# CrownFusion

## 3D Dental Crown Generation using Geometry Images and Latent Diffusion

MICCAI 2026

Johan Ziruo Ye[1,2*] Xingguang Yan[3*] Søren Hauberg[1] Angel X. Chang[3,4] Hao Zhang[3] Peter Lempel Søndergaard[2]

[1]Technical University of Denmark [2]3Shape [3]Simon Fraser University [4]CIFAR AI Chair, Amii [*]equal contribution

**60×**
LOWER RECONSTRUCTION ERROR

First **geometry-image** representation for dental crowns

**0.733**
BITE PROXIMITY · 127 INTERSECTIONS · BOTH BEST

First **latent diffusion** model for crown generation

**11,513**
ANNOTATED SCANS, RELEASED

**FDI 16 dataset** scans paired with surrounding dentition

## Why generate crowns?

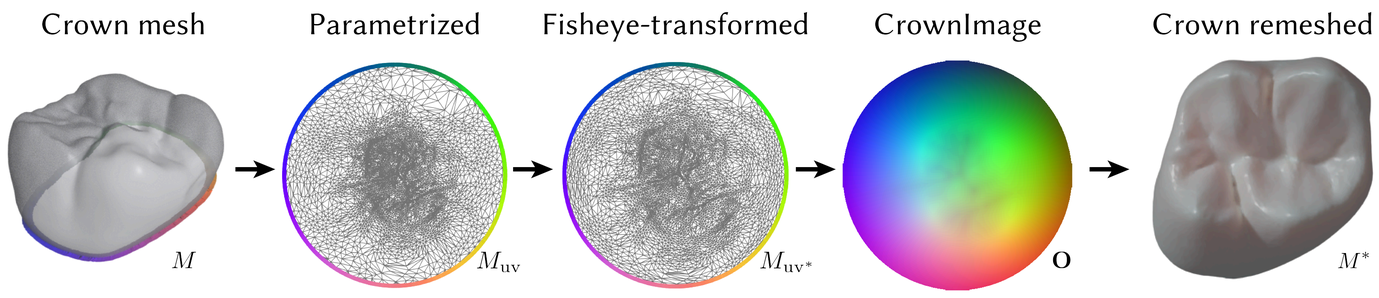
Crown design is labour-intensive and admits **multiple valid solutions** per patient, a natural fit for **generative** modelling. Every candidate must still **occlude correctly** with the opposing dentition and carry the natural fissures that guide food flow. But the shape representation has been the bottleneck: existing choices force a tradeoff between full 3D detail and 2D generative tooling, leaving no representation that supports both.

**Point clouds** smooth away the sharp fissures that define an occlusal surface, and still need a separate meshing stage.

**Depth images** project from a single viewpoint, discarding occluded regions like the lingual wall, or else demand multi-view fusion, adding registration complexity to the task.

**Key insight.** A crown is a dome with a rim, so its surface can be flattened onto a disk without any cutting or tearing. Storing each pixel's original 3D position turns the crown into an ordinary 2D image that a 2D generative model can learn, and the mesh is rebuilt exactly.

## The CrownImage representation




Pixel colour encodes 3D position (**R**= $x$, **G**= $y$, **B**= $z$): the rainbow disk *is* the crown surface, not a texture.

### 1 · Flatten the crown to a disk

A boundary-fixed harmonic map sends the mesh $M$ onto the unit disk, with no cuts.

### 2 · Zoom on the occlusal centre

A fisheye warp spends resolution on the occlusal centre and saves it at the axial rim.

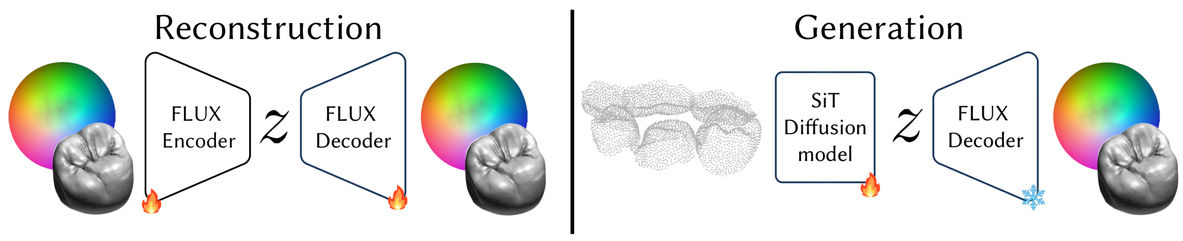
### 3 · Store 3D positions as pixels

Rasterise the inverse position map onto an $H \times W \times 3$ grid; each pixel stores $(x, y, z)$.

### 4 · Rebuild the mesh from pixels

Triangulate adjacent occupied pixels to recover $M^*$, no marching cubes, no post-hoc surfacing.

## The CrownFusion pipeline




### 1 · VAE: finetuned FLUX

Encodes CrownImages into compact latents. Finetuning matters: geometry images carry **coordinates**, not colour, so their distribution is far from FLUX's pretraining data. A small latent error is a **visible geometric distortion** on the reconstructed surface.

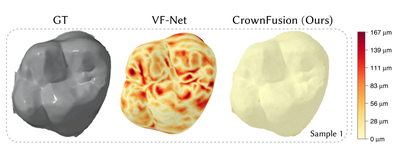
### 2 · Conditional SiT diffusion

A Scalable Interpolant Transformer trained from scratch in that latent space, cross-attended on the neighbouring and antagonist teeth via a jointly trained 3DShape2VecSet encoder, denoising towards a valid crown.

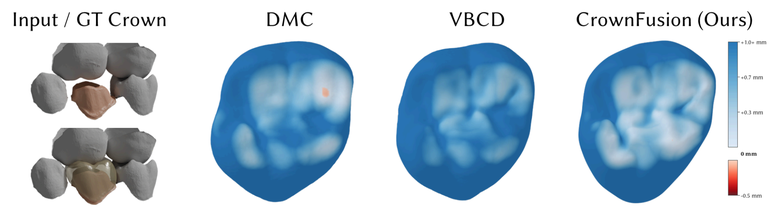
## Reconstruction

Generation is bounded by autoencoder fidelity. CrownImages reconstruct **two orders of magnitude** more accurately than point-cloud autoencoders, which blur the occlusal fissures and round off the **margin line**.

| | CD ×10²↓ | EMD ×10²↓ |
|---|---|---|
| LION | 5.35 | 22.85 |
| FoldingNet | 5.26 | 33.67 |
| VF-Net | 1.21 | 6.30 |
| **CrownFusion** | **0.02** | **0.45** |




## Generation




| | CD-L2↓ | Bite↓ | Int.#↓ |
|---|---|---|---|
| DMC | 0.433 | 0.734 | 256 |
| VBCD | **0.353** | 0.747 | 143 |
| **CrownFusion** | 0.412 | **0.733** | **127** |
| ground truth | — | 0.716 | 12 |

**Bite proximity**, distance to the antagonist, CrownFusion lands closest to ground truth and **halves** DMC's intersections.

## Variation

Each crown is a draw from the same conditioning, same neighbours, same antagonist, different noise, a shortlist of admissible crowns, not one averaged answer. Samples vary in fissure pattern, secondary fossae and oblique ridge, where clinicians have design freedom, while outline and ridge height stay stable so each still meets its neighbours.