

The orientation ambiguity is formulated as a four-class classification problem rather than as regression of an arbitrary three-dimensional rotation. The set of possible corrective rotations is

$$\mathcal{R} = \{\mathbf{I}, \mathbf{R}_x(\pi), \mathbf{R}_y(\pi), \mathbf{R}_z(\pi)\}. \tag{5.8}$$

The classifier follows a PointNet-style architecture, as shown in Fig. 5.2. A sequence of shared 1×1 convolutional layers processes the sampled three-dimensional points. Global maximum and average pooling produce complementary descriptors, which are concatenated and passed through a fully connected classification head that produces four logits. The corrective rotation associated with the highest logit is selected as $\mathbf{R}$. The combined PCA and learned orientation matrix is therefore

$$\mathbf{M} = \mathbf{BR}. \tag{5.9}$$

For the implementation used in this work, unit-sphere normalization is used internally for FPS, PCA, and orientation classification. The resulting orientation is then applied to the centered IOS geometry without changing its physical scale. When points are stored as row vectors, the canonical geometry can consequently be written as

$$\mathbf{X}' = (\mathbf{X} - \mathbf{c})\mathbf{BR}. \tag{5.10}$$

The centroid, scale used for orientation estimation, and orientation matrix are retained for transformation bookkeeping. When converting the implementation to the column-vector convention used by rigid registration, the corresponding rotation is represented consistently with the adopted matrix convention. The normalized geometry is then passed to the IOS landmark-detection pipeline.

5.4 IOS-Landmarks

After normalization, the IOS is processed by the *IOS-Landmarks* module. The module contains two stages: tooth-level semantic segmentation and landmark prediction on the segmented tooth surfaces. Both stages use Point Transformer V3, allowing the networks to operate directly on irregular three-dimensional point sets [20].

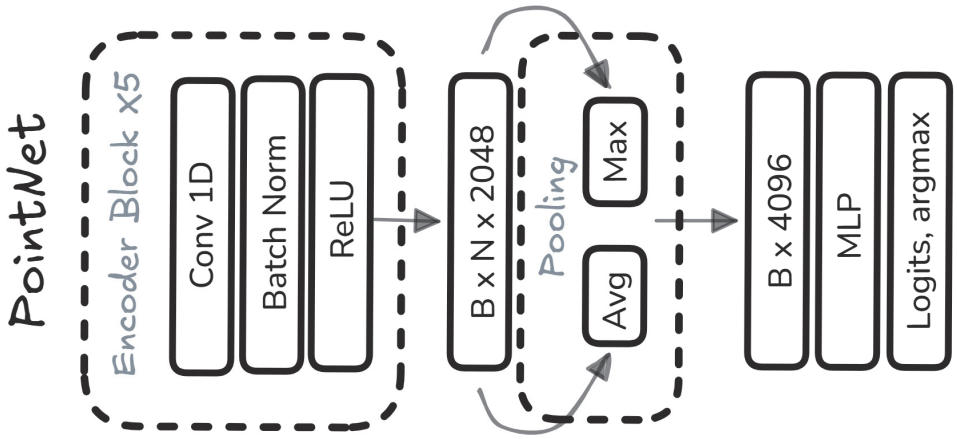


Figure 5.2: PointNet-style architecture.

The first stage assigns each input point to a dental structure, while the second stage predicts the position of the selected anatomical landmark for each detected tooth. The combination provides both semantic identity and three-dimensional landmark position.

5.4.1 Tooth Segmentation and Identification

Each normalized IOS vertex is assigned to one of 17 semantic classes, as shown in Fig. 5.3. Sixteen classes correspond to the possible teeth of the input arch and one class represents gingiva and other non-dental regions. Let the normalized IOS be

$$\mathbf{X}' = \{\mathbf{x}'_i\}_{i=1}^{N}. \tag{5.11}$$

The segmentation network predicts a class

$$\hat{y}_i \in \{0, \ldots, 16\} \tag{5.12}$$

for each vertex. The semantic classes are mapped to FDI tooth identifiers using the known identity of the input arch. This allows a tooth detected in the IOS to be associated with the anatomically corresponding class in the CBCT landmark detector.

The segmentation step therefore provides two functions. It isolates the dental structures from gingiva and noise, and it attaches anatomical identity to the extracted tooth regions.

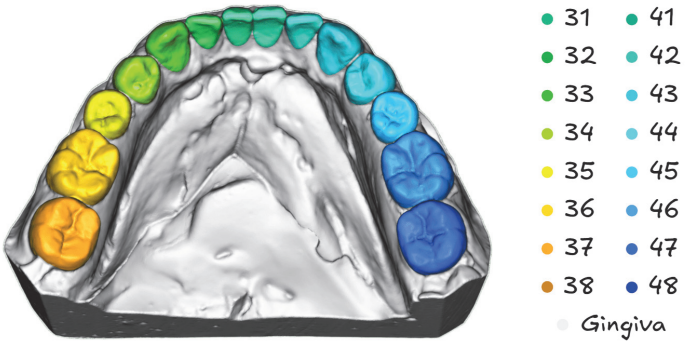


Figure 5.3: Segmentation Example of the Lower Jaw for Patient case 027.

5.4.2 Landmark Prediction and Selection

The vertices assigned to each tooth are processed independently by a second Point Transformer V3 model. Rather than directly regressing a coordinate, the network predicts a vertex-wise probability heatmap. For a tooth containing N_t vertices, the heatmap is written as

$$H = \{h_i\}_{i=1}^{N_t}, \qquad h_i \in [0, 1]. \tag{5.13}$$

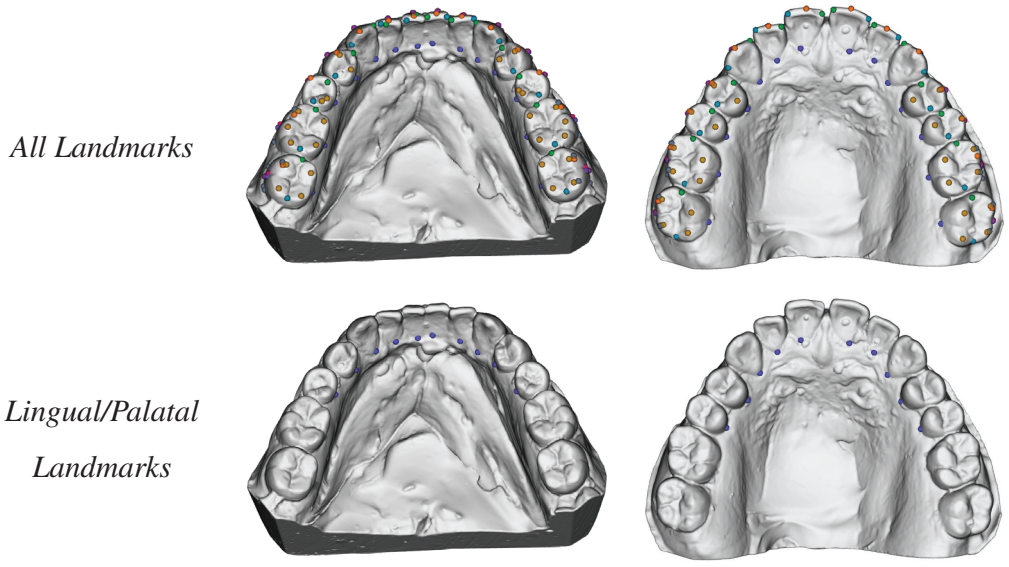
High-response vertices identify the region associated with the target landmark. Thresholding removes weak responses and K-means clustering groups the retained vertices into spatially coherent regions. The resulting cluster is used to determine the landmark coordinate.

The underlying landmark model can predict multiple categories, as shown in Fig. 2.5, but only the Lingual and Palatal landmarks are retained by the proposed registration pipeline as explained in Section 4.6. Fig. 5.4 shows the retained predictions for each case processed by the pipeline.

The resulting IOS output is therefore a set of anatomically identified three-dimensional points that can be associated with the corresponding CBCT landmarks.

5.5 CBCT-Landmarks

The CBCT branch operates directly on the volumetric acquisition using a dedicated landmark detector based on nnLandmark [7]. The network predicts one output heatmap for each of the 20 selected anatomical landmarks.



(a) Lower Jaw. (b) Upper Jaw.

Figure 5.4: Predicted Landmarks for Patient case 027.

5.5.1 Heatmap-Based Landmark Localization

For landmark class k, let the predicted heatmap be

$$H_k(\mathbf{v}), \qquad \mathbf{v} = (v_x, v_y, v_z), \tag{5.14}$$

where $\mathbf{v}$ is a voxel coordinate. During training, the target annotation is represented by a $3 \times 3 \times 3$ positive neighborhood around the reference voxel, as described in Section 4. At inference time, the CBCT volume is processed with a sliding-window strategy. For each output channel, the predicted landmark is the voxel associated with the maximum response,

$$\mathbf{v}_k^* = \arg\max_{\mathbf{v}} H_k(\mathbf{v}). \tag{5.15}$$

The predicted voxel coordinate is then converted to physical coordinates using the spatial metadata of the CBCT volume.

5.5.2 Landmark Confidence and Inference

The maximum heatmap response is also retained as a confidence value

$$c_k = H_k(\mathbf{v}_k^*). \tag{5.16}$$

CBCT landmarks with confidence below the minimum threshold specified in the experimental configuration are discarded before rigid registration.

The inference implementation was optimized to avoid reconstructing complete heatmap volumes when only the maxima are required. Instead, the maximum response and associated coordinate are tracked directly across sliding-window patches. Additional optimizations include batched patch processing, persistent GPU model loading, and GPU-based resampling. These modifications reduce unnecessary memory transfers and intermediate allocations while preserving the predicted landmark locations needed by the registration stage.

The final output of the CBCT branch is a set of anatomically identified physical coordinates together with confidence values.

5.6 Landmark-Based Coarse Registration

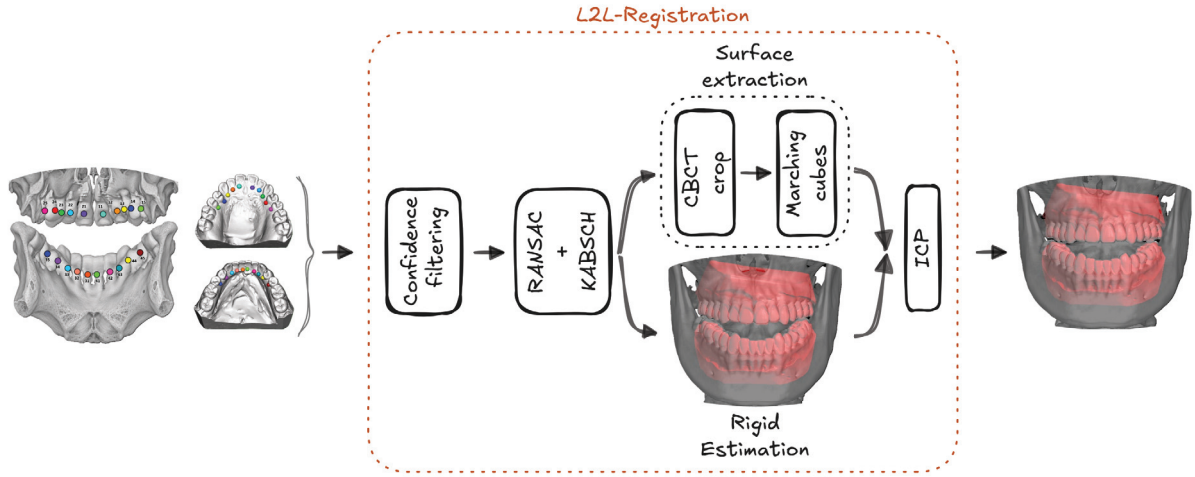


Figure 5.5: *L2L-Registration* module in detail.

The first stage of the Landmark-to-Landmark Registration, shown in Fig. 5.5, estimates the global rigid relative pose between IOS and CBCT. Because the landmarks are detected independently in the two modalities, correspondence can be established from anatomical identity rather than from initial spatial proximity.

The main adversity is that the predicted landmarks are not guaranteed to be accurate. The proposed solution combines confidence-weighted Kabsch estimation with a robust RANSAC pro-

cedure. The output is both a coarse rigid transformation and a geometrically consistent inlier set used by the subsequent surface-refinement stage.

5.6.1 Cross-Modal Landmark Correspondences

Let the IOS landmarks be

$$\mathcal{P} = \{\mathbf{p}_i\}_{i=1}^{K}, \qquad \mathbf{p}_i \in \mathbb{R}^3, \tag{5.17}$$

and the corresponding CBCT landmarks be

$$\mathcal{Q} = \{\mathbf{q}_i\}_{i=1}^{K}, \qquad \mathbf{q}_i \in \mathbb{R}^3. \tag{5.18}$$

A pair $(\mathbf{p}_i, \mathbf{q}_i)$ represents the same anatomical landmark. Correspondences are established from tooth identity and landmark type, not by nearest-neighbor matching.

The CBCT predictions are first converted from voxel indices to physical coordinates. Let c_i^{IOS} and c_i^{CBCT} denote the confidence values associated with the two predictions. When no IOS confidence is available, a unit value is used. CBCT predictions below the configured confidence threshold are rejected. The valid correspondence set is therefore

$$C = \left\{(\mathbf{p}_i, \mathbf{q}_i, c_i^{\mathrm{IOS}}, c_i^{\mathrm{CBCT}})\right\}_{i=1}^{K}. \tag{5.19}$$

The number of valid correspondences varies between cases, motivating the adaptive procedure described below.

5.6.2 Confidence-Weighted Kabsch Estimation

For corresponding points, the rigid transformation is defined by

$$\mathbf{q}_i = \mathbf{R}\mathbf{p}_i + \mathbf{t}, \tag{5.20}$$

where $\mathbf{R} \in \mathrm{SO}(3)$ and $\mathbf{t} \in \mathbb{R}^3$. The rotation satisfies

$$\mathbf{R}^T\mathbf{R} = \mathbf{I}, \qquad \det(\mathbf{R}) = 1. \tag{5.21}$$

The confidence associated with each correspondence is incorporated through

$$w_i = \sqrt{c_i^{\mathrm{IOS}} c_i^{\mathrm{CBCT}}}. \tag{5.22}$$

The weights are normalized as

$$\tilde{w}_i = \frac{w_i}{\sum_{j=1}^{K} w_j}. \tag{5.23}$$

The weighted centroids are then

$$\bar{\mathbf{p}}_w = \sum_{i=1}^{K} \tilde{w}_i \mathbf{p}_i, \qquad \bar{\mathbf{q}}_w = \sum_{i=1}^{K} \tilde{w}_i \mathbf{q}_i. \tag{5.24}$$

Centered points are defined as

$$\tilde{\mathbf{p}}_i = \mathbf{p}_i - \bar{\mathbf{p}}_w, \qquad \tilde{\mathbf{q}}_i = \mathbf{q}_i - \bar{\mathbf{q}}_w. \tag{5.25}$$

The weighted cross-covariance matrix is

$$\mathbf{H}_w = \sum_{i=1}^{K} \tilde{w}_i \tilde{\mathbf{p}}_i \tilde{\mathbf{q}}_i^T. \tag{5.26}$$

Its singular value decomposition is

$$\mathbf{H}_w = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T. \tag{5.27}$$

The optimal proper rotation is obtained with determinant correction,

$$\mathbf{R} = \mathbf{V}\,\mathrm{diag}\left(1, 1, \det(\mathbf{V}\mathbf{U}^T)\right)\mathbf{U}^T. \tag{5.28}$$

The translation is then

$$\mathbf{t} = \bar{\mathbf{q}}_w - \mathbf{R}\bar{\mathbf{p}}_w. \tag{5.29}$$

The weighted Kabsch estimator is optimal for the supplied correspondence set, but it remains sensitive to incorrectly localized points. The estimator is therefore embedded within the robust consensus procedure described next.

5.6.3 Confidence-Weighted RANSAC-Kabsch

RANSAC is used to identify a geometrically consistent subset of landmark correspondences. Each hypothesis is generated from three landmark pairs. Candidate triplets are rejected when the points in either modality are approximately collinear, since such a configuration does not sufficiently constrain a three-dimensional rigid transformation.

For each triplet, the spatial distribution is evaluated through singular value decomposition, and the candidate is accepted only if the second singular value satisfies the implementation threshold independently for IOS and CBCT.

If the total number of triplets does not exceed the maximum number of trials, all combinations are evaluated. Otherwise, a fixed number of triplets is sampled without replacement using a deterministic sampling configuration. For a hypothesis h, the weighted Kabsch estimator provides $(\mathbf{R}_h, \mathbf{t}_h)$. The residual of correspondence i is

$$e_i^{(h)} = \|\mathbf{R}_h \mathbf{p}_i + \mathbf{t}_h - \mathbf{q}_i\|_2 \,. \tag{5.30}$$

For threshold τ, a correspondence is classified as an inlier when

$$e_i^{(h)} \leq \tau. \tag{5.31}$$

The corresponding consensus set is

$$\mathcal{I}_h = \{i : e_i^{(h)} \leq \tau\}. \tag{5.32}$$

Candidate hypotheses are ranked lexicographically. First, the number of inliers is maximized. If this is equal, the sum of confidence weights of the inliers is used as the second criterion. If this is also equal, the hypothesis with the smallest median inlier residual is preferred.

5.6.4 Adaptive Consensus and Transformation Re-estimation

The available number and quality of landmarks vary across cases. A single fixed residual threshold and consensus size would therefore be unnecessarily restrictive in difficult cases. The registration procedure starts from an initial threshold τ_0 and progressively increases it by a fixed increment

$$\tau_{j+1} = \tau_j + \Delta\tau \tag{5.33}$$

until either a valid consensus is found or a maximum threshold $\tau_{\max}$ is reached. Let K be the number of valid correspondences and $N_{\min}$ the configured minimum consensus. The effective requirement is

$$N_{\text{req}} = \min\left(K, \max(3, N_{\min})\right) . \tag{5.34}$$

If the complete threshold sequence fails to produce a valid solution, a second fallback reduces the required consensus by one landmark while maintaining a minimum of three correspondences. The threshold sequence is then evaluated again.

If exactly three valid correspondences are available, conventional RANSAC sampling is skipped. The triplet is first checked for non-collinearity, the confidence-weighted Kabsch transformation is estimated directly, and the solution is accepted only when all three residuals satisfy the current threshold.

Once a valid hypothesis is selected, its original three-point transformation is not used as the final coarse solution. Instead, all correspondences in the selected consensus are used for weighted Kabsch re-estimation,

$$(\mathbf{R}^*, \mathbf{t}^*) = \arg\min_{\mathbf{R},\mathbf{t}} \sum_{i \in \mathcal{I}} \tilde{w}_i \, \|\mathbf{R}\mathbf{p}_i + \mathbf{t} - \mathbf{q}_i\|_2^2 \,. \tag{5.35}$$

The residuals are recomputed and the inlier mask is updated. This process is repeated until the inlier set remains unchanged between two consecutive iterations. The implementation limits the refinement to ten iterations. A solution is rejected when the consensus falls below the required minimum or fails to stabilize. The stable transformation is denoted by

$$\mathbf{T}_{\text{rigid}} = \begin{bmatrix} \mathbf{R}_{\text{rigid}} & \mathbf{t}_{\text{rigid}} \\ \mathbf{0}^T & 1 \end{bmatrix}. \tag{5.36}$$

The corresponding inlier set is retained for the surface-refinement stage.

5.7 Surface-Based Fine Registration

The coarse registration stage provides a global alignment but uses only a sparse subset of the available anatomical information. Indeed, the final alignment is therefore refined using dense surface geometry.

Direct reconstruction of the complete CBCT surface is not desirable because the volume contains roots, bone, and other structures that do not have corresponding IOS surfaces. The robust landmark consensus is consequently reused to define a restricted CBCT region before surface extraction.

5.7.1 Landmark-Guided CBCT Cropping

Let the final inlier CBCT landmarks be

$$Q_{\text{in}} = \{\mathbf{q}_i\}_{i \in \mathcal{I}}. \tag{5.37}$$

The minimum and maximum coordinates are computed independently along each axis,

$$\mathbf{q}_{\min} = \min_{\mathbf{q}_i \in Q_{\text{in}}} \mathbf{q}_i, \qquad \mathbf{q}_{\max} = \max_{\mathbf{q}_i \in Q_{\text{in}}} \mathbf{q}_i. \tag{5.38}$$

The extent of the landmark configuration is

$$\mathbf{E} = \mathbf{q}_{\max} - \mathbf{q}_{\min}. \tag{5.39}$$

A margin is added element-wise according to

$$\mathbf{m} = \max(m_{\min}, \alpha \mathbf{E}), \tag{5.40}$$

and the physical crop bounds become

$$\mathbf{b}_{\min} = \mathbf{q}_{\min} - \mathbf{m}, \qquad \mathbf{b}_{\max} = \mathbf{q}_{\max} + \mathbf{m}. \tag{5.41}$$

The resulting physical bounds are converted to voxel indices using the CBCT spatial metadata and clipped to the valid image domain.

5.7.2 Surface Reconstruction and Preprocessing

The cropped CBCT region is converted into a triangular surface using the Marching Cubes algorithm.

If $\mathbf{v}$ is a voxel-coordinate vertex, $\mathbf{s}$ the voxel spacing, $\mathbf{o}$ the image origin, and $\mathbf{D}$ the direction matrix, the corresponding physical coordinate is

$$\mathbf{x}_{\text{phys}} = \mathbf{o} + \mathbf{D}(\mathbf{v} \odot \mathbf{s}), \tag{5.42}$$

where $\odot$ denotes element-wise multiplication.

The coarsely aligned IOS and the reconstructed CBCT surface are then converted into point clouds. Surface sampling is followed by voxel downsampling to reduce computational cost and provide a more homogeneous point density. Surface normals are estimated from local neighborhoods and are used by the point-to-plane ICP objective.

5.7.3 Point-to-Plane ICP

ICP is used only to estimate a local correction. Let $\mathcal{P}_s = \{\mathbf{p}_i\}$ denote the coarsely aligned IOS point cloud and $\mathcal{P}_t = \{\mathbf{q}_j\}$ the point cloud reconstructed from the cropped CBCT region. For a point-to-plane objective, a source point $\mathbf{p}_i$ is paired with a target point $\mathbf{q}_i$ and target normal $\mathbf{n}_i$. The incremental transformation minimizes the displacement along the target normal,

$$(\mathbf{R}^*_{\text{ICP}}, \mathbf{t}^*_{\text{ICP}}) = \arg\min_{\mathbf{R},\mathbf{t}} \sum_i \left[\mathbf{n}_i^T(\mathbf{R}\mathbf{p}_i + \mathbf{t} - \mathbf{q}_i)\right]^2. \tag{5.43}$$

Only correspondences below the maximum correspondence distance are considered. The correspondence set and incremental transformation are updated iteratively until convergence or the iteration limit is reached. Because the coarse transformation has already been applied to the IOS, the ICP stage begins from the identity correction. The resulting homogeneous matrix is

$$\mathbf{T}_{\text{ICP}} = \begin{bmatrix} \mathbf{R}_{\text{ICP}} & \mathbf{t}_{\text{ICP}} \\ \mathbf{0}^T & 1 \end{bmatrix}. \tag{5.44}$$

5.7.4 Transformation Composition

The complete registration maps the original IOS coordinate system into the CBCT coordinate system. Let $\mathbf{T}_{\text{oriented}}$ denote the rigid orientation transform corresponding to the stored IOS centering and orientation parameters, expressed consistently in the column-vector convention used for registration, and let $\mathbf{T}_{\text{rigid}}$ and $\mathbf{T}_{\text{ICP}}$ be the rigid and ICP transformations defined above. The coarse transformation in the CBCT physical coordinate system is

$$\mathbf{T}^{\text{LPS}}_{\text{coarse}} = \mathbf{T}_{\text{rigid}}\mathbf{T}_{\text{oriented}}. \tag{5.45}$$

Since the ICP correction is estimated after the coarse transformation has been applied, the refined transformation is

$$\mathbf{T}^{\text{LPS}}_{\text{final}} = \mathbf{T}_{\text{ICP}}\mathbf{T}_{\text{rigid}}\mathbf{T}_{\text{oriented}}. \tag{5.46}$$

The physical CBCT coordinates used internally follow the LPS convention. For export, the final transformation is converted to the RAS convention through

$$\mathbf{T}_{\mathrm{LPS}\to\mathrm{RAS}} = \begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \tag{5.47}$$

and the exported transformation is

$$\mathbf{T}_{\mathrm{final}}^{\mathrm{RAS}} = \mathbf{T}_{\mathrm{LPS}\to\mathrm{RAS}}\mathbf{T}_{\mathrm{ICP}}\mathbf{T}_{\mathrm{rigid}}\mathbf{T}_{\mathrm{oriented}}. \tag{5.48}$$

This matrix is the final output of the proposed registration framework and maps the original IOS geometry into the target coordinate system required for evaluation.

6. Experiments and Results

6.1 Introduction

This chapter evaluates the proposed multimodal IOS–CBCT registration framework and analyzes the contribution of its principal registration components. The experiments were conducted within the MICCAI STS 2026 Challenge setting, using the 30 labeled training cases for controlled evaluation and ablation studies and the official 50-case validation set for final challenge evaluation.

The chapter first describes the experimental environment and training protocols, followed by the parameters of the geometric registration stages. The evaluation metrics are then introduced. An ablation study isolates the contribution of robust landmark selection and ICP refinement, after which quantitative and qualitative results are presented. The chapter concludes with the analysis of the systematic transformation-labeling inconsistency identified in the challenge data and a description of the computational optimizations used in the implementation.

6.2 Experimental Setup

The complete experimental configuration includes the fine-tuned *IOS-Normalizer*, the pretrained *IOS-Landmark*, the trained *CBCT-Landmarks*, and the *L2L-Registration* stage described in Chapter 5. All experiments were performed on a workstation equipped as reported in Tab. 6.1.

6.2.1 Training Protocols

The three learning-based components were trained or initialized differently. The *IOS-Normalizer* was fine-tuned from a pretrained checkpoint, the *IOS-Landmark* models were used with publicly available pretrained weights, and the *CBCT-Landmark* detector was trained from scratch using the pseudo-labeled strategy described in Chapter 4.

IOS-Normalizer. The IOS-Normalizer was initialized from a checkpoint pretrained on Teeth3DS, containing 1,800 consistently oriented intraoral scans. To adapt the classifier to the challenge data,

Table 6.1: Hardware and software environment used for the experiments.

Component	**Configuration**
Operating system	Ubuntu 22.04.5 LTS
Processor	Intel Xeon Silver 4216 @ 2.10 GHz
System memory	128 GB
Graphics processor	NVIDIA RTX A5000, 24 GB
CUDA	12.6
Programming language	Python 3.10
Deep learning framework	PyTorch 2.8
torchvision	0.23.0

60 IOS cases were selected from the unlabeled training subset and manually oriented. Fifty cases were used for fine-tuning and ten for validation. Each scan was sampled to 16,384 points using Farthest-Point Sampling. During training, an input was either kept in canonical orientation or transformed by a 180° rotation around one of the three cartesian axes. The network was trained to classify the corresponding orientation correction.

IOS-Landmarks. The IOS segmentation and landmark-prediction models described in Section 5 were used with publicly available pretrained weights and were not additionally fine-tuned on the STS 2026 dataset.

CBCT-Landmarks. The CBCT landmark detector was trained from scratch using the 201 pseudo-labeled cases retained after qualitative inspection. Training was restricted to the 20 Lingual/Palatal landmark classes associated with tooth positions 1–5 across the four hemi-arches. Standard spatial and intensity augmentation and foreground oversampling were used, while left–right mirroring was disabled to preserve anatomical class identity. Five-fold cross-validation was used during model development, and the checkpoint associated with the best validation performance was retained.

Table 6.2: Training protocols for IOS-Normalizer and CBCT-Landmarks modules.

Parameter	**IOS-Normalizer**	**CBCT-Landmarks**
Batch size	1	2
Input patch size	–	128^3 voxels
Epochs	3	280
Optimizer	AdamW	Nesterov SGD
Weight decay / momentum	10^{-4}	0.99
Initial learning rate	10^{-4}	10^{-2}
Learning-rate schedule	None	Polynomial ($p = 0.9$)
Training time	4.5 min	87 h (5-fold CV)
Loss	Cross-entropy	Top-20% BCE (logits)
Model parameters	13 M	102 M
FLOPs/step	91.3 GFLOPs	8.6 TFLOPs

6.2.2 Registration Parameters

The parameters of the final geometric registration configuration are summarized in Tab. 6.3. The landmark confidence threshold is applied before robust estimation. Candidate RANSAC hypotheses use three non-collinear correspondences, and the inlier threshold is progressively relaxed when no valid consensus is found.

6.3 Evaluation Metrics

Registration accuracy is evaluated using Mean Translation Error (MTE) and Mean Rotation Error (MRE), which separately measure translational and rotational differences between the predicted and reference transformations. Let the predicted transformation for case i be

$$\mathbf{T}_i = \begin{bmatrix} \mathbf{R}_i & \mathbf{t}_i \\ \mathbf{0}^T & 1 \end{bmatrix} \tag{6.1}$$

Table 6.3: L2L-Registration Parameters.

Stage	Parameter	Value
Landmark	Minimum landmark confidence	0.25
	Maximum RANSAC hypotheses	2000
	Landmarks per hypothesis	3 non-collinear
	Initial inlier distance	3.5 mm
	Maximum inlier distance	5.5 mm
	Initial minimum consensus	4 landmarks
	Second consensus fallback	Enabled
	Minimum fallback consensus	3 landmarks
Surface	Crop margin	$\max(15\,\text{mm}, 0.25E)$
	Marching Cubes threshold	1500
	Maximum sampled points	12,000 per surface
	Voxel size	1.0 mm
	Normal estimation radius	3.0 mm
ICP	Maximum correspondence distance	1.0 mm
	Maximum iterations	40
	Relative fitness tolerance	10^{-5}
	Relative RMSE tolerance	10^{-5}

and the reference transformation be

$$\mathbf{T}_i^{\mathrm{GT}} = \begin{bmatrix} \mathbf{R}_i^{\mathrm{GT}} & \mathbf{t}_i^{\mathrm{GT}} \\ \mathbf{0}^T & 1 \end{bmatrix}. \tag{6.2}$$

6.3.1 Mean Translation Error

For an individual case, the translation error is the Euclidean distance between predicted and reference translation vectors,

$$e_t^{(i)} = \left\| \mathbf{t}_i - \mathbf{t}_i^{\mathrm{GT}} \right\|_2 . \tag{6.3}$$

For a dataset containing N cases,

$$\mathrm{MTE} = \frac{1}{N} \sum_{i=1}^{N} e_t^{(i)}. \tag{6.4}$$

MTE is reported in millimeters.

6.3.2 Mean Rotation Error

The rotational discrepancy is computed from the relative rotation

$$\mathbf{R}_{\Delta}^{(i)} = \left(\mathbf{R}_i^{\mathrm{GT}}\right)^T \mathbf{R}_i. \tag{6.5}$$

The corresponding geodesic angular error is

$$e_r^{(i)} = \arccos\left(\mathrm{clip}\left(\frac{\mathrm{tr}(\mathbf{R}_{\Delta}^{(i)}) - 1}{2}, -1, 1\right)\right). \tag{6.6}$$

The Mean Rotation Error is

$$\mathrm{MRE} = \frac{1}{N} \sum_{i=1}^{N} e_r^{(i)}, \tag{6.7}$$

with the final value expressed in degrees.

6.4 Ablation Study

An ablation study was conducted on the 30 labeled training cases. The reference transformations were used only for evaluation and were not used to train any component of the proposed framework. Tab. 6.4 shows the ablation results and also the score on the hidden validation and test sets.

The results show a reduction in both metrics when robust correspondence selection is introduced. Landmark registration alone produced an MTE of 13.759 mm and an MRE of 3.993°, whereas adding RANSAC reduced the errors to 8.708 mm and 2.619°. Applying ICP without RANSAC resulted in an MTE of 7.959 mm and an MRE of 1.997°. The complete pipeline, combining robust landmark registration and ICP, reached 1.813 mm and 0.891° on the labeled subset.

These results indicate that the two components address different sources of error and improve the overall registration. RANSAC improves the reliability of the global transformation by removing geometrically inconsistent correspondences, while ICP exploits dense surface information to refine the residual alignment.

Table 6.4: Ablation study of the registration pipeline and official evaluation results.

Set	RANSAC	ICP	Flip ID > 269	MTE (mm) ↓	MRE (°) ↓
Train Labeled	✗	✗	✓	13.759	3.993
	✓	✗	✓	8.708	2.619
	✗	✓	✓	7.959	1.997
	✓	✓	✓	**1.813**	**0.891**
Validation	✓	✓	✗	10.037	22.809
	✓	✓	✓	**4.988**	**1.421**
Test corrected	✓	✓	✗	6.752	1.976

6.5 Systematic Transformation Labeling Error

During analysis of the challenge reference transformations, a systematic inconsistency was identified in a subset of the labeled data. Inspection of the 3×3 linear component of the provided 4×4 transformations showed that all labeled training cases with identifiers greater than 269 had

$$\det(\mathbf{A}) = -1, \tag{6.8}$$

indicating the presence of a reflection rather than a proper pure rigid rotation. Visual inspection of the transformed IOS meshes confirmed that this corresponded to a reflection across the YZ plane. The issue became relevant during validation. An initial submission in which all predictions were restricted to proper rotations produced an MTE of 10.037 mm and an MRE of 22.809°, as shown in Tab. 6.4. Predictions associated with identifiers greater than 269 were subsequently composed with the corresponding YZ-plane reflection, based on the same identifier-dependent convention observed in the training labeled data. The resulting evaluation was 4.988 mm MTE and 1.421° MRE. The presence and interpretation of the inconsistency were subsequently confirmed by the challenge organizers. The correction therefore concerns the reference-transformation convention used for evaluation rather than the registration algorithm itself. The proposed framework estimates proper rigid transformations and does not intrinsically introduce reflections.

6.6 Quantitative Results on Validation Set

The complete pipeline was evaluated on the official validation and test sets.

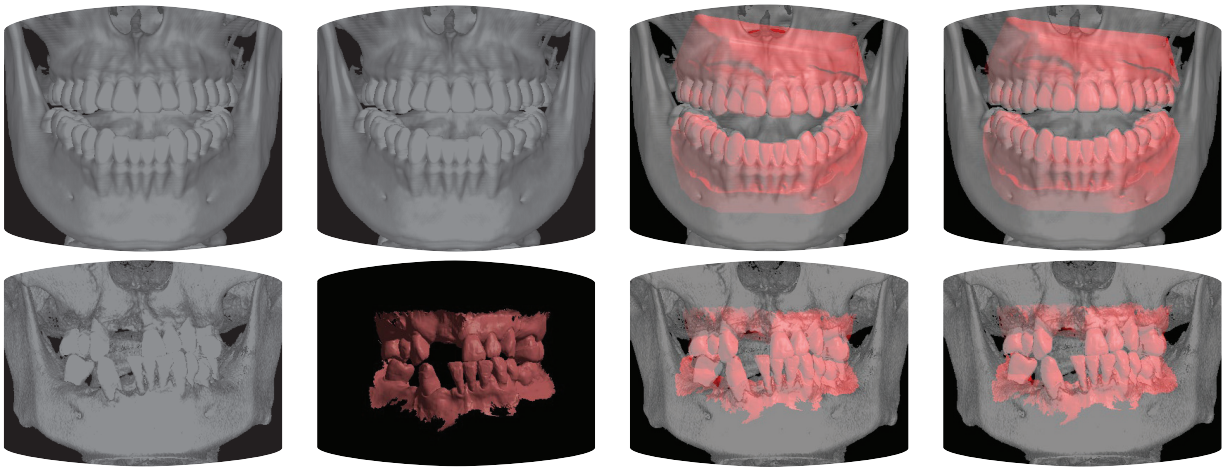
6.6.1 Validation Set

Since the reference transformations of the 50 validation cases were hidden, the predicted transformations were submitted to the challenge evaluation system. As reported in Tab. 6.4, the complete method achieved an MTE of 4.988 mm and an MRE of 1.421° on the challenge validation set, following confirmation by the challenge organizers of the systematic transformation-labeling error discussed in Section 6.5.

The validation translation error is higher than that observed on the labeled training subset, while the rotation error remains substantially lower than the translation error. The results therefore provide a quantitative assessment of the complete pipeline on cases whose reference transformations were not available during development.

6.6.2 Test Set

The final evaluation was performed by the challenge organizers on a hidden test set of 99 cases following submission of the Docker container. The evaluation metrics achieved were an MTE of 10114.828 mm and an MRE of 36.122° with 38 failed pairs. Since the scores were very different from the performance reported in Tab. 6.4, we requested clarification from the organizers. According to the confirmation provided by the challenge organizers, the unusual scores were caused by the fact that the ID-dependent inconsistency identified in the validation set was not correctly resolved in the test set. The corrected evaluation results, shown in Tab. 6.4, yielded an MTE of 6.752 mm and an MRE of 1.976°, with 2 failed pairs after excluding the 18 cases affected by the remaining reflection issue.



(a) CBCT (b) IOS (c) Kabsch (d) ICP

Figure 6.1: Qualitative results, one patient per row (P027 and P214 from the top).

6.7 Qualitative Results

Qualitative inspection was used to analyze the behavior of the method under different landmark configurations. Registrations generally appeared more consistent when several reliable landmark correspondences were available and distributed across the dental arch. Cases relying on only three or four correspondences appeared more sensitive to individual localization errors. The spatial distribution of the landmarks was also relevant: points covering a broad portion of the arch provide stronger geometric constraints than points concentrated within a small region. Approximately collinear configurations are explicitly avoided during RANSAC sampling. In cases where the landmark-based initialization retained a residual local mismatch, point-to-plane ICP often improved the surface superimposition by exploiting the reconstructed dental geometry. Fig. 6.1 provides registration results for patient cases 027 and 214 from the Validation Set.

6.8 Computational Performance

Computational efficiency was considered during development, particularly for CBCT landmark inference. The original nnLandmark inference procedure was modified to track the maximum heatmap response and its corresponding coordinate directly across sliding-window patches instead of reconstructing and transferring complete heatmap volumes.

Further optimizations included batched inference patches, persistent GPU model loading, and GPU-based volume resampling. These modifications reduce memory transfers and unnecessary intermediate allocations while preserving the information required by subsequent registration stages.

The geometric stages also limit computation by reconstructing only the landmark-guided CBCT crop and by downsampling both surfaces before ICP. Exact end-to-end runtime was not used as a primary evaluation metric in the experiments reported here; quantitative assessment focuses on registration accuracy through MTE and MRE.

7. Discussion

7.1 Discussion of Main Findings

The results presented in Chapter 6 show that the proposed coarse-to-fine strategy benefits from combining semantic anatomical information with local surface geometry. The two registration stages address different aspects of the IOS–CBCT problem: landmark-based alignment provides a global estimate of the relative pose, while surface refinement improves the local agreement after the two modalities have been brought into approximate correspondence. The ablation study reported in Tab. 6.4 supports this interpretation. RANSAC improves the landmark-based transformation by reducing the influence of inconsistent correspondences, while ICP contributes an additional refinement based on dense geometric information. The combined configuration provides a substantially lower error on the labeled training subset than either component used independently.

7.2 Landmark Reliability and Spatial Distribution

Anatomical landmarks are useful because their identity is independent of the coordinate system and representation used by the acquisition modality. A landmark associated with a specific tooth and anatomical location can therefore be paired between IOS and CBCT even when the two inputs are initially far apart.

The main limitation of this representation is the sensitivity of rigid transformation estimation to landmark quality. When only three or four correspondences are available, the localization error of one point can have a comparatively large influence on the estimated transformation. Additional valid correspondences reduce this dependence and provide redundancy against isolated prediction errors.

Spatial distribution is equally important. Landmarks spread across a large portion of the dental arch provide stronger constraints on the rotational component of the transformation than landmarks concentrated in a small region. Approximately collinear configurations are particularly weak because rotation around the direction defined by the points is poorly constrained. The explicit

non-collinearity check in the RANSAC stage is therefore consistent with the geometric requirements of the problem. Confidence information provides a complementary mechanism for handling uncertainty. In the proposed method, weak CBCT predictions are filtered before registration, and the confidence values of the remaining IOS and CBCT predictions affect the weighted Kabsch estimate. Confidence is nevertheless combined with geometric consensus rather than used as a substitute for it. A high-confidence landmark that is geometrically inconsistent with the other correspondences can still be rejected by RANSAC. The adaptive consensus strategy also reflects the variability of landmark availability across cases. A fixed large consensus would make registration more difficult when only a limited number of landmarks are detected, whereas accepting arbitrary three-point configurations would increase sensitivity to incorrect correspondences. Progressive threshold relaxation and the secondary consensus fallback provide a mechanism for balancing these effects.

7.3 Role of Surface Refinement

Landmark registration provides a semantically meaningful but sparse description of the shared anatomy. The surface-refinement stage complements this representation by exploiting many more geometric samples after a global initialization has been obtained. Tab. 6.4 shows that ICP improves registration even when RANSAC is not used. This indicates that dense surface geometry can compensate for part of the residual landmark error. The strongest improvement is obtained when ICP follows robust landmark initialization, which is consistent with the local nature of the algorithm.

The landmark-guided CBCT crop is an important part of this refinement. The full CBCT volume contains anatomical structures that are not represented in the IOS, including roots, bone, and surrounding tissues. Restricting reconstruction to the region defined by the robust landmark consensus reduces the amount of unrelated geometry presented to ICP and focuses the refinement on the portion most likely to overlap with the IOS. The two registration stages can therefore be interpreted as complementary levels of information. The landmark stage uses sparse semantic information to recover the global transformation, while the surface stage uses dense local geometry to reduce the remaining alignment error.

7.4 Robustness and Failure Modes

The coarse-to-fine organization is particularly relevant when the initial relative pose between IOS and CBCT is unknown. Because the landmark branches operate independently in their respective coordinate systems, the global transformation does not require the two surfaces to be spatially close before correspondence is established.

This property reduces the dependence of the global registration stage on an initial geometric overlap. Once the landmark-based transformation has been estimated, the residual displacement is expected to be sufficiently small for local surface matching to be meaningful.

However, the two stages remain coupled through the quality of the coarse transformation. If landmark localization errors produce an incorrect global alignment, the CBCT crop may contain an inappropriate region and ICP may start outside a suitable convergence basin. Likewise, when fewer than three non-collinear and geometrically consistent correspondences remain available, a unique rigid transformation cannot be recovered by the proposed landmark stage.

The method is therefore robust to isolated correspondence errors but not to complete failure of the anatomical representation. This distinction is important when interpreting the role of RANSAC: robust estimation can reject inconsistent observations when sufficient correct evidence is present, but it cannot create anatomical information that is absent from both modalities.

7.5 Limitations

The first limitation is the dependence on landmark availability. The current pipeline uses only the Lingual and Palatal landmarks associated with tooth positions 1–5 because these provided the required coverage and consistency in the considered data. This restriction limits the amount of information available to the global registration stage and does not guarantee uniform coverage of the complete dental arch in every case.

A second limitation is the dependence on the learned landmark detectors. Orthodontic appliances, restorative materials, incomplete IOS acquisitions, and CBCT artifacts can affect landmark localization. Robust estimation reduces the influence of isolated inaccurate predictions, but it cannot fully compensate when many correspondences fail simultaneously.

The surface-refinement stage also has limitations. ICP remains a local optimization procedure and can converge incorrectly when the coarse transformation is insufficiently accurate. In addition, the reconstructed CBCT surface is an approximation whose quality depends on image resolution, artifacts, the selected intensity threshold, and the chosen region of interest. The IOS and CBCT surfaces therefore do not necessarily have identical geometric fidelity.

Finally, the quantitative evaluation was performed within the MICCAI STS 2026 Challenge setting. The observed transformation-labeling inconsistency described in Chapter 6 further illustrates the importance of verifying coordinate and transformation conventions when evaluating multimodal registration systems. Independent evaluation on additional datasets would be required to study generalization beyond the current challenge data.

7.6 Practical Relevance

Accurate IOS–CBCT registration provides a common three-dimensional reference in which detailed crown geometry can be related to roots, bone, and surrounding structures. Automatic estimation of this relationship can reduce dependence on manually selected correspondences and interactive alignment procedures.

The modular structure of the proposed framework also separates learned anatomical analysis from classical geometric registration. This makes it possible to improve landmark detection, correspondence handling, or surface refinement independently. The approach therefore provides a basis for further development of automated multimodal dental workflows, while its clinical utility would require additional validation under broader acquisition and patient conditions.

8. Conclusions and Future Work

8.1 Conclusions

This thesis investigated the automatic rigid registration of Intraoral Scans (IOS) and Cone-Beam Computed Tomography (CBCT) as a multimodal three-dimensional dental imaging problem. The two modalities provide complementary information but differ substantially in representation, anatomical coverage, and image characteristics. These differences make direct surface or intensity matching unreliable, particularly when the initial relative pose is unknown.

To address this problem, a landmark-guided coarse-to-fine registration framework was developed. The IOS is first transformed into a canonical orientation using PCA and a learned orientation correction. The normalized surface is then processed by tooth segmentation and landmark prediction, while a dedicated volumetric detector identifies the corresponding landmarks in CBCT. Anatomical identities provide the cross-modal correspondence required for global alignment without relying on initial spatial proximity. The landmark correspondences are filtered using confidence information and processed with a confidence-weighted Kabsch estimator embedded in an adaptive RANSAC procedure. The resulting geometric consensus provides a robust global transformation and simultaneously identifies the landmark subset used to determine the relevant CBCT region. A surface is reconstructed from this region and refined against the coarse IOS alignment through point-to-plane ICP. The final transformation is obtained by composing the orientation, landmark-registration, and surface-refinement stages. The ablation study demonstrated the complementary behavior of the two geometric components. On the 30 labeled training cases, direct landmark registration produced an MTE of 13.759 mm and an MRE of 3.993°. Adding RANSAC reduced these values to 8.708 mm and 2.619°, while ICP alone resulted in 7.959 mm and 1.997°. The complete pipeline achieved 1.813 mm and 0.891° on the same subset. On the official 50-case validation set, the complete method achieved an MTE of 4.988 mm and an MRE of 1.421°. These results support the use of semantic anatomical landmarks for global multimodal alignment and dense surface geometry for local refinement. Overall, the work demonstrates a practical strategy for combining learning-based anatomical analysis with classical geometric registration. The main

methodological principle is to use semantic information to establish a global relationship between heterogeneous acquisitions and to introduce dense geometric matching only after an appropriate initialization has been obtained.

8.2 Future Work

Several directions remain open for improving the proposed framework. A first direction concerns the anatomical landmark set. The current method uses the Lingual and Palatal landmarks associated with tooth positions 1–5 because of their observed availability across the considered data. Additional reliably detectable landmarks could increase the number of valid correspondences and improve their spatial coverage over the dental arch. This may be particularly useful in cases with missing teeth or incomplete scans. A second direction is the improvement of landmark localization and uncertainty estimation. Explicit confidence calibration, uncertainty-aware landmark models, or improved handling of orthodontic and restorative artifacts could make the global registration stage less dependent on a small number of predictions. The surface-refinement stage could also be extended. The current implementation reconstructs a CBCT surface using a fixed intensity threshold and applies point-to-plane ICP. Alternative surface-extraction methods, robust local objectives, or correspondence strategies could reduce sensitivity to CBCT artifacts and reconstruction quality. Another possible development is a closer integration of the coarse and fine stages. Rather than treating landmark alignment and surface refinement as strictly sequential procedures, future methods could jointly optimize anatomical correspondence reliability, rigid transformation parameters, and local surface agreement. Finally, evaluation on larger and more heterogeneous datasets would provide a broader assessment of generalization. Data acquired with different scanners, reconstruction protocols, and clinical conditions would be particularly useful for determining how the method behaves beyond the specific challenge setting considered in this thesis.

8.3 Final Remarks

Multimodal registration provides the geometric foundation required to combine heterogeneous three-dimensional dental acquisitions within a common reference frame. The framework presented

in this thesis uses anatomical landmarks to bridge the representational gap between IOS and CBCT and subsequently exploits dense surface information to refine the alignment. The resulting coarse-to-fine formulation provides a basis for further work on robust and fully automatic integration of multimodal dental data.

A. Challenge Paper Publication

The research activity conducted in the context of this thesis has led to the submission and acceptance of a paper for the STS 2026 challenge, along with an oral presentation at the ODIN workshop within MICCAI 2026 in Strasbourg, France. For the sake of completeness and to provide the reader with direct access to the core results and methodology, the paper is included in full in the following pages.

A. Veronese, M. Lugli, O. Carpentiero, K. Marchesini, L. Lumetti, and F. Bolelli.
"Landmark-Guided Coarse-to-Fine Registration of Intraoral Scans and Cone-Beam CT"

Landmark-Guided Coarse-to-Fine Registration of Intraoral Scans and Cone-Beam CT

Alex Veronese, Matteo Lugli, Omar Carpentiero,
Kevin Marchesini, Luca Lumetti, and Federico Bolelli ✉

University of Modena and Reggio Emilia, Modena, Italy
{name.surname}@unimore.it

Abstract. Accurate registration of Intraoral Scans (IOS) with Cone-Beam Computed Tomography (CBCT) enables integration of dental crown surfaces with the tooth roots and the surrounding alveolar bone. However, IOS-CBCT registration remains challenging because of limited anatomical overlap, cross-modal differences, and unreliable correspondences. Task 2 of the MICCAI STS 2026 Challenge formulates this setting as a semi-supervised rigid registration problem, in which IOS meshes must be aligned with their corresponding CBCT volumes. For this task, we propose a landmark-driven coarse-to-fine framework in which modality-specific networks predict corresponding dental landmarks from the IOS mesh and CBCT volume. A confidence-weighted RANSAC-Kabsch procedure estimates a robust initial transformation, later refined using point-to-plane Iterative Closest Point. On the challenge validation set, the method achieved a *Mean Translation Error* of 4.988 mm and a *Mean Rotation Error* of 1.421°, ranking first on the registration leaderboard at the time of evaluation. The code is available on GitHub: https://github.com/AImageLab-zip/L2L-Registration

Keywords: IOS-CBCT registration · Dental imaging · Landmark detection · RANSAC · Iterative Closest Point

1 Introduction

Cone-Beam Computed Tomography (CBCT) and Intraoral Scans (IOS) provide complementary representations of dentoalveolar anatomy [3,6,18]. CBCT captures tooth roots and alveolar bone, but its representation of dental crowns may be degraded by limited resolution, noise, partial-volume effects, and metal artifacts. In contrast, IOS provides high-resolution surface models of the visible crowns and gingiva [4], but does not capture roots or underlying bone and may exhibit stitching errors in full-arch scans [17,23]. Accurate IOS-CBCT registration therefore enables the construction of integrated crown-root-bone models for orthodontic diagnosis, implant planning, orthognathic surgery, and longitudinal assessment of tooth movement. This task is challenging because the modalities

✉ Corresponding author: federico.bolelli@unimore.it

differ in data representation and share only a limited anatomical region: the exposed dental elements. Reliable correspondence estimation may be hindered by missing teeth, similar or symmetric tooth morphology, restorations, orthodontic appliances, metal artifacts, and incomplete scans. Prior studies have also shown that registration performance depends strongly on the accuracy of tooth segmentation and identification. Moreover, large initial pose differences may cause local surface-registration methods to converge to anatomically incorrect alignments [10,23]. Learning-based approaches face the additional limitation of scarce paired data with accurate transformations.

Conventional workflows typically estimate an initial transformation from manually selected landmarks or dental surfaces and refine the alignment using Iterative Closest Point (ICP) [2]. Although computationally efficient, ICP is sensitive to initialization, partial overlap, outliers, and repetitive dental geometry. Recent automatic methods have therefore adopted coarse-to-fine pipelines combining semantic preprocessing, robust global alignment, and local surface refinement. Kim et al. [14] combined deep learning-based tooth segmentation with RANSAC-based coarse registration [8] and ICP refinement. Jang et al. [10] incorporated individual-tooth segmentation and identification into a global-to-local framework while addressing cumulative deformation in full-arch IOS scans. More recently, point-cloud-based approaches have been investigated, including FPFH-based geometric matching [15], PointNet++ feature encoding with differentiable SVD-based rigid-transformation estimation [11], and PointNetLK-based iterative alignment [9,19]. However, these approaches may remain sensitive to cross-modal geometric discrepancies, limited anatomical overlap, and large initial pose differences.

Task 2 of the MICCAI STS 2026 Challenge formulates IOS-CBCT alignment as a semi-supervised rigid-registration problem and provides labeled and unlabeled dental data for standardized evaluation. Building on recent advances in modality-specific anatomical landmark prediction [7], we propose a landmark-guided coarse-to-fine registration framework. Our main contribution is the robust use of predicted anatomical correspondences: a confidence-weighted adaptive RANSAC-Kabsch procedure identifies a reliable landmark consensus for global alignment, and the resulting inliers guide the subsequent local surface refinement. This design reduces the influence of inaccurate landmark predictions and unrelated CBCT anatomy while improving robustness to limited overlap and large initial pose differences.

2 Method

We propose a landmark-driven pipeline for rigid IOS-CBCT registration. As illustrated in Fig. 1, the IOS and CBCT data are processed in parallel to predict anatomically corresponding dental landmarks. In the IOS branch, the input mesh is centered and oriented by the *IOS-Normalizer* module according to the reference frame required by the subsequent *IOS-Landmarks* module, which estimates the three-dimensional coordinates of the selected dental landmarks. In

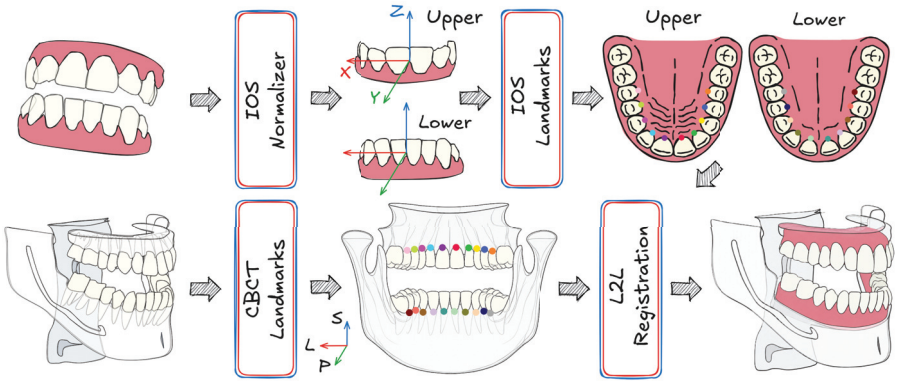


Fig. 1: Overview of the proposed IOS-CBCT registration pipeline.

parallel, the *CBCT-Landmarks* module localizes the corresponding landmarks directly in the CBCT volume. The predicted correspondences are provided to the Landmarks-to-Landmarks registration module (*L2L-Registration* in short), which estimates an initial rigid IOS-CBCT transformation using a confidence-weighted RANSAC-Kabsch procedure [8,12,13]. The resulting alignment is subsequently refined using point-to-plane ICP [2]. This coarse-to-fine strategy reduces sensitivity to large initial pose differences and improves the final alignment between the IOS surface and the corresponding CBCT dental surface.

IOS-Normalizer. Given an input dental mesh, this module[1] aligns each IOS with the reference frame used to train the IOS-Landmarks module. The normalization procedure combines principal component analysis (PCA) with a learned rotation correction. A fixed number of points is sampled from the unit-sphere normalized vertices using farthest-point sampling (FPS). PCA is then applied to the covariance matrix of the sampled point cloud. The ordered eigenvectors form an orthonormal basis B that aligns the scan with its principal geometric axes. If $\det(B) < 0$, the sign of the final eigenvector is reversed to ensure a right-handed coordinate system. Because PCA does not uniquely determine the signs of the eigenvectors, residual orientation ambiguities may remain. To resolve these ambiguities, the PCA-aligned point cloud is processed by a small PointNet-style network [5], which predicts a single corrective rotation R selected from four predefined classes: the identity rotation and 180° rotations about the x-, y-, and z-axes. The transformed mesh vertices are computed as $X' = (X - c)BR$, where c denotes the mesh centroid.

IOS-Landmarks. Given a single IOS, this module predicts dental landmarks using a two-stage deep learning architecture.[2] First, a semantic segmentation network classifies each mesh vertex into one of 17 classes, representing the 16 teeth of the input arch and a combined gingiva/noise class. At inference time,

[1] https://github.com/AImageLab-zip/IOS-Normalizer/commit/ecebe11
[2] https://github.com/AImageLab-zip/IOS-Landmarks/commit/3ee1ee6

Table 1: L2L-Registration Parameters

Stage	Parameter	Setting
(i) Landmark	Min. landmark confidence	0.25
	RANSAC hypotheses	2,000
	Landmark pairs per hypothesis	3 non-collinear pairs
	Initial inlier distance	3.5 mm
	Max. inlier distance	5.5 mm
	Initial min. consensus	4 landmarks[a]
	Fallback min. consensus	3 non-collinear landmarks
(ii) Surface	Crop margin	max(15 mm, $0.25E$)
	Marching-cubes threshold	1,500
	Max. sampled points	12,000 per surface
	Voxel size	1.0 mm
	Normal-estimation radius	3.0 mm
(iii) ICP	Max. correspondence distance	1.0 mm
	Max. iterations	40
	Relative fitness tolerance	10^{-5}
	Relative RMSE tolerance	10^{-5}

E denotes the landmark extent along the corresponding spatial axis.
[a]Applied when at least four landmark correspondences were available.
RMSE: root-mean-square error.

the tooth classes are mapped to their corresponding FDI numbers using the known arch identity (lower or upper). The vertices assigned to each tooth are processed independently by a landmark-prediction network that generates a per-vertex probability heatmap. Final landmark coordinates are obtained through thresholding and K-means clustering. Both segmentation and landmark prediction models are based on Point Transformer V3 [22]. Among the predicted landmarks, only the *Inner* ones, illustrated in Fig. 1, were used directly in the subsequent stages. This choice was motivated by the variable quality of the IOS scans observed in the training and validation sets. Some scans contained residual orthodontic brackets or dental impression material in the anterior region, potentially affecting landmarks located near the external tooth surface. To assess landmark availability across the dataset, we measured the proportion of tooth instances with a valid *Inner* prediction in each of the four hemi-arches. Applying a minimum prediction coverage threshold of 94% in each hemi-arch retained tooth positions 1–5 for subsequent processing.

CBCT-Landmarks. This module was developed using nnLandmark [7], an extension of the nnU-Net infrastructure to 3D landmark localization via heatmap regression. Landmark annotations are encoded as a $3\times3\times3$ cube around the target voxel. During inference, the input volume is processed using sliding-window prediction. For each landmark channel, the location of the maximum response defines the predicted coordinate, while the corresponding response value is retained as a confidence score. For each tooth previously selected for the IOS, only the corresponding *Inner* landmark was retained.

L2L-Registration. The final IOS-CBCT registration was performed in two stages: *(i)* a landmark-based rigid alignment, followed by *(ii)* surface-based refinement. Hyperparameters for these stages are summarized in Tab. 1. We se-

Table 2: Development environments and requirements.

Component	**Configuration**
Operating system	Ubuntu 22.04.5 LTS
Processor	Intel Xeon Silver 4216 CPU @ 2.10 GHz
System memory	128 GB
Graphics processor	1 × NVIDIA RTX A5000, 24 GB
CUDA version	`12.6`
Programming language	`Python 3.10`
Deep learning framework	`PyTorch 2.8`
Computer vision library	`torchvision 0.23.0`

lected the final configuration through a grid search that considered registration failures, combined translation and rotation errors, and computational cost. Predicted CBCT landmarks below the confidence threshold were excluded, and the remaining landmarks were converted from voxel indices to physical coordinates using the image metadata. An initial rigid transformation was estimated using a confidence-weighted RANSAC-Kabsch procedure [8,12,13]. For each hypothesis, three non-collinear landmark pairs were sampled and used to estimate a rigid transformation. The transformation was evaluated on all available correspondences, and the best solution was selected according to the inlier consensus, combined confidence, and residual error. When no valid consensus was obtained, the inlier threshold was progressively relaxed and, as a final fallback, the minimum consensus size was reduced to three non-collinear landmarks. The selected transformation was then re-estimated using all inliers until the consensus set stabilized and was applied to the normalized IOS mesh. For surface refinement, the landmark inliers were used to define an axis-aligned crop of the CBCT volume. A CBCT surface was extracted from the cropped region using marching cubes, after which the IOS and CBCT surfaces were sampled, voxel-downsampled, and assigned surface normals. The landmark-aligned IOS point cloud was then refined against the CBCT point cloud using point-to-plane ICP. The resulting ICP transformation was composed with the landmark-based transformation to obtain the final IOS-CBCT registration.

3 Experiments

Dataset and Evaluation Metrics. All experiments were conducted using the official MICCAI STS 2026 Challenge dataset, which comprises: *(i)* 30 cases with ground-truth 4×4 transformations and 300 unlabeled cases (*training set*) and *(ii)* 50 cases with ground-truth transformations withheld from participants and accessible only to the challenge organizers (*validation set*). Performance was evaluated using two metrics: *(i) Mean Translation Error* (MTE) and *(ii) Mean Rotation Error* (MRE). The former is defined as the Euclidean distance between the predicted and ground-truth translation vectors; the latter is computed as the geodesic distance between the predicted and reference rotation matrices.

Environments and Requirements. See Tab. 2.

Training Protocols. For this challenge, IOS-Normalizer was fine-tuned from a released checkpoint, IOS-Landmarks was used with pretrained weights, and CBCT-Landmarks was trained from scratch. The corresponding training protocols are described below and summarized in Tab. 4.

IOS-Normalizer. The model was initialized from a publicly available checkpoint pretrained on Teeth3DS [1], which includes 1,800 consistently oriented scans. It was then fine-tuned on 60 IOS from the unlabeled training set of the STS 2026 Challenge. All 60 scans were manually oriented; 50 were used for training and 10 for validation. During training, each oriented scan was downsampled to 16,384 points via FPS and either left unchanged or rotated by 180° about one of the three coordinate axes. The model was trained to identify the applied transformation, and, because a 180° rotation about an axis is self-inverse, reapplying the same transformation restores the scan to its original orientation.

IOS-Landmarks. Publicly available weights were used for both stages (segmentation and landmark prediction).

CBCT-Landmarks. We constructed a pseudo-labeled dataset from the 300 unlabeled CBCT cases in the MICCAI STS 2026 training set. Reference landmarks were obtained by transferring predictions from IOS-Landmarks to the paired CBCT volumes using an internal IOS-CBCT registration model.[3] Only 201 cases were retained after qualitative evaluation. The challenge validation set and the labeled training set (30 samples) were reserved exclusively for evaluation. Training of the CBCT landmark predictor was restricted to the *Inner* landmarks of tooth positions 1-5 in each hemi-arch, resulting in 20 output classes. Standard augmentations and foreground oversampling were applied, while left-right mirroring was disabled. The checkpoint from the fold achieving the best validation performance was used for inference. To accelerate inference, nnLandmark [7] was modified to track the maximum response and corresponding coordinate for each landmark directly across sliding-window patches, avoiding full heatmap reconstruction and CPU transfer. Batched patch processing, persistent GPU model loading, and GPU-based resampling with `resample_torch_for_nnunet` further reduced memory usage and latency.

Table 3: Ablation study on the labeled training subset and final registration performance on the validation set of the MICCAI STS 2026 Challenge.

Eval. Set	RANSAC	ICP	MTE [mm]	MRE [°]
Train	✗	✗	13.759	3.993
Train	✓	✗	8.708	2.619
Train	✗	✓	7.959	1.997
Train	✓	✓	1.813	0.891
Val.	✓	✓	4.988	1.421

4 Results and Discussion

The contribution of the individual registration components was assessed through an ablation study on the labeled training subset, for which reference IOS-CBCT

[3] Such a registration model cannot be released due to confidentiality obligations; therefore, it has not been used in the challenge itself.

Table 4: Training protocols for IOS-Normalizer and CBCT-Landmarks.

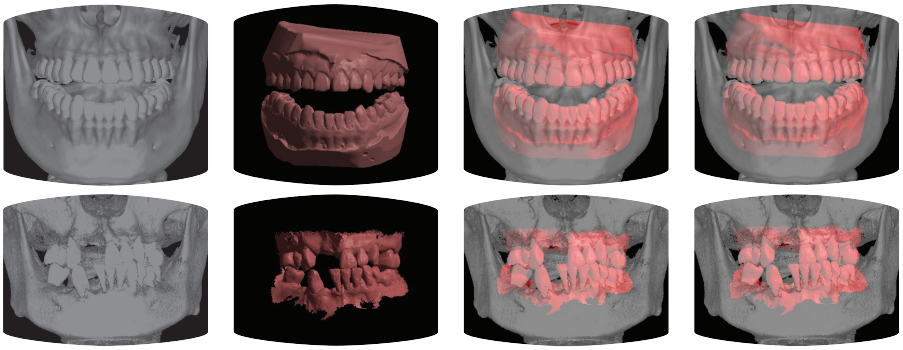
Parameter	**IOS-Normalizer**	**CBCT-Landmarks**
Batch size	1	2
Input patch size	—	128^3 voxels
Epochs	3	280
Optimizer	AdamW ($\lambda = 1e^{-4}$)	Nesterov SGD ($\mu = 0.99$)
Initial LR	$1e^{-4}$	$1e^{-2}$
LR schedule	None	Polynomial ($p = 0.9$)
Training time	4.5 min	87 h (5-fold CV)
Loss function	Cross-entropy over 4 classes	Top-20% BCE (logits)
Model parameters	13 M	102 M
FLOPs	91.3 GFLOPs/step	8.6 TFLOPs/step

transformation matrices were available. These matrices were not used to train or fine-tune any component of the proposed pipeline and were used only to quantify registration accuracy. As reported in Tab. 3, RANSAC improved registration performance both with and without ICP refinement. Point-to-plane ICP provided an additional improvement by refining the landmark-based transformation using the reconstructed surface geometry.

Quantitative Results on Validation. The proposed method achieved an MTE of 4.988 mm and an MRE of 1.421° on the challenge validation set, as reported in Tab. 3. At the submission deadline, these results placed the method first on the challenge registration leaderboard. The comparatively low rotation error indicates that the rotational component of the IOS-CBCT transformation was recovered accurately. Overall, the results support the combination of robust landmark-based initialization and local point-to-plane ICP refinement.

Qualitative Results on Validation. Qualitative inspection indicated that registration was generally more accurate when several reliable IOS-CBCT landmark correspondences were available and spatially distributed across the dental arch. Registrations based on only three or four correspondences appeared more sensitive to landmark localization errors. In these cases, ICP provided a significant improvement by refining the initial transformation using the reconstructed surface geometry. Visual examples are provided in Fig. 2.

Identification of a Systematic Transformation Labeling Error. Inspection of the 4×4 transformation matrices showed that all training cases with IDs above 269 had 3×3 linear-transformation submatrices with determinant -1, indicating a reflection across the YZ plane, as confirmed by visual comparison of the original and transformed IOS meshes. During validation, an initial submission using only positive-determinant rotations yielded an MTE of 10.037 mm and an MRE of 22.809°. Assuming the same ID-dependent inconsistency affected the validation set, we composed predictions for cases with IDs above 269 with the YZ-plane reflection; the resulting evaluation is the one previously reported in Tab. 3. The organizers later confirmed the labeling issue and our interpretation via email. Evaluation transparency was further limited by the post hoc inclusion of inference time as a weighted ranking component, despite the



(a) CBCT (b) IOS (c) Kabsch (d) ICP

Fig. 2: Qualitative results, one patient per row (P027 and P214 from the top).

absence of a runtime metric or formal time limit in the structured submission document [21] or on the challenge website. To our knowledge, neither issue was publicly communicated, limiting reproducibility and evaluation fairness.

Limitations and Future Work. Although no visually apparent registration failures were observed on the validation set, the proposed method may fail when fewer than three geometrically consistent correspondences are available, when the selected landmarks form a nearly collinear configuration, or when localization errors place the initial alignment outside the convergence basin of ICP. Future work will investigate additional anatomical landmarks that are consistently present and can be reliably localized in both IOS and CBCT data. Increasing the number and spatial coverage of cross-modal correspondences may improve robustness in cases involving missing teeth, incomplete scans, imaging artifacts, or inaccurate landmark predictions.

5 Conclusion

This work presented a robust coarse-to-fine framework for rigid IOS-CBCT registration that combines modality-specific anatomical landmark detection, confidence-weighted RANSAC-Kabsch alignment, and point-to-plane ICP refinement. The proposed approach reduces sensitivity to large initial pose differences and inconsistent cross-modal correspondences by combining robust anatomical initialization with local surface refinement. On the challenge validation set, the method achieved an MTE of 4.988 mm and an MRE of 1.421°. These results demonstrate the effectiveness of integrating anatomical landmark correspondences with geometric surface refinement for IOS-CBCT alignment.

Acknowledgments. This work was developed and evaluated as part of the ODIN 2026 Challenge Cluster, including the STS, ToothFairy, and Bite2Text challenges [16,17,19,20].

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Ben-Hamadou, A., Neifar, N., Rekik, A., Smaoui, O., Bouzguenda, F., Pujades, S., Boyer, E., Ladroit, E.: Teeth3DS+: An Extended Benchmark for Intraoral 3D Scans Analysis. arXiv preprint arXiv:2210.06094 (2022)
2. Besl, P., McKay, N.: A method for registration of 3-D shapes. IEEE Transactions on Pattern Analysis and Machine Intelligence **14** (1992)
3. Bolelli, F., Lumetti, L., van Nistelrooij, N., Vinayahalingam, S., Di Bartolomeo, M., Marchesini, K., Pellacani, A., Candeloro, E., Rosati, G., Xi, T., Isensee, F., Kirchhoff, Y., Krämer, L., Rokuss, M., Ulrich, C., Maier-Hein, K., Jiang, Y., Liu, Y., Wang, L., Wang, H., Chen, S., Cui, Z., Shi, P., Pan, Z., Liang, X., Ma, Q., Konukoglu, E., Wodzinski, M., Müller, H., Mai, H., Dang, X., Bhandary, S., Grosu, R., Bergé, S., Anesi, A., Grana, C.: Multi-Structure Segmentation in CBCT Volumes: the ToothFairy2 Challenge. Medical Image Analysis (2026)
4. Borghi, L., Lumetti, L., Cremonini, F., Rizzo, F., Grana, C., Lombardo, L., Bolelli, F.: Bits2Bites: Intra-oral Scans Occlusal Classification. In: International Workshop on Oral and Dental Image Analysis. pp. 45–54 (2025)
5. Charles, R.Q., Su, H., Kaichun, M., Guibas, L.J.: PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2017)
6. Di Bartolomeo, M., Pellacani, A., Bolelli, F., Cipriano, M., Lumetti, L., Negrello, S., Allegretti, S., Minafra, P., Pollastri, F., Nocini, R., et al.: Inferior Alveolar Canal Automatic Detection with Deep Learning CNNs on CBCTs: Development of a Novel Model and Release of Open-Source Dataset and Algorithm. Applied Sciences **13**(5) (2023)
7. Ertl, A., Denner, S., Peretzke, R., Xiao, S., Zimmerer, D., Fischer, M., Bujotzek, M.R., Yang, X., Neher, P., Isensee, F., Maier-Hein, K.: nnLandmark: A Self-Configuring Method for 3D Medical Landmark Detection. In: Medical Imaging with Deep Learning (2026)
8. Fischler, M.A., Bolles, R.C.: Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. In: Fischler, M.A., Firschein, O. (eds.) Readings in Computer Vision. San Francisco (CA) (1987)
9. George, A.B., Lekshmi, P.G., Bathini, S., Govind, A.: Semi-supervised CBCT–IOS registration using PointNetLK. In: Oral and Dental Image Analysis. Lecture Notes in Computer Science, vol. 16473 (2026)
10. Jang, T.J., Yun, H.S., Hyun, C.M., Kim, J.E., Lee, S.H., Seo, J.K.: Fully automatic integration of dental CBCT images and full-arch intraoral impressions with stitching error correction via individual tooth segmentation and identification. Medical Image Analysis **93** (2024)
11. Ji, C., Liu, Y., Jiang, Y., Wang, L.: Learning-Based CBCT–IOS Registration with PointNet++ and SVD. In: Oral and Dental Image Analysis. Lecture Notes in Computer Science, vol. 16473 (2026)
12. Kabsch, W.: A solution for the best rotation to relate two sets of vectors. Acta Crystallographica Section A **32**(5) (1976)
13. Kabsch, W.: A discussion of the solution for the best rotation to relate two sets of vectors. Acta Crystallographica Section A **34**(5) (1978)
14. Kim, Y.J., Ahn, J.H., Lim, H.K., Nguyen, T.P., Jha, N., Kim, A., Yoon, J.: Novel Procedure for Automatic Registration between Cone-Beam Computed Tomography and Intraoral Scan Data Supported with 3D Segmentation. Bioengineering **10**(11) (2023)

15. Liu, J., Hao, J., Lin, H., Pan, W., Yang, J., Feng, Y., Wang, G., Li, J., Jin, Z., Zhao, Z., Liu, Z.: Deep Learning-Enabled 3D Multimodal Fusion of Cone-Beam CT and Intraoral Mesh Scans for Clinically Applicable Tooth-Bone Reconstruction. Patterns **4**(9) (2023)
16. Lumetti, L., Di Bartolomeo, M., Pellacani, A., Anesi, A., Grana, C., Bolelli, F.: Ontology-Grounded Structured Prediction for Dental CBCT Reporting. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2026 (2026)
17. Lumetti, L., Rizzo, F., Cremonini, F., Candeloro, E., Lombardo, L., Grana, C., Bolelli, F.: Do Multimodal LLMs Understand Intraoral Dental Data? Dataset, Platform, and Baselines. In: European Conference on Computer Vision (ECCV) (2026)
18. Lumetti, L., Tan, Z.Q., Borghi, L., van Nistelrooij, N., Rosati, G., Addison, O., Li, Y., Vinayahalingam, S., Grana, C., Bolelli, F.: ToothFairy3: Scaling CBCT Maxillofacial Segmentation to 77 Classes with U-Mamba2. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2026 (2026)
19. Wang, Y., Zhang, Y., Chen, X., Wang, S., Qian, D., Ye, F., Xu, F., Zhang, H., Dan, R., Zhang, Q., et al.: MICCAI 2023 STS Challenge: A retrospective study of semi-supervised approaches for teeth segmentation. Pattern Recognition **170** (2026)
20. Wang, Y., Li, Z., Wu, C., Liu, J., Zhang, Y., Ni, J., Luo, Q., Chen, J., Zhang, H., Liu, J., Han, C., Fu, K., Ji, C., Cai, X., Hao, J., Zheng, Z., Xu, S., Chen, J., Yu, X., Zhang, Q., Qian, D., Wang, S., Zhou, H.: MICCAI STS 2024 challenge: Semi-supervised instance-level tooth segmentation in panoramic X-ray and CBCT images. Medical Image Analysis **110** (2026)
21. Wang, Y., Wang, S., Li, Z.: STS 2026: The 4th Semi-supervised Teeth Segmentation Challenge on Metal Artifact Reduction and Beyond (2026)
22. Wu, X., Jiang, L., Wang, P.S., Liu, Z., Liu, X., Qiao, Y., Ouyang, W., He, T., Zhao, H.: Point Transformer V3: Simpler, Faster, Stronger. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2024)
23. Zheng, Q., Wu, Y., Chen, J., Wang, X., Zhou, M., Li, H., Lin, J., Zhang, W., Chen, X.: Automatic multimodal registration of cone-beam computed tomography and intraoral scans: a systematic review and meta-analysis. Clinical Oral Investigations **29** (2025)

Acknowledgements

First and foremost, I would like to express my deepest gratitude to Prof. Federico Bolelli and Prof. Costantino Grana for their invaluable guidance and for giving me the opportunity to join the AImageLab-Zip research group. I have greatly enjoyed the time spent both inside and outside the laboratory with my colleagues, from whom I have learned a great deal.

In particular, I would like to thank Matteo Lugli, Omar Carpentiero, Kevin Marchesini, and Luca Lumetti for their support throughout this journey. We have shared many great moments, both professionally and personally, and the results we achieved together are a testament to the value of this experience. A heartfelt thank you also goes to all the other Zippers for the memorable moments in the lab, the traditional '*Zona*" lunches on Fridays, and the wonderful experience at MICCAI 2026 in Strasbourg. I look forward to starting new projects and sharing new experiences with all of you.

I would also like to thank all my university, lifelong, and recent friends. Over the past two years, I have realized even more how important you are to me and how much your support has helped me through demanding times.

My deepest thanks go to my parents Raffaella and Vanni, my brother Mattia, and to all my family. My mind is often elsewhere when I am at home, and perhaps I have never fully shared with you all the details of this journey. Nevertheless, I hope that I have made you proud of this achievement and of everything I have accomplished over the past two years, just as I am.

Last but not least, I would like to give a very special thank you to Gemma, or, better, Nana. You have been by my side throughout this entire journey, always encouraging me to keep going and never give up, making you my safe place, while also being the person who knows when to challenge my stubbornness and push me to do better. As our favourite artist sings, "*Riempimi i muri come vuoi*": you have done the same with all the time we have spent together, filling my life with moments, memories, and meaning. I cannot wait to paint many more pictures with you, filling our own story, or, as the song says, our own "new home".

Finally, great work Alex. I know you have not fully realized it yet, and I am sure you will still find something to be dissatisfied with. Perhaps now you will finally have some time to learn that sometimes, you simply have to stop and appreciate how far you have come.

Bibliography

[1] Achraf Ben-Hamadou et al. “Teeth3DS+: An Extended Benchmark for Intraoral 3D Scans Analysis”. In: *arXiv preprint arXiv:2210.06094* (2022).

[2] Paul Besl and Neil McKay. “A method for registration of 3-D shapes”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 14 (1992).

[3] Federico Bolelli et al. “Multi-Structure Segmentation in CBCT Volumes: the ToothFairy2 Challenge”. In: *Medical Image Analysis* (2026).

[4] Lorenzo Borghi et al. “Bits2Bites: Intra-oral Scans Occlusal Classification”. In: *International Workshop on Oral and Dental Image Analysis*. 2025, pp. 45–54.

[5] R Qi Charles et al. “PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation”. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2017.

[6] Mattia Di Bartolomeo et al. “Inferior Alveolar Canal Automatic Detection with Deep Learning CNNs on CBCTs: Development of a Novel Model and Release of Open-Source Dataset and Algorithm”. In: *Applied Sciences* 13.5 (2023).

[7] Alexandra Ertl et al. “nnLandmark: A Self-Configuring Method for 3D Medical Landmark Detection”. In: *Medical Imaging with Deep Learning*. 2026.

[8] Martin A. Fischler and Robert C. Bolles. “Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography”. In: *Readings in Computer Vision*. Ed. by Martin A. Fischler and Oscar Firschein. San Francisco (CA), 1987. ISBN: 978-0-08-051581-6.

[9] Ajo Babu George et al. “Semi-Supervised CBCT–IOS Registration Using PointNetLK”. In: *Oral and Dental Image Analysis*. Vol. 16473. Lecture Notes in Computer Science. 2026.

[10] Fabian Isensee et al. “nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation”. In: *Nature methods* 18.2 (2021), pp. 203–211.

[11] Tae Jun Jang et al. “Fully automatic integration of dental CBCT images and full-arch intraoral impressions with stitching error correction via individual tooth segmentation and identification”. In: *Medical Image Analysis* 93 (2024). ISSN: 1361-8415.

[12] Changkai Ji et al. “Learning-Based CBCT–IOS Registration with PointNet++ and SVD”. In: *Oral and Dental Image Analysis*. Vol. 16473. Lecture Notes in Computer Science. 2026.

[13] W. Kabsch. “A discussion of the solution for the best rotation to relate two sets of vectors”. In: *Acta Crystallographica Section A* 34.5 (1978).

[14] W. Kabsch. “A solution for the best rotation to relate two sets of vectors”. In: *Acta Crystallographica Section A* 32.5 (1976).

[15] Yoon-Ji Kim et al. “Novel Procedure for Automatic Registration between Cone-Beam Computed Tomography and Intraoral Scan Data Supported with 3D Segmentation”. In: *Bioengineering* 10.11 (2023). ISSN: 2306-5354.

[16] Jiaxiang Liu et al. “Deep Learning-Enabled 3D Multimodal Fusion of Cone-Beam CT and Intraoral Mesh Scans for Clinically Applicable Tooth-Bone Reconstruction”. In: *Patterns* 4.9 (2023).

[17] Luca Lumetti et al. “ToothFairy3: Scaling CBCT Maxillofacial Segmentation to 77 Classes with U-Mamba2”. In: *Medical Image Computing and Computer Assisted Intervention – MICCAI 2026*. 2026.

[18] Charles Ruizhongtai Qi et al. “PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space”. In: *Advances in Neural Information Processing Systems*. Ed. by I. Guyon et al. Vol. 30. 2017.

[19] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. “U-Net: Convolutional Networks for Biomedical Image Segmentation”. In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. Ed. by Nassir Navab et al. 2015, pp. 234–241.

[20] Xiaoyang Wu et al. “Point Transformer V3: Simpler, Faster, Stronger”. In: *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2024.

[21] Qianhan Zheng et al. “Automatic multimodal registration of cone-beam computed tomography and intraoral scans: a systematic review and meta-analysis”. In: *Clinical Oral Investigations* 29 (2025).