Morphology-Driven Deep Watershed Transform for 3D Tooth Segmentation

Tomasz Szczepański[1][0000−0001−6189−478X] and Szymon Płotka[2][0000−0001−9411−820X]

[1] Sano Centre for Computational Medicine, Cracow, Poland
t.szczepanski@sanoscience.org
[2] Jagiellonian University, Cracow, Poland

Abstract. Segmentation of dentomaxillofacial structures in Cone-Beam Computed Tomography (CBCT) remains challenging, particularly for fine details such as root apices and nerve canals, which are crucial for evaluating root resorption in digital dentistry or to make surgical planning more precise. We present an approach that unifies instance detection and multi-class dentomaxillofacial structure segmentation in CBCT scans, in the scope of the ToothFairy3 Challenge. We adapt a Deep Watershed method, modeling each anatomical structure as a continuous 3D energy basin encoding voxel distances to class boundaries. This instance-aware representation ensures accurate segmentation of narrow, complex dentomaxillofacial structures. We train and evaluate our solution on the ToothFairy3 dataset, comprising 532 CBCT scans with voxel-wise annotations. Our method achieved a mean Dice coefficient of 0.742 and HD95 of 111.13 on the test set. We provide implementation at https://github.com/tomek1911/TF3.

Keywords: CBCT segmentation · ToothFairy3 Challenge · Morphological inductive bias · Deep Watershed

1 Introduction

In this report, we describe our solution for Task 1, "Multi-class segmentation" of the ToothFairy3 challenge. Automatic tooth segmentation in dental CBCT volumes is a critical step for various clinical applications, including orthodontic planning, endodontics, and surgical guidance. Building upon previous efforts in the ToothFairy challenges, we present a method adapted to the increased complexity of ToothFairy3. Compared to ToothFairy2, the new dataset contains 52 additional CBCT volumes acquired with a different scanner, and annotations have been substantially expanded to include 35 new labels, covering pulpy cavities for all 32 teeth, left and right incisive canals, and the lingual canal. The quality of annotations has also been improved, offering a richer resource for developing robust segmentation algorithms.

The task requires accurate voxel-wise labeling of all tooth structures and internal anatomical features within high-resolution CBCT volumes. It presents

several challenges: the small size and variability of pulp cavities, the complex shape of incisive and lingual canals, and the presence of noise and artifacts in CBCT scans. Furthermore, inter-patient anatomical variations and differences in scanner acquisition parameters increase the difficulty of generalizing segmentation models.

Several methods have been proposed for tooth segmentation in previous challenges and research field [4,1]. Classical approaches include atlas-based registration, graph-based techniques, or multi-stage approaches but the common part is that all recent advances leverage deep learning for volumetric segmentation. Notably, approaches such as SGANET [5], TSG-GCN [6], ToothSeg [3] and GEPAR3D [9] have demonstrated the effectiveness of combining volumetric convolutional networks with morphology-aware guidance. What is more, incorporating geometry-related features has been shown to enhance the model's generalization to external datasets [8].

Our approach extends the methodology proposed in GEPAR3D, incorporating a 3D Deep Watershed Transform guided by a direction map to enable morphology-aware learning of more than 32 teeth classes. This design allows the network to leverage both volumetric context and fine-grained morphological cues, leading to precise delineation of teeth and internal structures such as pulp cavities or nerve canals. To accommodate the high-resolution CBCT volumes within challenge memory constraints, we adapt a sliding window inference strategy, improving upon the MONAI-based sliding window used in the original GEPAR3D implementation. By combining morphology-guided learning with efficient volumetric inference, our solution effectively addresses the increased label complexity, variability, and inherent challenges of ToothFairy3.

2 Methods

An overview of our pipeline is presented in Fig. 1. The proposed solution builds upon the GEPAR3D method [9], extending it to the multi-class setting required by ToothFairy3. Our model jointly addresses multi-class semantic segmentation and instance-level regression, enabling it to separate individual teeth while also capturing their internal anatomical structures. To support both multi-class and binary segmentation objectives, we integrate strategies such as majority voting across classes and pulp fusion to ensure consistent labeling of internal cavities. During training, we introduce auxiliary objectives to enhance morphological awareness: an Energy Direction loss to model complex apex geometries and elongated nerve canals (see Fig. 2), and an instance regression task to generate energy maps that guide the 3D Deep Watershed Transform. These components together encourage the network to learn both local morphological details and global structural consistency.

Deep Watershed Instance Regression. To produce the inputs required by the Deep Watershed algorithm we train the network to solve two complementary volumetric regression tasks: (i) a continuous energy-basin regression that encodes each *pulp-free* tooth instance as a smooth scalar field and (ii) a per-voxel

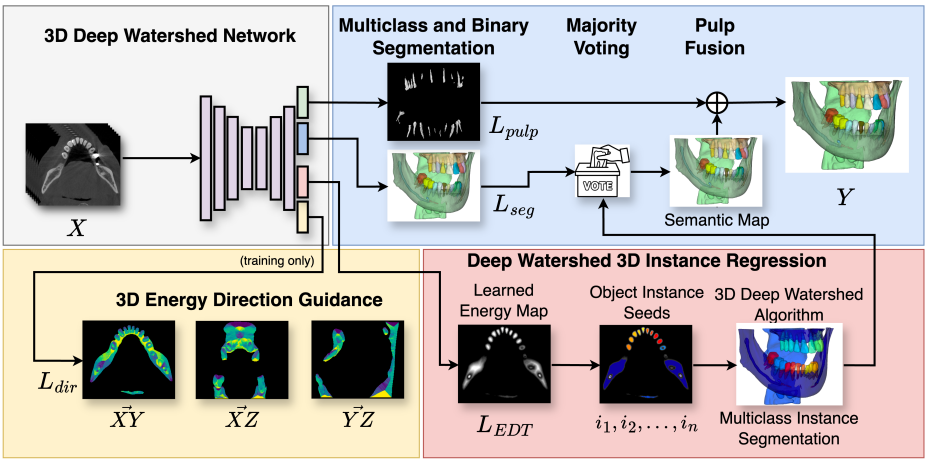

Fig. 1. An overview of the proposed solution, which unifies instance detection and multi-class segmentation for dentomaxillofacial structures in CBCT scans. Our model simultaneously performs multi-class segmentation and instance regression (gray). It also handles both multi-class and binary segmentation, incorporating techniques like majority voting and pulp fusion (blue). During training, we capture complex apex geometries via an Energy Direction loss (yellow) and use an instance regression task to generate energy maps for the Deep Watershed Algorithm (red).

direction (descent) estimate that refines boundary localization, especially in regions with steep gradients such as root apices and elongated nerve canals (see Fig. 3). We first create a secondary set of instance labels in which all pulp voxels have been removed from tooth instances (this guarantees that tooth instances are disjoint and suitable for watershed processing). Ground-truth energy basins $E_{GT}(\mathbf{r})$ are computed on these pulp-free instances using the Euclidean Distance Transform (EDT) to the each instance boundary separately (based on semantic classes of GT) and then normalized to $[0, 1]$ for numerical stability. The network regresses a continuous energy map $\widehat{E}(\mathbf{r})$ (single-channel) using a mean squared error objective:

$$L_{EDT} = \frac{1}{N} \sum_{\mathbf{r}} \left(E_{GT}(\mathbf{r}) - \widehat{E}(\mathbf{r}) \right)^2.$$

For directional supervision we compute the gradient field of the ground truth energy $\nabla E_{GT}(\mathbf{r})$ (implemented via a 3D Sobel-Feldman operator along x, y, z) and form unit direction vectors

$$\mathbf{u}_{GT}(\mathbf{r}) = \frac{\nabla E_{GT}(\mathbf{r})}{\max\{\|\nabla E_{GT}(\mathbf{r})\|_2, \varepsilon\}},$$

with a small ε to avoid division by zero. The model predicts a 3-channel direction vector $\widehat{\mathbf{u}}(\mathbf{r})$ which we normalize voxelwise. We supervise the directions with an

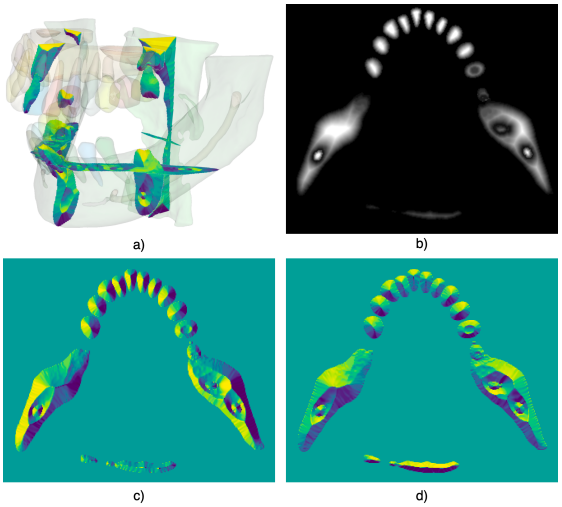

Fig. 2. We provide slices of the 3D Energy Direction Map (a) overlaid with semi-transparent (opacity 0.5) segmentation labels, enabling visualization of structural boundaries within spatial context. The direction map, derived by applying a 3D Sobel kernel to the distance map, assists the model in segmenting elongated and thin structures. While the distance map (b) approaches zero at the nerve canal–bone boundary, the direction map shows contrasting values, highlighting regions that are difficult to segment. Boundary regions between individual teeth (c, d) are similarly marked by abrupt vector changes, where regression errors are heavily penalized through the angular loss L_{dir}, enforcing directional consistency.

angular loss:

$$L_{\text{dir}} = \frac{1}{N} \sum_{i=1}^{N} \left(\frac{\cos^{-1}\left(\langle \mathbf{u}_{GT}^{(i)}, \widehat{\mathbf{u}}^{(i)} \rangle \right)}{\pi} \right)^2,$$

where N is the total number of voxels. We clip $\cos^{-1}$ inputs to $[-1, 1]$ for stability and divide by π to scale the angular error to $[0, 1]$. To focus the direction learning on anatomically relevant boundaries we mask N, see Fig. 2c to include voxels belonging to tooth instances and to thin/elongated semantic classes (e.g. nerve canals) but exclude pulp voxels.

Deep Watershed Instance Classification via Majority Voting. At inference, we first obtain voxel-wise semantic predictions for all classes (i.a. teeth without pulp, nerve canals, pulp binary map, jaw/skull bones) and the predicted continuous energy map $\widehat{E}$. To isolate found instances we binarize the semantic outputs into a *objects mask*. Seed points for watershed are extracted from predicted Energy Map basins by thresholding basin depth (empirically $\beta = 0.5$). The Watershed Transform is then run on $\widehat{E}$ constrained to *objects mask* and using the extracted seeds. This yields disjoint 3D objects instances V_j.

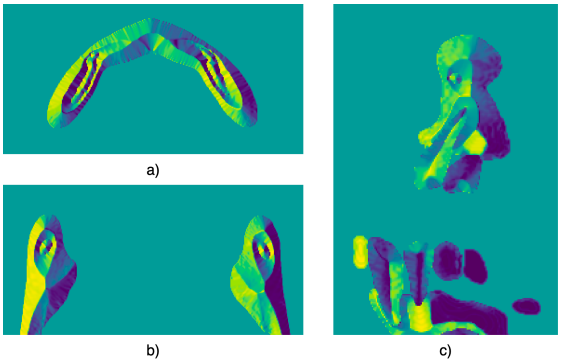

Fig. 3. Slices of the 3D Energy Direction Map with the inferior alveolar nerve visualized (a, b) show that the map clearly delineates the boundary between nerve and bone, both in perpendicular cross-sections and along the canal. In (c), the nerve canal and root apices are visible, with rapid angular transitions in the vector field highlighting anatomically complex regions. These transitions are particularly pronounced at the root apices, where fine, tapering structures curve sharply and diverge from surrounding bone.

Each resulting instance is assigned a semantic class by majority voting on the multi-class semantic branch:

$$\text{class}(V_j) \;=\; \arg\max_{c} \sum_{\mathbf{r}\in V_j} \mathbf{1}\{S(\mathbf{r}) = c\},$$

where $S(\mathbf{r})$ is the per-voxel semantic prediction and $\mathbf{1}\{\cdot\}$ is the indicator function.
Pulp fusion. During training, pulp voxels are optimized independently through L_{pulp}. Pulp segmentation is trained separately as a binary segmentation problem. We optimize a composite loss $L_{pulp} = L_{BCE}^{(w)} + L_{Dice}$, where $w_p = 5$ is for positive voxels to counteract severe imbalance. Since ToothFairy3 provides pulp annotations for all 32 teeth, but evaluation metrics treat pulp as a single aggregated class, we collapse these labels into one fused pulp mask. The final prediction is obtained by first running instance segmentation via deep watershed and majority voting for tooth and canal classes, followed by assigning pulp voxels on top of the corresponding multi-class predictions. This ensures consistency with the challenge evaluation protocol while still leveraging detailed pulp annotations during learning.
Overall training objective. The final loss function combines the contributions from semantic segmentation, pulp segmentation, and instance regression. Specifically, we use a weighted sum of four components: (i) multi-class semantic segmentation loss L_{seg}, implemented as a combination of cross-entropy and Dice; (ii) binary pulp segmentation loss L_{pulp}, formulated as weighted BCE plus Dice to address strong class imbalance; (iii) energy basin regression loss L_{EDT}, which drives accurate continuous energy map prediction for watershed separation; and (iv) direction field loss L_{dir}, which regularizes geometric consistency by

enforcing alignment between predicted and ground-truth descent directions. This design balances voxel-level classification with morphology-aware instance regression, ensuring robust segmentation of both large structures (e.g., jaw bones) and fine-scale anatomy (nerve canals, root apices).

Memory-efficient sliding-window inference. Large 3D volumes exceed GPU memory limits during dense prediction, so inference is typically performed with a sliding-window approach with overlapping patches. The default MONAI implementation accumulates intermediate patch predictions in lists before merging, which leads to high memory consumption proportional to the number of overlapping patches. To address this, we implemented a memory-efficient variant that directly accumulates predictions into preallocated output tensors, avoiding intermediate storage.

For each patch, we apply the model to obtain multi-class logits, energy distance maps, and pulp probabilities. Predictions are weighted by an importance map (constant or Gaussian blending) and accumulated on the fly into global tensors: voxel-wise probability sums on the CPU for multi-class segmentation, and GPU-accumulated maps for distance and pulp outputs. A separate weight accumulator ensures correct normalization. This design prevents redundant storage of overlapping patches while retaining smooth blending across patch boundaries. The memory-efficient approach reduces inference RAM memory usage substantially while preserving identical prediction quality to the original MONAI sliding window inferer.

3 Experimental design

3.1 Dataset

We train and evaluate our method on the novel ToothFairy3 dataset [7,1,2], which consists of multi-center data from centers A, B, and C, comprising 417, 63, and 52 cases, respectively. For training, we randomly selected 10 cases from each center for validation (30 in total), while the remaining 502 cases were used for training.

3.2 Implementation details

All scans are resampled to an isotropic resolution of $0.3 \times 0.3 \times 0.3$ mm^3, with Hounsfield Unit intensities clipped to $[0, 3000]$ and normalized to $[0, 1]$. During training, we randomly crop $288 \times 288 \times 160$ patches and pad with zeros if necessary. The model is trained for 400 epochs with AdamW, batch size of 2, and a cosine annealing scheduler. The loss function is defined as:

$$L = \Lambda_1 L_{EDT} + \Lambda_2 L_{seg} + \Lambda_3 L_{dir} + \Lambda_4 L_{pulp}, \tag{1}$$

with empirically set weights $\Lambda_1 = 10$, $\Lambda_2 = 0.1$, $\Lambda_3 = 1.0$, $\Lambda_4 = 1.0$ for balance. The initial learning rate and weight decay are set to $1e^{-3}$ and $1e^{-4}$, respectively.

Table 1. Official top 8 leaderboard test phase results for Task 1 - Multi-class segmentation of ToothFairy3 challenge.

Position	Team	mDSC (%)	mHD95 (mm)
1.	Black_Myth	79.81±6.4	88.72±32.33
2.	TAIR Lab	79.20±6.5	93.18±30.43
3.	sjtu_eiee	77.05±7.5	104.59±37.21
4.	ring821	76.84±9.7	104.40±47.98
5.	DLaBella29	73.86±7.1	97.71±33.20
6.	SMIR (ours)	74.22±8.1	111.13±39.40
7.	LAVIA Lab	69.70±9.4	144.97±48.90
8.	gagaha	55.1±17.6	172.49±63.60

Our implementation was developed with PyTorch 2.4.0 and MONAI 1.4.0. Training was performed on a single NVIDIA A100 GPU (80 GB) using float32 precision, while inference employed mixed precision (float16) and was executed on an NVIDIA T4 GPU (16 GB).

3.3 Evaluation metrics

The segmentation performance was quantitatively evaluated using two metrics: the Dice Similarity Coefficient (DSC, %) to measure volumetric overlap and the 95th percentile Hausdorff Distance (HD95, mm) to assess boundary accuracy. A third evaluation criterion, segmentation time, will be reported by the organizers following publication of the final ranking board.

4 Results

This section presents the quantitative and qualitative results from the official test phase leaderboard for "Task 1 - Multi-class Segmentation".

Quantitative results. Our solution participated in the "Task 1 - Multi-class Segmentation" challenge. Table 1 shows the official test phase leaderboard of the best eight submissions. Overall, we achieved a mDSC of 74.22±8.1% and a mHD95 of 111.13±39.40 mm across all 50 test cases. In the final leaderboard, we ranked 6th overall, and 5th in terms of mDSC among the 12 teams.

Qualitative results As shown in Fig. 4, our method produces generally accurate segmentations. Some errors remain, primarily undersegmentation of jaw bone structures or omission of the lingual nerve. Nonetheless, the method successfully delineated most of the challenging inferior alveolar nerve canal and correctly classified individual tooth instances.

5 Conclusions

In this work, we presented our solution for the ToothFairy3 challenge, addressing multi-class segmentation of CBCT scans including tooth instances, pulp cavities,

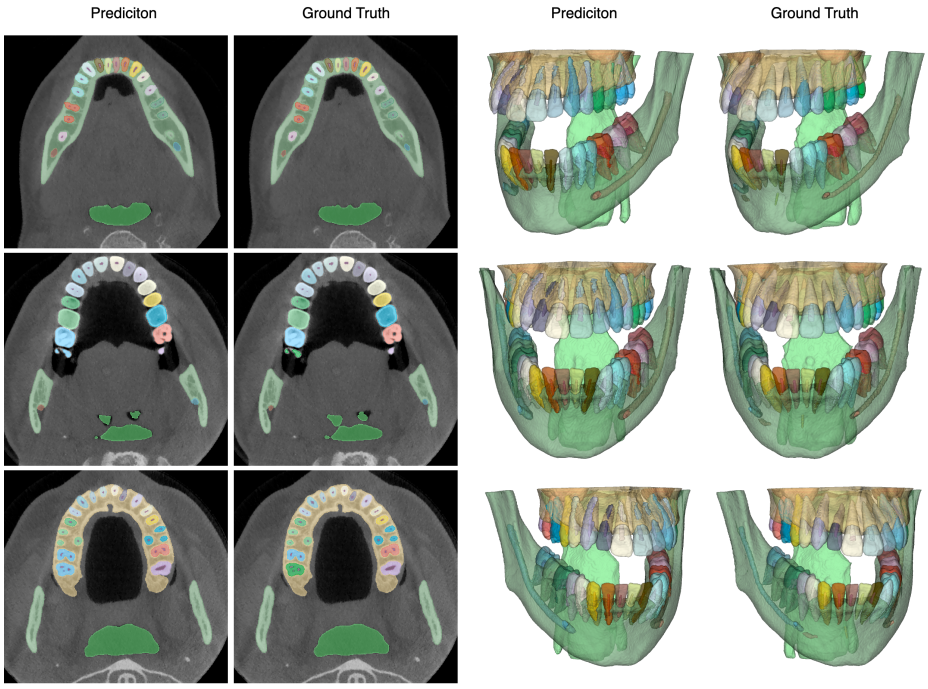

Fig. 4. Qualitative results of our method, proposed as a solution to the ToothFairy3 challenge. We visualize sample from validation set, center A. Ground truth is shown on the right, with both a 3D rendering and a representative 2D slices, while corresponding predictions are shown on the left. Our method yields precise nerve canal segmentation, as shown in the top-row slices and 3D transparent volumes, but shows reduced accuracy in matching the ground truth upper and lower jaw bone.

nerve canals, and jaw structures. Our method extends the GEPAR3D framework with a 3D deep watershed transform guided by direction maps, enabling morphology-aware learning and robust instance separation adapted to 45 dentomaxillofacial classes. Handling pulp as a separate binary task allowed effective fusion with Deep Watershed-based instances while avoiding label overlap.

We further introduced a memory-efficient sliding-window inference to process large CBCT volumes and optimized a combined loss comprising multi-class, pulp, and instance regression components to balance geometric precision with fine-structure accuracy. This design improved delineation of challenging anatomical features, such as root apices, narrow nerve canals, and pulp cavities.

Unfortunately, our method achieved results inferior to those reported in GEPAR3D. Unlike that approach, we did not leverage a geometrical prior to regularize the loss function, as a Statistical Shape Model was not available for the ToothFairy3 dentomaxillofacial labels. Furthermore, after submission we discovered that our Direction Map labels had been discretized, which substantially

reduced the information they carried. We plan to address this issue in future iterations.

Future work will integrate pulp directly into the multi-class segmentation branch and refine the direction-map auxiliary task to better capture narrow pulp fragments and fine canal structures, aiming to further enhance segmentation accuracy and anatomical fidelity.

Acknowledgments. Tomasz Szczepański is supported by the EU's Horizon 2020 programme (grant no. 857533, Sano) and the Foundation for Polish Science's International Research Agendas programme (MAB PLUS/2019/13), co-financed by the EU under the European Regional Development Fund and the Polish Ministry of Science and Higher Education (contract no. MEiN/2023/DIR/3796).

Disclosure of Interests. The authors have no competing interests to declare.

References

1. Bolelli, F., Lumetti, L., Vinayahalingam, S., et al.: Segmenting the inferior alveolar canal in CBCT volumes: the toothfairy challenge. IEEE Transactions on Medical Imaging (2024)
2. Bolelli, F., Marchesini, K., van Nistelrooij, N., Lumetti, L., Pipoli, V., Ficarra, E., Vinayahalingam, S., Grana, C.: Segmenting maxillofacial structures in cbct volumes. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 5238–5248 (2025)
3. Cui, Z., Zhang, B., Lian, C., et al.: Hierarchical morphology-guided tooth instance segmentation from CBCT images. In: Information Processing in Medical Imaging. pp. 150–162. Springer (2021)
4. Isensee, F., Kirchhoff, Y., Kraemer, L., Rokuss, M., Ulrich, C., Maier-Hein, K.H.: Scaling nnu-net for cbct segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 13–20. Springer (2024)
5. Li, P., Liu, Y., Cui, Z., et al.: Semantic graph attention with explicit anatomical association modeling for tooth segmentation from CBCT images. IEEE Transactions on Medical Imaging **41**(11), 3116–3127 (2022)
6. Liu, Y., Zhang, S., Wu, X., et al.: Individual graph representation learning for pediatric tooth segmentation from dental CBCT. IEEE Transactions on Medical Imaging (2024)
7. Lumetti, L., Pipoli, V., Bolelli, F., Ficarra, E., Grana, C.: Enhancing patch-based learning for the segmentation of the mandibular canal. IEEE Access **12**, 79014–79024 (2024)
8. Szczepański, T., Grzeszczyk, M.K., Płotka, S., et al.: Let me DeCode you: Decoder conditioning with tabular data. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 228–238. Springer (2024)
9. Szczepański, T., Płotka, S., Grzeszczyk, M.K., Adamowicz, A., Fudalej, P., Korzeniowski, P., Trzciński, T., Sitek, A.: Gepar3d: Geometry prior-assisted learning for 3d tooth segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 218–228. Springer (2025)