MELBA Journal Machine Learning for Biomedical Imaging

Deep Spectral Models for Robust Dental Shape Generation

Tibor **Kubík** [1,2], François Guibault [1], Michal Španěl [2], Hervé Lombaert [1],

1 Polytechnique Montréal, Montréal, Canada
2 Brno University of Technology, Brno, Czech Republic

Abstract

Accurate modeling of dental crown morphology is fundamental for diagnosis, orthodontic planning, and computer-aided restoration design. However, datasets suitable for training such models are typically limited in size. We present ToothForge, a deep spectral generative framework that models dental crown geometries from compact, intrinsic representations. By operating in the spectral domain, ToothForge learns a latent manifold of 3D tooth shapes through synchronized spectral embeddings, ensuring consistent modeling across samples with varying connectivity. Spectral synchronization mitigates the instability of Laplace-Beltrami eigenbases and enables efficient learning in a low-dimensional space. The framework is thoroughly evaluated through robustness analysis, ablation studies, and benchmarking against PCA-based statistical shape models and point-based generative frameworks. Results show that synchronized spectral modeling achieves reconstruction and generative performance comparable to or exceeding spatial approaches, while maintaining compactness and geometric interpretability. Together, the compact synchronized coefficients and low-dimensional learning space make the framework particularly suitable for limited datasets, as often encountered in dental and medical domains, and applicable in real-world scenarios where guaranteeing consistent connectivity across shapes from various clinics is unrealistic.

Keywords

3D Tooth Shape Generation, Digital Dentistry, Spectral Shape Learning, Geometric Deep Learning

Article informations

https://doi.org/10.59275/j.melba.2026-522e
Received: 2025-10-31, Published 2026-06-08
Corresponding author: ikubik@fit.vut.cz
Special issue: Information Processing in Medical Imaging (IPMI) 2025
Guest editors: Ipek Oguz, Shireen Elhabian, Ismail Ben Ayed

1. Introduction

Digital dentistry relies heavily on 3D shape analysis for tasks such as morphology assessment and computer-aided dental restorations. Among morphology assessment applications, accurate modeling of tooth and root morphology plays a critical role in several clinical domains. In dentofacial orthopaedics, evaluating palatal shape morphology has the potential to aid in the evaluation and outcome prediction of maxillary expansion (Matsuyama et al.; Nauwelaers et al.; Primožič et al.). In endodontics, dental pulp treatment, knowledge of root canal anatomy is crucial for guiding canal detection and preventing procedural complications such as perforation (Cleghorn et al.; Peiris et al.). Similarly, in implantology, knowledge of crown and root morphology determines implant trajectory, proximity to the sinus floor, and the feasibility of immediate implant placement after extraction (Bhola et al.; Testori et al.).

Beyond diagnostic and surgical applications, automated generation of dental restorations represents key domain where accurate 3D modeling of tooth morphology is critical (Kong and Kim). In restorative dentistry, digital workflows increasingly rely on data-driven crown design, where prosthetic shapes must adapt to the patient's existing anatomy and occlusal context. The ability to synthesize realistic and anatomically feasible crown restorations can substantially reduce manual design time and thus improve clinical and laboratory practice (Hosseinimanesh et al.; Tian et al.; Yang et al.). However, these frameworks require large, diverse, and well-annotated datasets of dental shapes to achieve generalization across patients and tooth classes. Such requirement is rarely met in practise due to privacy constraints and costly annotations.

Collectively, these applications highlight two complementary needs in digital dentistry: (i) shape models that

capture population-level variability of dental morphology for analysis and diagnosis, and (ii) generative frameworks capable of synthesizing realistic shapes for restorative design and data augmentation.

As a traditional baseline, statistical shape models (SSMs) model shape variation by principal component analysis (PCA) of training data shapes (Wold et al.). By representing a population of shapes through low-dimensional linear modes of variation, PCA-based SSMs provide interpretable and efficient frameworks for morphological analysis and reconstruction. They also possess generative power by modifying the variables within given shape population. However, due to their inherently linear formulation, these models struggle to capture complex anatomical geometries. Examples include the non-linear evolution of palatal dimensions during dentition development and the non-convex relationships among occlusal surfaces and cusp morphologies.

The recent emergence of 3D deep learning has extended the paradigm of SSMs toward non-linear representations, usually operating over spatial coordinates (Adams and Elhabian; Pan et al.; Yang et al.; Yang et al.; Gutiérrez-Becker et al.; Luo and Hu). Autoencoder (AE) networks can be described as a non-linear generalization of PCA. They are capable of capturing population-level variability while serving as a generative model for synthetizing novel shapes, both in a non-linear setting. Moreover, AEs are easily expanded to provide unified modeling across multiple classes, integrate contextual or multimodal inputs, or perform additional downstream analysis like classification, regression or pathology detection. Despite their flexibility, autoencoders also introduce several practical limitations when applied to anatomical shape modeling. First, they scale poorly with resolution. Training directly on high-dimensional spatial coordinates or dense meshes requires large datasets and considerable computational resources. This is particularly challenging in medical domains such as digital dentistry, where datasets are small but each sample contains tens or hundreds of thousands of vertices (Hosseinimanesh et al.; Kubík and Španěl; Tan et al.; Wang et al.). This dimensionality leads to a severe imbalance between data size and representational dimensionality. In such settings, overfitting becomes a critical risk, and training stability often depends on extensive downsampling, which may compromise geometric fidelity. Second, unlike PCA-based SSMs, the latent representations learned by AEs are not inherently orthogonal or ordered by explained variance. As a result, latent dimensions can be entangled, making it difficult to attribute specific anatomical meaning to individual dimensions without additional constraints (Chen et al.).

To address these challenges, recent research (Reuter et al.; Reuter et al.; Lemeunier et al.; Lemeunier et al.; Biffi et al.) has explored spectral representations of 3D surfaces as a compact lternative to spatial coordinates. By decomposing a surface into frequency coefficients of the Laplace-Beltrami operator, shapes can be described through intrinsic, ordered features, where low frequencies capture global form and higher modes encode fine details. Such a decomposition facilitates more efficient learning while maintaining geometric interpretability. The applicability of spectral decomposition to deep learning was demonstrated by Lemeunier et al. (2022), who trained an autoencoder on truncated Laplacian coefficients and showed that networks can learn geometry in the spectral domain. However, their approach requires all training meshes to share identical connectivity, making it unsuitable for clinical datasets with heterogeneous mesh connectivity and varying vertex counts. A similar requirement for known point-wise correspondences is present in recent point-based diffusion models, such as that of Zhu et al. (2025). In earlier preliminary work (Kubík et al.), we explored this limitation by introducing spectral synchronization, which aligns the spectral embeddings of all shapes to a common reference basis. This alignment removes the bias introduced by eigenbasis instability and enables modeling of a shared spectral manifold across arbitrary connectivities.

1.1 Contributions

The present work develops a complete and rigorously evaluated framework for dental shape generation based on synchronized spectral embeddings. In summary, the major contributions of our work are as follows:

- A deep generative framework for dental shape analysis operating on compact, synchronized spectral coefficients trainable on datasets with variable triangulations.
- Evidence that spectral synchronization is essential for achieving high-fidelity reconstructions and that latent regularization performed entirely in the spectral domain is sufficient for effective learning. Optimization in spatial coordinates provides no significant benefit, reinforcing the hypothesis that intrinsic spectral representations are sufficient to model anatomical variability.
- A comprehensive experimental evaluation demonstrating the robustness of spectral synchronization with respect to the reference shape, number of modes, and tooth class. The results confirm that the proposed framework generalizes across different tooth families and remains effective under varying spectral truncations, enabling modeling at multiple levels of geometric detail.
- Direct benchmarks against PCA-based statistical shape models and point-based VAEs and diffusion frameworks, supported by a thorough discussion of the observed results and implications for future extensions.

This study builds upon and significantly extends the preliminary version presented at the 29th International Conference on Information Processing in Medical Imaging, IPMI 2025 (Kubík et al.). Beyond expanding the experimental design across multiple tooth classes and truncation levels, the present work adds ablation analyses isolating the roles of synchronization and latent regularization, and introduces comprehensive comparative evaluations against established baselines. Together, these contributions advance well beyond the preliminary study, establishing a rigorous understanding of the behavior, clinical motivation, and more details of the evaluated methods. Access the codebase here: https://github.com/tiborkubik/toothForge.

2. Materials and Methods

This study benchmarks four approaches to generative modeling of dental shapes, concentrating on individual crown geometries without incorporating root structures or anatomical context. A classical PCA-based statistical shape model provides a linear baseline that is compact and interpretable but limited in expressiveness. Second, we consider spatial deep generative models operating directly on point clouds, which overcome linearity by exploiting nonlinear neural mappings but remain constrained by the unordered and extrinsic nature of point sets. Finally, we describe *ToothForge*, a spectral generative framework that leverages intrinsic geometry to achieve compact, stable, and connectivity-independent modeling. The following subsections introduce each method in detail, followed by the dataset and evaluation metrics description.

2.1 PCA: Statistical Shape Model

Principal component analysis (PCA) provides the foundation of classical statistical shape modeling (SSM). The main idea is to capture population variability by projecting shapes onto a low-dimensional linear subspace spanned by orthogonal modes of variation. SSM is constructed by first bringing all meshes of a given class into dense correspondence, ensuring that each vertex across the population encodes the same anatomical landmark. Each shape with n vertices is then represented as a column vector $x \in \mathbb{R}^{3n}$ obtained by concatenating its x, y, z coordinates. Collecting m such shapes yields a data matrix $X = [x_1, \ldots, x_m] \in \mathbb{R}^{3n \times m}$, which is centered by subtracting the mean shape

$$\bar{x} = \frac{1}{m} \sum_{j=1}^{m} x_j.$$

PCA is performed on the covariance matrix of X, yielding eigenvectors $U = [u_1, \ldots, u_r]$ and eigenvalues $\lambda_1 \geq \ldots \geq \lambda_r$, where u_i defines the i-th principal mode of deformation. Any shape x can be approximated in this basis as

$$x \approx \bar{x} + U_r b,$$

where $U_r \in \mathbb{R}^{3n \times r}$ contains the leading r eigenvectors and $b \in \mathbb{R}^r$ are the shape parameters. The variance explained by each mode is determined by λ_i, and the coefficients follow a Gaussian distribution $b_i \sim \mathcal{N}(0, \lambda_i)$.

New shapes can be synthesized by sampling coefficients from the Gaussian prior and reconstructing

$$x_{\text{new}} = \bar{x} + U_r b.$$

This linear model is compact, interpretable, and data-efficient. Each eigenvector corresponds to a distinct mode of anatomical variability, and varying b_i allows controlled exploration of the associated deformation.

However, this linear modeling assumption restricts the ability of the model to capture complex non-linear anatomical variations. In practice, it means that generated shapes are confined to lie within an affine subspace around the mean, which can lead to overly smooth or implausible reconstructions. This makes PCAs simple and interpretable, but less practical for representing rich anatomical details.

2.2 PointVAE and PointDiffusion: Spatial Deep Generative Models

While PCA-based statistical shape models provide an interpretable baseline, they are fundamentally limited by their *linearity*. Deep generative models, and in particular variational autoencoders (VAEs), offer a natural extension by learning a nonlinear mapping between data and a low-dimensional latent space. An encoder e_θ processes the input signal into a global representation, parameterizing a Gaussian distribution

$$e_\theta(x) \to \mu, \Sigma, \quad z \sim \mathcal{N}(\mu, \Sigma),$$

where $z \in \mathbb{R}^d$ denotes the latent code. A decoder d_γ then maps this latent vector back to the data domain, producing a reconstruction $\hat{x} = d_\gamma(z)$. The latent space is regularized to follow a prior distribution, which enables both faithful reconstructions and sampling of new shapes. Unlike PCA, the encoder and decoder are modeled by neural networks with non-linear activations, allowing them to capture complex and highly non-linear modes of anatomical variation.

Spatial deep generative models directly operate on vertex coordinates and connectivity information of 3D shapes. Among the various spatial representations explored in the literature, point clouds have emerged as a particularly prominent choice due to their flexibility and efficiency. A point cloud representation discards explicit connectivity information and instead models each mesh as an unordered set of points in $\mathbb{R}^3$. This eliminates the requirement of consistent

mesh connectivity across the dataset and allows a unified treatment of shapes with varying discretizations. Here, a 3D mesh is represented by a point cloud $S \subset \mathbb{R}^3$, obtained by uniformly sampling m points from the surface. The generative process is modeled using a variational autoencoder (VAE), which maps this set into a latent distribution and reconstructs it back into a set of points approximating the original surface. The encoder e_θ learns to project the input point cloud into the parameters of a Gaussian latent distribution,

$$e_\theta(S) \to \mu, \Sigma, \quad z \sim \mathcal{N}(\mu, \Sigma),$$

where $z \in \mathbb{R}^d$ denotes a d-dimensional latent vector. The decoder d_γ maps the latent code back into an unordered set of points,

$$\hat{S} = d_\gamma(z) \subset \mathbb{R}^3, \quad |\hat{S}| = m,$$

aiming to recover the original surface geometry.

As point clouds are unordered, architectures operating on them must be designed to be invariant to input permutations and extrinsic transformations such as translation, rotation, or scale. This invariance is typically enforced by building local neighborhoods through k-nearest-neighbor searches or ball queries, and then aggregating features within these neighborhoods using symmetric functions such as max- or mean-pooling. PointNet (Qi et al.) achieves this by applying shared multi-layer perceptrons to each point independently and using a global pooling operator to produce an order-invariant embedding. Such models scale poorly with input data resolution, which is crucial when analyzing geometries of anatomical shapes. More recent architectures such as Point Transformers (Wu et al.) extend this idea by learning attention weights over local neighborhoods defined through spatial proximity, which again requires explicit neighbor queries but allows for richer contextual modeling. The latest versions of point transformers improve scalability by fast point serialization. Despite these advances, such networks still operate directly on extrinsic coordinates. The geometry of the shape must therefore be inferred from the data, while invariance to global transformations has to be explicitly encoded in the architecture. The reconstruction loss is defined on unordered sets, most commonly using the Chamfer distance:

$$d_{\mathsf{CD}}(S, \hat{S}) = \sum_{x \in S} \min_{y \in \hat{S}} \|x - y\|_2^2 \; + \; \sum_{y \in \hat{S}} \min_{x \in S} \|x - y\|_2^2,$$

which measures the proximity of each point in one set to its nearest neighbor in the other. The full training objective combines this reconstruction loss with the Kullback–Leibler divergence that regularizes the latent space:

$$\mathcal{L} = d_{\mathsf{CD}}(S, \hat{S}) \; + \; \beta \, \mathsf{KL}\big(\mathcal{N}(\mu, \Sigma) \,\|\, \mathcal{N}(0, I)\big),$$

with β weighting the relative importance of regularization.

This framework allows for both reconstruction of input shapes and generation of novel samples. However, because point clouds are unordered, the network must repeatedly perform expensive neighborhood searches in high-dimensional feature spaces in order to capture local geometric relations. When large point clouds are required for anatomical fidelity, the receptive fields must expand proportionally, making training and inference computationally demanding and training is less stable with small datasets at hand. Thus, while point-based VAEs provide flexibility, their scalability to high-resolution shapes in low data regimes is limited.

Recently, diffusion probabilistic models have emerged as a strong alternative to VAEs for point cloud generation. In particular, Luo and Hu (2021) formulate a point cloud as a set of particles undergoing a gradual forward diffusion process that corrupts data into Gaussian noise, and then learn the corresponding reverse denoising dynamics to generate new shapes. Unlike VAEs, diffusion models do not require a single-shot mapping from the latent vector to a full point set. Instead, they refine an initial noise point cloud over many steps, which often improves sample fidelity and diversity in point cloud synthesis. The main drawback is computational as sampling requires executing the denoiser for multiple reverse steps. This makes diffusion-based generation substantially slower than a single decoder forward pass, and typically more time demanding in training and inference-time generation.

2.3 ToothForge: Spectral Deep Generative Model

ToothForge is a framework that utilizes spectral coefficients as input features generative deep modeling of dental shape variability. Instead of operating directly on vertex coordinates, each mesh is projected into the eigenspace of the Laplace–Beltrami operator, providing a compact, intrinsic description of surface geometry. See Figure 1a for a visual outline. Given a closed manifold mesh with vertices $V \in \mathbb{R}^{n \times 3}$, its truncated spectral coefficients are defined as

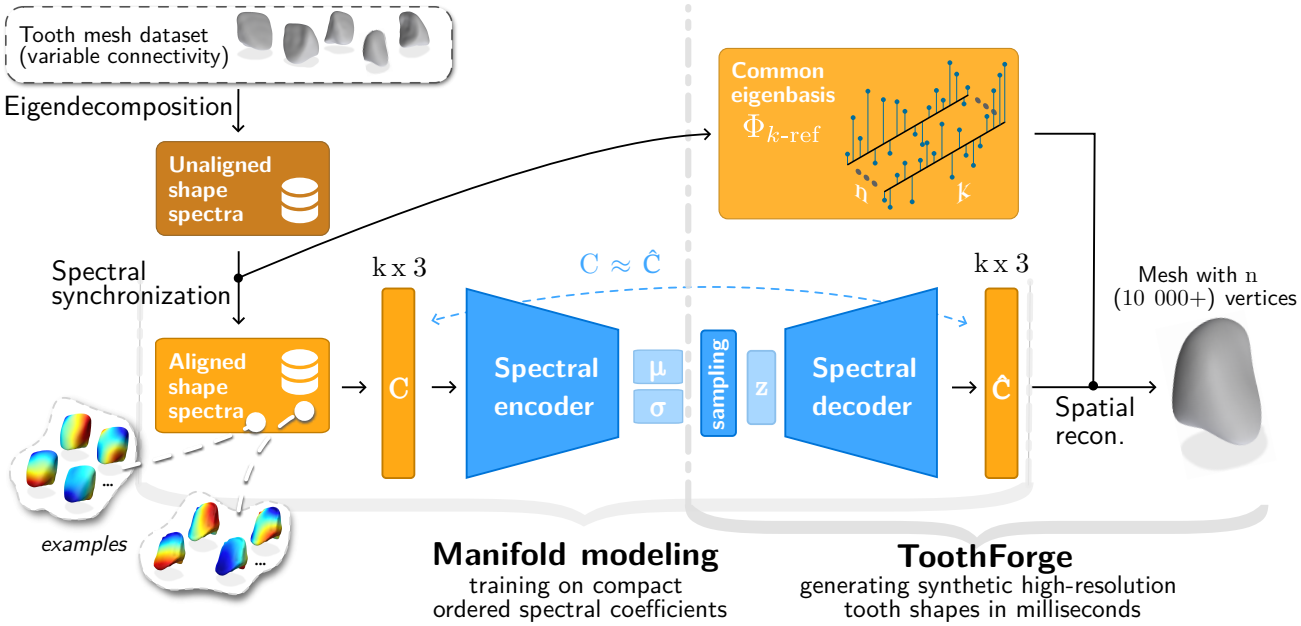
$$C_k = \Phi_k^\top V,$$

where $\Phi_k \in \mathbb{R}^{n \times k}$ contains the first k eigenvectors of the Laplacian, ordered by their associated eigenvalues. Lower frequencies capture coarse anatomical structure, such as crown length, while higher frequencies encode fine details, such as molar cusp morphologies, exhibiting rapid oscillations.

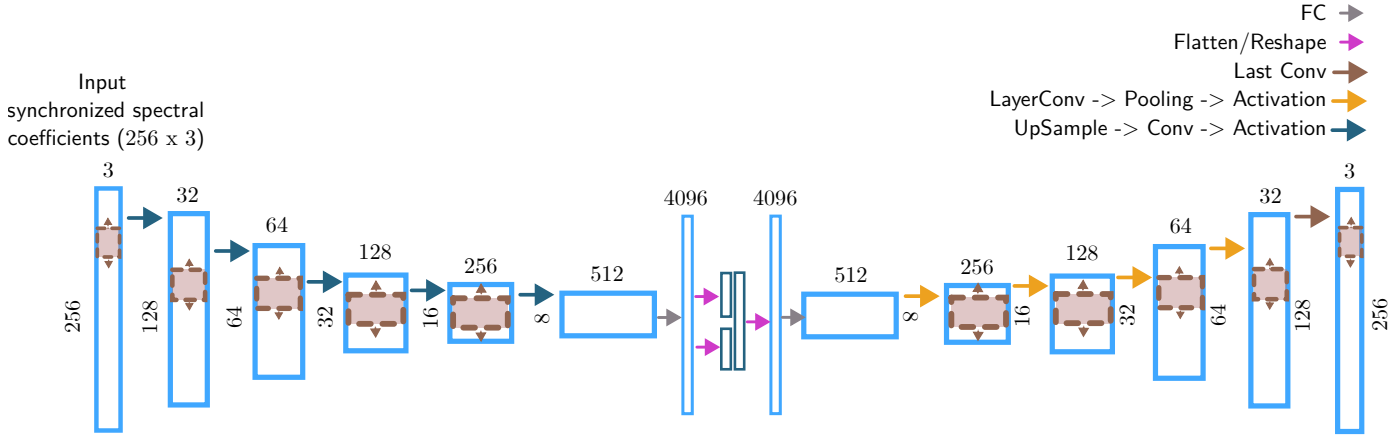
A shape can be approximated by reconstruction from these coefficients as

$$V_k = \Phi_k C_k,$$

with increasing k yielding higher-fidelity reconstructions. This natural ordering of coefficients provides a canonical



(a) Our framework for generating synthetic shapes of teeth via sampling on a latent manifold. Such manifold is modeled using synchronized spectral coefficients of tooth shapes, denoted as C. For novel data sampling using *ToothForge*, two ingredients are necessary: decoder weights for infering novel modal coefficient $\hat{C}$ and common eigenbasis $\Phi_{k,\text{ref}}$ to project it to spatial domain.



(b) The framework adopts a five-stage 1D convolutional architecture with learnable pooling and upsampling (Lemeunier et al.).

Figure 1: Outline and architecture details of ToothForge framework.

axis along which neighboring frequencies can be pooled, enabling convolutional and hierarchical operations that are not readily applicable in the unordered spatial domain.

Despite the natural ordering advantage, spectral representations also introduce specific challenges. Eigenfunctions are not uniquely defined and may flip signs or swap order across shapes due to instabilities of the decomposition. ToothForge addresses this by introducing *spectral synchronization*, aligning all spectral embeddings to a common reference basis $\Phi_{k,\text{ref}}$. For each shape M_i with coefficients $C_{k,i}$, a transformation $R_i \in \mathbb{R}^{k\times k}$ is estimated such that

$$\tilde{C}_{k,i} = R_i C_{k,i},$$

minimizing the discrepancy between the aligned coefficients and the reference basis. This procedure eliminates instabilities of the decomposition and provides consistent, connectivity-independent spectral features across the dataset. The reference template is chosen from the training set and defines the common spectral basis $\Phi_{k,\text{ref}}$ used for synchronization and for projecting decoded coefficients back to the spatial domain. Importantly, synchronization aligns the eigenbases (i.e., the spectral coordinate system) rather than deforming shapes toward the reference geometry, so the template does not act as a geometric prior. In practice, template can be selected either at random from the training set or as a representative *medoid* shape minimizing average distance to other training samples.

The aligned spectral coefficients are modeled with a β-VAE. The encoder e_θ maps coefficients into a Gaussian

latent distribution

$$e_\theta(\tilde{C}_k) \to \mu, \Sigma, \quad z \sim \mathcal{N}(\mu, \Sigma),$$

and the decoder d_γ reconstructs coefficients $\hat{C}_k = d_\gamma(z)$. The training objective is a weighted sum of reconstruction error in spectral space and a KL divergence term:

$$\mathcal{L} = \|\tilde{C}_k - \hat{C}_k\|_2^2 + \beta\, \mathrm{KL}(\mathcal{N}(\mu, \Sigma) \,\|\, \mathcal{N}(0, I)).$$

Although projecting coefficients back into the spatial domain is computationally inexpensive, the optimization is carried out entirely in the spectral domain. This reflects a fundamental distinction between *extrinsic* and *intrinsic* learning. Spatial generative models must operate on raw coordinates in $\mathbb{R}^3$, and therefore spend capacity accounting for extrinsic factors such as global scale, orientation, translation, and the permutation of input points. In contrast, spectral coefficients are intrinsic to the surface and already embed the geometry of the mesh into the representation. The network does not need to learn the underlying geometry from data, but can instead focus directly on modeling anatomical variability in a compact and stable space.

After training, novel shapes are generated by sampling latent vectors $z \sim \mathcal{N}(0, I)$, decoding them into spectral coefficients $\hat{C}_k$, and projecting back to the spatial domain through the common basis:

$$\hat{V} = \Phi_{k,\mathrm{ref}}\, \hat{C}_k.$$

Coupled with the template shape connectivity information, this produces high-resolution meshes with consistent correspondence across samples, as visualized in Figure 2.

ToothForge thereby combines the compactness of spectral embeddings with the generative power of variational autoencoders. Compared to point-based VAEs operating on unordered coordinates, ToothForge explicitly leverages intrinsic geometric structure via synchronized spectral coefficients, which is particularly advantageous in bandwidth-matched, low-information regimes.

We adopt a β-VAE as the generative backbone because ToothForge is designed to learn a compact, low-dimensional latent manifold and to support efficient sampling for downstream use (e.g., augmentation). Diffusion models are a promising alternative for high-fidelity synthesis, but they typically require iterative sampling and substantially higher training/sampling cost, and are often most effective when trained with larger datasets. We therefore focus on VAEs in this work, and leave diffusion-based generation in the spectral domain as an interesting direction for future research.

2.4 Data

All experiments were performed on a private dataset of 430 dental crown shapes: 149 incisors, 161 premolars and 120 molars. Data was provided by an industrial partner. The dataset was divided into three anatomical categories: incisors, premolars, and molars. For each category, an independent model was trained under all evaluated frameworks. We train separate models per crown category to maintain anatomically homogeneous shape distributions and to enable a direct, class-wise comparison to standard PCA baselines, which are commonly constructed per anatomical class.

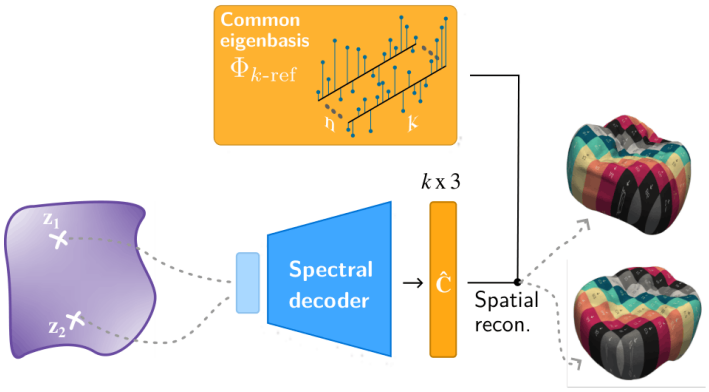


Figure 2: ToothForge samples new shapes by generating spectral coefficients and reconstructing them through the shared reference eigenbasis. When combined with the template mesh connectivity, all synthesized meshes inherit the same vertex indexing, yielding guaranteed point-to-point correspondence across samples.

All crowns were digitally modeled by experienced dental technicians with the specific purpose of serving as patient-specific dental prostheses. This ensures clinically realistic anatomical morphology while avoiding biases such as tooth wear or surface irregularities that would be present if shapes were segmented from patient optical scans.

All dental crowns are represented as triangular manifold meshes with various triangulations. No texture or color information is included, and the average resolution is 11484 ± 535 vertices. Since the crowns were digitally modeled for use as dental prostheses, the meshes are inherently watertight and free of holes. This guarantees that spectral decompositions and correspondence procedures are not affected by artifacts such as holes, self-intersections, or non-manifold edges, ensuring stable downstream processing.

2.5 Evaluation Metrics

All reported geometric distances and reconstruction errors are measured and reported in millimeters (mm).

2.5.1 Accuracy

Accuracy evaluates how well the model can reconstruct shapes from the *training set* after compression. When two representations are ordered, such as spatial point clouds

with point-to-point correspondence, accuracy is measured using the mean squared error (MSE):

$$d_{\text{MSE}}(C_1, C_2) = \|C_1 - C_2\|_2^2,$$

where $C_1, C_2 \subseteq \mathbb{R}^3$ denote the ordered sets. For unordered vertex sets, $S_1 \subseteq \mathbb{R}^3$ and $S_2 \subseteq \mathbb{R}^3$, we use the Chamfer distance (CD), defined as

$$d_{\text{CD}}(S_1, S_2) = \sum_{x \in S_1} \min_{y \in S_2} \|x - y\|_2^2 + \sum_{y \in S_2} \min_{x \in S_1} \|x - y\|_2^2.$$

Lower values of d_{MSE} and d_{CD} indicate more faithful reconstructions of the training samples. The overall accuracy of a model, denoted $d_{\text{trn-MSE}}$ or $d_{\text{trn-CD}}$, is obtained as the average of the corresponding measure across all shapes in the training set.

2.5.2 Generalization

Generalization assesses how well the model can reconstruct *unseen test samples*, thereby measuring its ability to extend beyond the training distribution. For point sets with consistent ordering, we compute reconstruction error using the mean squared error (MSE). For unordered vertex sets, we instead employ the Chamfer distance (CD), as defined in the previous section. Smaller errors correspond to reconstructions that remain faithful to the anatomical details of held-out shapes. The reported generalization score of a model, denoted $d_{\text{tst-MSE}}$ or $d_{\text{tst-CD}}$, is obtained by averaging the chosen distance measure across all shapes in the test set.

2.5.3 Minimum Matching Distance

While accuracy and generalization quantify how faithfully a model can reconstruct individual shapes after compression, they do not evaluate the quality of the model as a generative distribution. To assess the distributional fidelity of the learned shape space, we evaluate the Minimum Matching Distance (MMD). MMD evaluates how well the learned generative model captures the fidelity of the training samples using a nearest-neighbor matching score. Concretely, we draw random latent codes from the prior distribution, decode them into shapes, and obtain a generated set S_g with $|S_g| = |S_r|$, where S_r is the reference set of real shapes. We measure the distance of S_g to the reference set S_r of real shapes. For each shape in S_r, given a set of real crowns S_r and a set of generated crowns S_g, we compute the matching error

$$\text{MMD}(S_r, S_g) = \frac{1}{|S_r|} \sum_{r \in S_r} \min_{g \in S_g} d_{CD}(r, g).$$

Low values indicate that generated samples are on-manifold and do not deviate into unrealistic shapes.

2.5.4 Coverage

To complement MMD, we additionally measure the diversity as the fraction of reference samples that are selected as the nearest neighbor of at least one generated sample:

$$\text{COV}(S_r, S_g) = \frac{|\{\arg\min_{r \in S_r} d_{CD}(r, g) | g \in S_g\}|}{|S_r|}.$$

Higher values indicate that the generated set covers a larger portion of the reference set (Naeem et al.).

3. Results

3.1 Experimental Setup

Dense correspondence was first established among all meshes to enable construction of the PCA-based statistical shape model. A single template was selected, simplified to 4096 vertices, and normalized to unit scale. For each remaining mesh, 4096 surface points were sampled and aligned to the template using rigid ICP Rusinkiewicz and Levoy (2001), followed by non-rigid coherent point drift from Myronenko and Song (2010) with parameters $\beta = 2.0$, $\lambda = 3.0$, and 150 iterations. This produced a set of meshes sharing the same template connectivity. The registered meshes were vectorized by concatenating their x, y, z coordinates and stacked into a shape matrix. Shapes were centered and normalized before PCA was applied with k retained components. Novel shapes were generated by sampling coefficients $b_i \sim \mathcal{N}(0, \lambda_i)$ and forming linear combinations of principal components with the mean shape.

As nonlinear spatial baselines, a point-based variational autoencoder (PointVAE) was used, similar to what is presented in Gutiérrez-Becker et al. (2021). The encoder and decoder was replaced by PointNet++ (Qi et al.) and implemented with architecture matched in depth to ToothForge. The network follows a 5-stage encoder-decoder design with a Gaussian bottleneck of dimension $d = 16$, optimized using Chamfer distance on vertex positions combined with a β-weighted KL divergence term. A cyclical annealing of β between 0 and 0.05 was applied. Input point clouds were sampled using Poisson disk sampling with m points per mesh. Training was performed with batch size 16 using AdamW (initial learning rate 10^{-4}, cosine annealing restarts every 10000 iterations) for approximately two hours on a single Tesla T4 GPU. Second point-based framework is a point diffusion model presented by Luo and Hu. More specifically, we employ their point cloud generator model with normalizing flows. We keep most of the hyperparameters as proposed defaults, and only change the latent dimensionality to 64. Input point clouds were sampled using Poisson disk sampling with m points per mesh. We optimized the model for approximately two hours using single Tesla T4

Table 1: Reconstruction and generative quality across spectral truncations k of *ToothForge*. *SR* (shape recovery) is the average truncation error of the k-harmonic reconstruction of the ground-truth meshes for given class and k combination computed between ground-truth vertices V and truncated reconstruction $V_k = \Phi_k \Phi_k^\top V$. Relative errors ($^{\text{rel}}$) compare predictions to the k-truncated target, absolute errors ($^{\text{abs}}$) compare to the full-resolution ground truth. All values are measured with spatial reconstructions using CD as the distance metric.

	Incisors					**Premolars**					**Molars**				
k	SR	$d^{\text{rel}}_{\text{trn}}$	$d^{\text{abs}}_{\text{trn}}$	$d^{\text{rel}}_{\text{tst}}$	$d^{\text{abs}}_{\text{tst}}$	SR	$d^{\text{rel}}_{\text{trn}}$	$d^{\text{abs}}_{\text{trn}}$	$d^{\text{rel}}_{\text{tst}}$	$d^{\text{abs}}_{\text{tst}}$	SR	$d^{\text{rel}}_{\text{trn}}$	$d^{\text{abs}}_{\text{trn}}$	$d^{\text{rel}}_{\text{tst}}$	$d^{\text{abs}}_{\text{tst}}$
32	0.232	0.022	0.241	0.061	0.234	0.221	0.031	0.248	0.061	0.243	0.244	0.033	0.250	0.075	0.246
64	0.110	0.041	0.152	0.066	0.155	0.101	0.027	0.127	0.038	0.129	0.123	0.024	0.159	0.038	0.155
128	0.083	0.046	0.123	0.060	0.128	0.066	0.030	0.079	0.037	0.102	0.096	0.033	0.093	0.033	0.108
256	0.069	0.043	0.117	0.059	0.122	0.063	0.028	0.075	0.049	0.091	0.089	0.029	0.104	0.037	0.092
512	0.063	0.059	0.115	0.067	0.114	0.061	0.043	0.081	0.054	0.102	0.084	0.030	0.099	0.043	0.093
1024	0.062	0.055	0.129	0.060	0.117	0.060	0.049	0.076	0.063	0.100	0.084	0.033	0.094	0.071	0.101

GPU. We refer to this model as PointDiffusion in the text. To encourage robustness in both spatial approaches, small random perturbations in global rotation and isotropic scale were applied to the input point clouds during training. Separate models were trained for incisors, premolars, and molars, using the same 80/20 train–test split as with PCA.

ToothForge operates in the spectral domain, where each mesh is projected into the k-truncated eigenbasis of the Laplace–Beltrami operator. Spectral embeddings are synchronized to a common reference basis using the method of Lombaert et al. (2015), ensuring stable and consistent coefficients across shapes. A β-VAE with a 5-stage 1D convolutional encoder–decoder, latent size $d = 16$, and cyclical annealing of β between 0 and 0.05 was trained. The architecture uses pooling and unpooling operators as presented by Lemeunier et al. (2022). See Figure 1b for details in scenario where 256 modes are used during training. The reconstruction loss was defined purely in the spectral domain, combining mean squared error on coefficients with the KL divergence term. Training was performed with batch size 16 using AdamW (initial learning rate 10^{-4}, cosine annealing restarts every 10,000 iterations) for approximately two hours on a Tesla T4 GPU. Separate models were trained for incisors, premolars, and molars using the same 80/20 train–test split as for the other baselines.

3.2 Results on ToothForge

3.2.1 Reconstruction Quality Across Tooth Classes and Spectral Truncations

A central design choice in spectral generative modeling is the number of Laplace–Beltrami eigenfunctions retained when projecting a mesh into the spectral domain. Lower values of k yield a more compact representation but may discard high-frequency information, while larger values preserve fine geometric detail at the cost of higher dimensionality. In practice, k is selected based on the target trade-off between compactness and geometric detail, e.g., by choosing the smallest k for which reconstruction/coverage metrics (and the corresponding qualitative fidelity) have saturated for the intended application. To assess how this trade-off influences ToothForge, reconstruction quality was quantified across different truncation levels.

Reconstruction errors are reported in two complementary ways. First, predictions are compared against the k-truncated meshes used to generate the training coefficients. This relative error isolates the fidelity of the network. Second, reconstructions are compared against the original high-resolution ground-truth meshes. This absolute error captures both truncation and network reconstruction losses, reflecting the overall clinical fidelity of the model. The results are summarized in Table 1.

The relative reconstruction errors indicate that the model generalizes well to unseen samples across tooth classes. As the number of spectral harmonics increases, the reconstruction quality remains stable, with only a slight rise in the relative training error. This trend suggests that while the model effectively captures the global geometric structure, it becomes increasingly sensitive to the higher-frequency spectral components. These components are inherently less stable and more susceptible to noise. Consequently, the spectral representation of fine geometric details may exhibit minor irregularities, which can challenge the network's ability to model them consistently. The absolute reconstruction errors remain consistently low across all tooth classes, reflecting high fidelity to the full-resolution ground truth. Their gradual reduction with increasing k indicates that additional spectral components improve geometric detail recovery without introducing instability in the spatial domain. Although minor noise appears at higher k values, the absolute errors continue to approach the SR values. This indicates that most of the remaining discrepancy originates from the intrinsic low-pass filtering rather than from network reconstruction errors. See Figure 5 for

Table 2: Ablation at k=256 for molar class. All values are measured with spatial reconstructions using CD as the distance metric. **Bold** values denote the best-performing results and underlined values indicate the second-best within each column.

Variant	$d_{\text{trn}}^{\text{rel}}$	$d_{\text{tst}}^{\text{rel}}$	$d_{\text{tst}}^{\text{abs}}$	MMD
ToothForge	0.029	**0.037**	**0.092**	0.099
$\beta = 0$	0.066	0.125	0.232	0.212
No spec. sync.	0.101	0.127	0.160	0.186
$+\lambda_{\text{sp}}{=}1\times10^{-2}$	0.027	0.041	0.105	0.111
$+\lambda_{\text{sp}}{=}5\times10^{-2}$	**0.022**	0.040	0.112	0.102
$+\lambda_{\text{sp}}{=}1\times10^{-3}$	0.026	0.044	0.108	**0.098**

randomly selected tooth reconstructions.

3.2.2 Ablation Study

A quantitative ablation study quantified the effect of individual design choices, with results summarized in Table 2. The parameter k was fixed at 256, and analysis focused on the molar class, noting that similar trends hold across other truncation values and tooth classes. Setting $\beta = 0$ eliminated latent regularization and led to reconstructions that collapsed toward overly smooth mean-shape-like geometries, with little variability across latent space exploration. Removing spectral synchronization consistently degraded reconstructions: in the spatial domain, generated outputs were often semantically far from dental structures, producing implausible shapes. As an additional test, the training loss was modified to include a spatial regularization term, implemented as an auxiliary Chamfer distance on reconstructed vertices,

$$\mathcal{L} = \|\tilde{C}_k - \hat{C}_k\|_2^2 + \beta\,\mathrm{KL}\Big(\mathcal{N}(\mu,\Sigma)\,\|\,\mathcal{N}(0,I)\Big) + \lambda_{\text{sp}}\, d_{\text{CD}}(V,\hat{V}),$$

where V and $\hat{V}$ denote the ground-truth and reconstructed vertex sets, respectively, and $d_{\text{CD}}(\cdot,\cdot)$ is the Chamfer distance as defined earlier. Across all tooth classes, this additional spatial term did not yield consistent improvements and in several cases slightly worsened results. Overall differences were neglectable. These findings support that the intrinsic smoothness encoded by the spectral representation is sufficient, and that adding explicit spatial regularization does not benefit ToothForge.

3.2.3 Robustness to Template Shape Selection

The robustness of ToothForge to the choice of reference template was evaluated by training models on different randomly selected template shapes and testing them across all possible train–test template combinations. The results are presented in Figure 4. The diagonal entries, where training and test data were synchronized to the same template, consistently produced the lowest reconstruction errors and plausible reconstructions. In contrast, the off-diagonal cases, where different templates were used for training and testing, showed that the generated shapes were less accurate, underlining that mismatched bases introduce variability into the predictions. While this degradation was not as severe as in the "NoAlign" setting, the results indicate that stable performance is primarily achieved when train and test shapes are synchronized to the same reference template. We observed similar trends across tooth classes and k values. Importantly, this experiment indicates that the reference template does not bias the generated geometry toward a particular shape. Instead, it establishes a common spectral basis used to express coefficients consistently. The main failure mode arises from using different bases between training and inference, rather than from any particular choice of template. In practice, reference can be chosen from the training set as a representative shape (e.g., a medoid) or at random, and then kept fixed throughout training and generation to ensure a stable synchronization convention.

3.3 Comparative Analysis

First, quantitative comparisons are provided in Figure 3 and Table 3. ToothForge consistently achieves low minimal matching distance values on molar class. Same trend was observed on incisors and premolars. For experiments with low truncations ($k = 32, 64$), the limited number of coefficients constrains the representation capacity, leading to reconstructions that are coarse and not anatomically plausible. PointVAE and PointDiffsion show higher error in low k setup because its inputs are sparse point clouds (matched in bandwidth to the coefficient dimensionality), providing substantially less geometric information for reconstruction than the spectral representation. For higher k values, both spectral and point-based methods capture tooth shape variability well. PointVAE, however, often produces locally noisy surfaces. PointDiff exhibits strong generative behavior, but sampling is substantially slower (milliseconds for a single forward pass of ToothForge vs. seconds for iterative diffusion), which limits practical throughput. Moreover, these point-based frameworks produce unordered point sets without connectivity, so a separate surface reconstruction step (e.g., Poisson reconstruction) is required. When applied to sparse point clouds, this post-processing can yield implausible or anatomically distorted meshes. In contrast, eventhough spectral reconstructions tend to be slightly smoothed, they directly yield coherent surfaces, which better preserves anatomical fidelity and results in more plausible crown geometries. The difference in reconstruction quality is visualized in Figure 6. As for the comparison with PCA, both PCA and ToothForge capture global crown characteristics well. However, PCA-generated samples show reduced

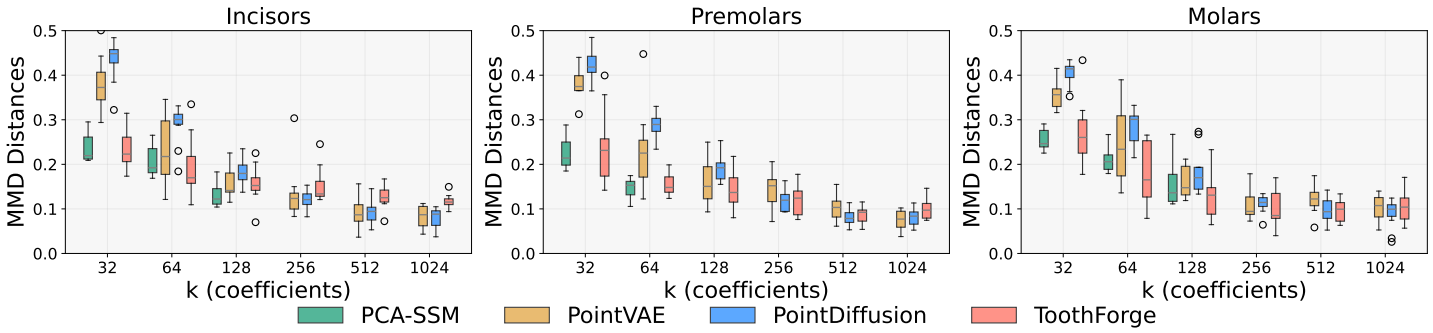


Figure 3: Reconstruction fidelity across spectral truncations k for molars using PCA, PointVAE, PointDiffusion and ToothForge. Note that PCA results are not reported for higher k values due to the insufficient number of available samples to support the corresponding number of principal components. For the PointVAE and PointDiffusion, k denotes the number of input spatial points (x, y, z).

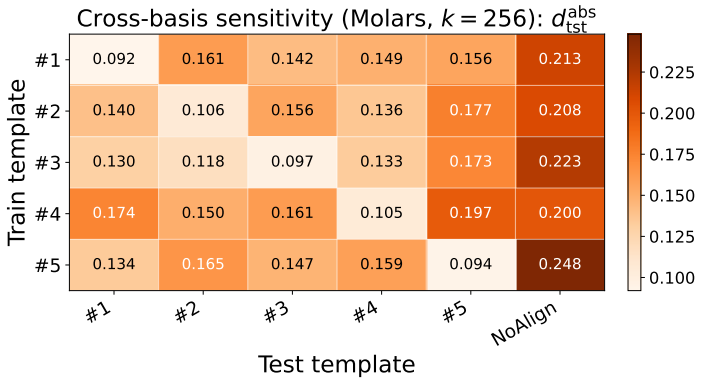


Figure 4: Cross-basis sensitivity for the molar class at $k = 256$. Heatmap entries show the absolute test reconstruction error $d_{\text{tst}}^{\text{abs}}$ when training with one reference template (rows) and evaluating test shapes aligned to another template (columns). Diagonal entries correspond to matched train–test templates and yield the lowest errors. Off-diagonal entries show modest degradation when train and test templates differ. The "NoAlign" column indicates performance without spectral synchronization, resulting in substantially higher errors.

diversity in localized anatomical features, most notably in fine-scale regions such as lingual/buccal grooves, whereas ToothForge yields a broader range of such morphological variations.

Jointly, the quantitative and qualitative results show at least two things. First, in the context of population-level tooth modeling, the benefits of non-linear latent spaces are not markedly pronounced compared to classical statistical shape models such as PCA. Nevertheless, our findings demonstrate that non-linear spectral modeling can achieve comparable reconstruction fidelity while leveraging a compact, frequency-ordered intrinsic coefficient representation that is stable under changes in discretization through synchronization. ToothForge concentrates learning on a fixed-length intrinsic descriptor whose dimensionality is decoupled from mesh resolution, while remaining compatible with future extensions such as conditional generative modeling (Zhang et al.). This could enable more practical frameworks, for example by learning a single latent space for all tooth classes or by introducing attribute-guided shape synthesis, text-driven generating (Xu et al.), or geometry modifying (Achlioptas et al.). We also show that the distributional coverage and fidelity is on par with point-based frameworks, but without its limitations in representing fine anatomical details in low-data settings. While even more advanced point-based generative methods for dental structures (Chanintonsongkhla et al.) can capture geometric complexity through point-flow mechanisms (Yang et al.), they require extensive optimization (30 days for anterior teeth) and large datasets. This compactness is therefore a key distinction from spatial generative models, which typically operate directly on high-dimensional, unordered coordinates and therefore require substantially more model capacity, computational time, and bigger datasets to learn invariances and local structure.

3.4 Downstream Task Experiment

To assess whether ToothForge is useful beyond shape reconstruction, we evaluate it as a data augmentor in a tooth class classification (incisor/premolar/molar). Because this task is relatively easy on the full dataset, we deliberately restrict the classifier training set to 100 randomly selected shapes while preserving the original class distribution, making the setting data-limited and thus more sensitive to augmentation. To prevent any data contamination, ToothForge is trained only on the remaining training split, excluding the 100 shapes used to train the classifier. The test set is identical for generative and classification task.

We compare setups with no augmentation, conventional mesh-space augmentations only, ToothForge samples only,

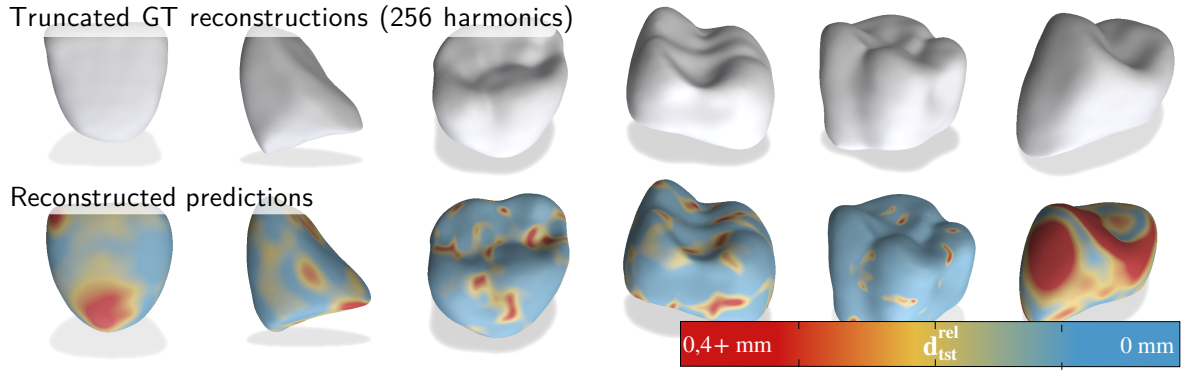


Figure 5: Reconstructions of unseen tooth shapes. The predictions accurately capture the overall tooth shape. The reconstructions may appear smoothed at times (rightmost premolar). This is due to possible inaccuracies in predicting high-frequency components.

Table 3: Quantitative comparison and runtime of the evaluated generative methods on the molar tooth class. We report Minimum Matching Distance (MMD, ↓) and Coverage (COV, ↑) computed using Chamfer Distance (CD), together with training and sampling time. PCA uses the top 128 input features, while deep learning baselines use 256 features. Best results are highlighted in **bold**, and second-best results are underlined.

Method	MMD	COV (%)	Training Time	Sampling Time
PCA	0.1541	37.04	∼ **10 s**	∼ 10 ms
PointVAE	0.1090	42.63	∼ 121 mins	∼ 22 ms
PointDiffusion	0.1112	**47.00**	∼ 143 mins	∼ 9 s
ToothForge	**0.0997**	43.78	∼ 100 mins	∼ **1 ms**

and both combined. For conventional augmentations, we apply a rigid transform, random rotation around the z-axis up to $45°$ and, a deformation augmentation, vertex jittering with Gaussian noise ($\sigma = 0.005$). For ToothForge augmentation, wer consider two budgets 100 and 1000 generated shapes added to the training set.

Table 4 shows that ToothForge can serve as an effective augmentor in this low-data setting. Training without augmentation yields 90.36% accuracy, while conventional mesh-space augmentations improve this to 95.18%. Using 1000 ToothForge samples leads to a larger gain, reaching 98.79%, and adding mesh transformations on top does not further change performance, indicating near-ceiling accuracy on this simple task and relatively small test set.

To better probe complementarity before saturation, we also evaluate a smaller generative budget. With only 100 ToothForge samples, accuracy increases to 92.77%, i.e., better than no augmentation but below conventional augmentations. Importantly, combining these 100 generated samples with mesh transformations improves accuracy to 96.38%, outperforming mesh transformations alone (95.18%). This suggests that ToothForge provides additional, non-trivial shape variability that complements standard rigid and deformation augmentations.

Table 4: Downstream tooth type classification (incisor/premolar/molar) under different augmentation strategies in the low-data regime (100 training shapes, class-balanced). Mesh transforms denotes random z-axis rotations and vertex jitter. ToothForge augmentation adds 100 or 1000 generated crowns sampled from the learned manifold. Accuracy is reported on a fixed held-out test set. Best results are highlighted in **bold**, and second-best results are underlined.

Variant	Accuracy (%)
No augmentation	90.36
Mesh transforms (rotation + vertex jitter)	95.18
ToothForge ($N = 1000$)	**98.79**
Mesh transforms + ToothForge ($N = 1000$)	**98.79**
Mesh transforms (rotation + vertex jitter)	95.18
ToothForge ($N = 100$)	92.77
Mesh transforms + ToothForge ($N = 100$)	**96.38**

4. Conclusion

This study presented ToothForge, a deep generative framework that learns dental crown geometry from synchronized spectral representations. By operating in the spectral domain, the method models intrinsic shape variability through compact and interpretable latent spaces that remain stable across heterogeneous mesh connectivities. Comprehensive

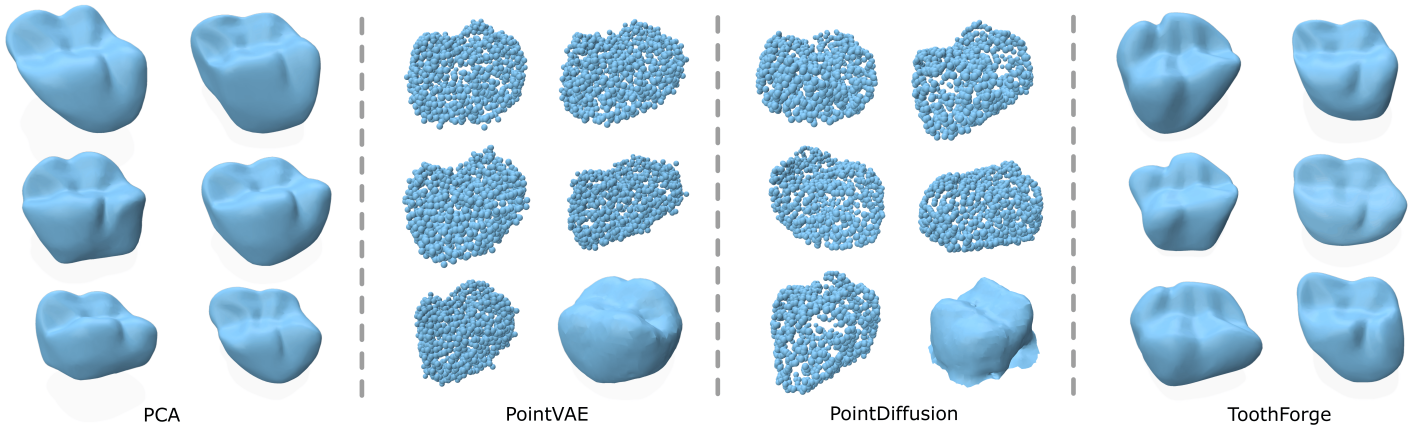


Figure 6: Samples from the evaluated methods. For PCA and ToothForge, we visualize reconstructed meshes directly using the shared template connectivity and per-vertex correspondence. For point-based methods, we display raw point clouds, with one representative sample additionally reconstructed via Poisson surface reconstruction. Note that while both PCA and ToothForge capture global crown characteristics well, PCA-generated samples exhibit reduced diversity in localized anatomical features like lingual/buccal grooves, whereas ToothForge yields a broader range of fine-scale morphological variations. Point-based samples tend to show more pronounced local surface noise.

experiments across incisor, premolar, and molar classes demonstrated consistent reconstruction accuracy, valided across varying number of spectral modes used during training. Our findings indicate that synchronized spectral modeling exhibits robustness to reference selection, as long as a consistent synchronization framework is maintained between training and testing. Compared to PCA and point-based generative models, ToothForge provides a good balance of compactness, fast sampling, and strong generative performance. Finally, we showed that ToothForge is effective as a generative data augmentor, improving performance in a low-data downstream tooth-type classification task and complementing standard mesh-space augmentations. Collectively, these findings validate the efficacy of spectral embedding methodologies for medical shape generation tasks, particularly under the constraints of limited data availability and high-resolution representation characteristic of digital dentistry applications. We emphasize that the main value of ToothForge is not solely improved reconstruction over PCAs under ideal correspondence, but a practical generative representation for clinical datasets where consistent connectivity and discretization cannot be guaranteed. By combining spectral synchronization with compact, frequency-ordered coefficients, ToothForge preserves PCA-like efficiency while enabling non-linear generative modeling in a space that is readily extensible. This creates a direct path toward (i) unified latent space that can model multiple classes like incisors, premolars, and molars jointly, (ii) conditional generation where we can steer synthesis using explicit attributes like target crown size or cusp prominence), and (iii) controllable shape editing workflows that modify an existing crown using additional inputs such as partial geometry, occlusal constraints, or other multimodal clinical signals.

Ethical Standards

The work follows appropriate ethical standards in conducting research and writing the manuscript, following all applicable laws and regulations regarding treatment of animals or human subjects.

Conflicts of Interest

We declare we don't have conflicts of interest.

Data availability

The data used in this study consist of patient-specific dental crown prostheses provided by an industrial partner. Due to privacy regulations and data protection agreements, these data cannot be publicly shared.

References

Panos Achlioptas, Ian Huang, Minhyuk Sung, Sergey Tulyakov, and Leonidas Guibas. ShapeTalk: A language dataset and framework for 3d shape edits and deformations. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023.

Jadie Adams and Shireen Y. Elhabian. Can point cloud networks learn statistical shape models of anatomies? In *Medical Image Computing and Computer Assisted Intervention (MICCAI)*, 2023.

Monish Bhola, Anthony L Neely, and Shilpa Kolhatkar.

Immediate implant placement: clinical decisions, advantages, and disadvantages. *Journal of Prosthodontics: Implant, Esthetic and Reconstructive Dentistry*, 2008.

Carlo Biffi, Juan J. Cerrolaza, Giacomo Tarroni, Wenjia Bai, Antonio de Marvao, Ozan Oktay, Christian Ledig, Loic Le Folgoc, Konstantinos Kamnitsas, Georgia Doumou, Jinming Duan, Sanjay K. Prasad, Stuart A. Cook, Declan P. O'Regan, and Daniel Rueckert. Explainable anatomical shape analysis through deep hierarchical generative models. *IEEE Transactions on Medical Imaging*, 39(6):2088–2099, 2020. .

Chawalit Chanintonsongkhla, Varin Chouvatut, Chumphol Bunkhumpornpat, and Pornpat Theerasopon. A latent variable deep generative model for 3d anterior tooth shape. *Journal of Prosthodontics*, 2025.

Ricky TQ Chen, Xuechen Li, Roger B Grosse, and David K Duvenaud. Isolating sources of disentanglement in variational autoencoders. *Neural Information Processing Systems (NeurIPS)*, 2018.

Blaine M Cleghorn, William H Christie, and Cecilia CS Dong. Root and root canal morphology of the human permanent maxillary first molar: a literature review. *Journal of endodontics*, 2006.

Benjamín Gutiérrez-Becker, Ignacio Sarasua, and Christian Wachinger. Discriminative and generative models for anatomical shape analysis on point clouds with deep neural networks. *Medical Image Analysis (MedIA)*, 2021.

Golriz Hosseinimanesh, Farnoosh Ghadiri, Francois Guibault, Farida Cheriet, and Julia Keren. From mesh completion to ai designed crown. In *Medical Image Computing and Computer Assisted Intervention (MICCAI)*, 2023.

Golriz Hosseinimanesh, Ammar Alsheghri, Julia Keren, Farida Cheriet, and Francois Guibault. Personalized dental crown design: A point-to-mesh completion network. *Medical Image Analysis (MedIA)*, 2025.

Hyun-Jun Kong and Yu-Lee Kim. Application of artificial intelligence in dental crown prosthesis: a scoping review. *BMC Oral Health*, 2024.

Tibor Kubík, François Guibault, Michal Španěl, and Hervé Lombaert. Toothforge: Automatic dental shape generation using synchronized spectral embeddings. In *Information Processing in Medical Imaging (IPMI)*, 2026.

Tibor Kubík and Michal Španěl. Lmvsegrnn and poseidon3d: Addressing challenging teeth segmentation cases in 3d dental surface orthodontic scans. *Bioengineering*, 2024.

Clément Lemeunier, Florence Denis, Guillaume Lavoué, and Florent Dupont. Representation learning of 3d meshes using an autoencoder in the spectral domain. *Computers & Graphics*, 2022.

Clément Lemeunier, Florence Denis, Guillaume Lavoué, and Florent Dupont. Spectrhums: Spectral transformer for human mesh sequence learning. *Computers & Graphics*, 2023.

Herve Lombaert, Michael Arcaro, and Nicholas Ayache. Brain transfer: Spectral analysis of cortical surfaces and functional maps. In *Information Processing in Medical Imaging (IPMI)*, 2015.

Shitong Luo and Wei Hu. Diffusion probabilistic models for 3d point cloud generation. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021.

Yosuke Matsuyama, Mitsuru Motoyoshi, Niina Tsurumachi, and Noriyoshi Shimizu. Effects of palate depth, modified arm shape, and anchor screw on rapid maxillary expansion: a finite element analysis. *European Journal of Orthodontics*, 2015.

Andriy Myronenko and Xubo Song. Point set registration: Coherent point drift. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 2010.

Muhammad Ferjad Naeem, Seong Joon Oh, Youngjung Uh, Yunjey Choi, and Jaejun Yoo. Reliable fidelity and diversity metrics for generative models. In *International Conference on Machine Learning (ICML)*, 2020.

Nele Nauwelaers, Harold Matthews, Yi Fan, Balder Croquet, Hanne Hoskens, Soha Mahdi, Ahmed El Sergani, Shunwang Gong, Tianmin Xu, Michael Bronstein, Mary Marazita, Seth Weinberg, and Peter Claes. Exploring palatal and dental shape variation with 3d shape analysis and geometric deep learning. *Orthodontics & Craniofacial Research*, 2021.

Liang Pan, Xinyi Chen, Zhongang Cai, Junzhe Zhang, Haiyu Zhao, Shuai Yi, and Ziwei Liu. Variational relational point completion network. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021.

HRD Peiris, TN Pitakotuwage, M Takahashi, K Sasaki, and E Kanazawa. Root canal morphology of mandibular permanent molars at different ages. *International Endodontic Journal*, 2008.

Jasmina Primožič, Maja Ovsenik, Stephen Richmond, Chung How Kau, and Alexei Zhurov. Early crossbite correction: a three-dimensional evaluation. *European Journal of Orthodontics*, 2009.

Charles R Qi, Hao Su, Kaichun Mo, and Leonidas J Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017a.

Charles R. Qi, Li Yi, Hao Su, and Leonidas J. Guibas. Pointnet++: deep hierarchical feature learning on point sets in a metric space. In *Neural Information Processing Systems (NeurIPS)*, 2017b.

Martin Reuter, Franz-Erich Wolter, and Niklas Peinecke. Laplace–beltrami spectra as 'shape-dna'of surfaces and solids. *Computer-Aided Design*, 2006.

Martin Reuter, Silvia Biasotti, Daniela Giorgi, Giuseppe Patanè, and Michela Spagnuolo. Discrete laplace–beltrami operators for shape analysis and segmentation. *Computers & Graphics*, 2009.

S. Rusinkiewicz and M. Levoy. Efficient variants of the icp algorithm. In *International Conference on 3D Digital Imaging and Modeling*, 2001.

Yuwen Tan, Xiang Xiang, Yifeng Chen, Hongyi Jing, Shiyang Ye, Chaoran Xue, and Hui Xu. Coupling bracket segmentation and tooth surface reconstruction on 3d dental models. In *Medical Image Computing and Computer Assisted Intervention (MICCAI)*, 2023.

Tiziano Testori, Tommaso Weinstein, Fabio Scutellà, Hom-Lay Wang, and Giovanni Zucchelli. Implant placement in the esthetic area: criteria for positioning single and multiple implants. *Periodontology*, 2018.

Sukun Tian, Renkai Huang, Zhenyang Li, Luca Fiorenza, Ning Dai, Yuchun Sun, and Haifeng Ma. A dual discriminator adversarial learning approach for dental occlusal surface reconstruction. *Journal of Healthcare Engineering*, 2022.

Shaofeng Wang, Changsong Lei, Yaqian Liang, Jun Sun, Xianju Xie, Yajie Wang, Feifei Zuo, Yuxin Bai, Song Li, and Yongjin Liu. A 3d dental model dataset with pre/post-orthodontic treatment for automatic tooth alignment. *Scientific Data*, 2024.

Svante Wold, Kim Esbensen, and Paul Geladi. Principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, 1987.

Xiaoyang Wu, Li Jiang, Peng-Shuai Wang, Zhijian Liu, Xihui Liu, Yu Qiao, Wanli Ouyang, Tong He, and Hengshuang Zhao. Point transformer v3: Simpler faster stronger. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024.

Runsen Xu, Xiaolong Wang, Tai Wang, Yilun Chen, Jiangmiao Pang, and Dahua Lin. Pointllm: Empowering large language models to understand point clouds. In *European Conference on Computer Vision (ECCV)*, 2025.

Guandao Yang, Xun Huang, Zekun Hao, Ming-Yu Liu, Serge Belongie, and Bharath Hariharan. Pointflow: 3d point cloud generation with continuous normalizing flows. In *International Conference on Computer Vision (ICCV)*, 2019.

Su Yang, Jiyong Han, Sang-Heon Lim, Ji-Yong Yoo, Su-Jeong Kim, Dahyun Song, Sunjung Kim, Jun-Min Kim, and Won-Jin Yi. Dcrownformer: Morphology-aware point-to-mesh generation transformer for dental crown prosthesis from 3d scan data of antagonist and preparation teeth. In *Medical Image Computing and Computer Assisted Intervention (MICCAI)*, 2024.

Yaoqing Yang, Chen Feng, Yiru Shen, and Dong Tian. Foldingnet: Point cloud auto-encoder via deep grid deformation. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.

Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. Adding conditional control to text-to-image diffusion models. In *International Conference on Computer Vision (ICCV)*, 2023.

Shen Zhu, Yinzhu Jin, Ifrah Zawar, and Tom Fletcher. Point-based shape representation generation with a correspondence-preserving diffusion model. In *Medical Imaging with Deep Learning (MIDL)*, 2025.