U-Mamba2: Scaling State Space Models for Dental Anatomy Segmentation in CBCT

Zhi Qin Tan, Xiatian Zhu, Owen Addison, Yunpeng Li

ODIN Workshop, MICCAI 2025

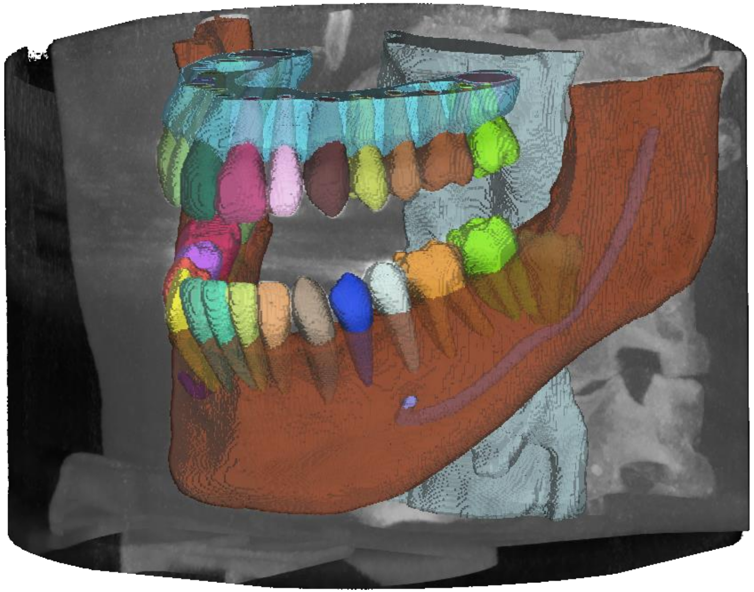
Background & Motivation

- CBCT: widely-used image modality
- Provides 3D volumetric information of the orofacial region
- Accurate segmentation of anatomical structures is crucial for downstream applications

Background & Motivation

ToothFairy3 Challenges – MICCAI 2025

1. Fast Multi-Structure Segmentation in CBCT Volumes
2. Interactive Segmentation in CBCT Volumes



Related Works

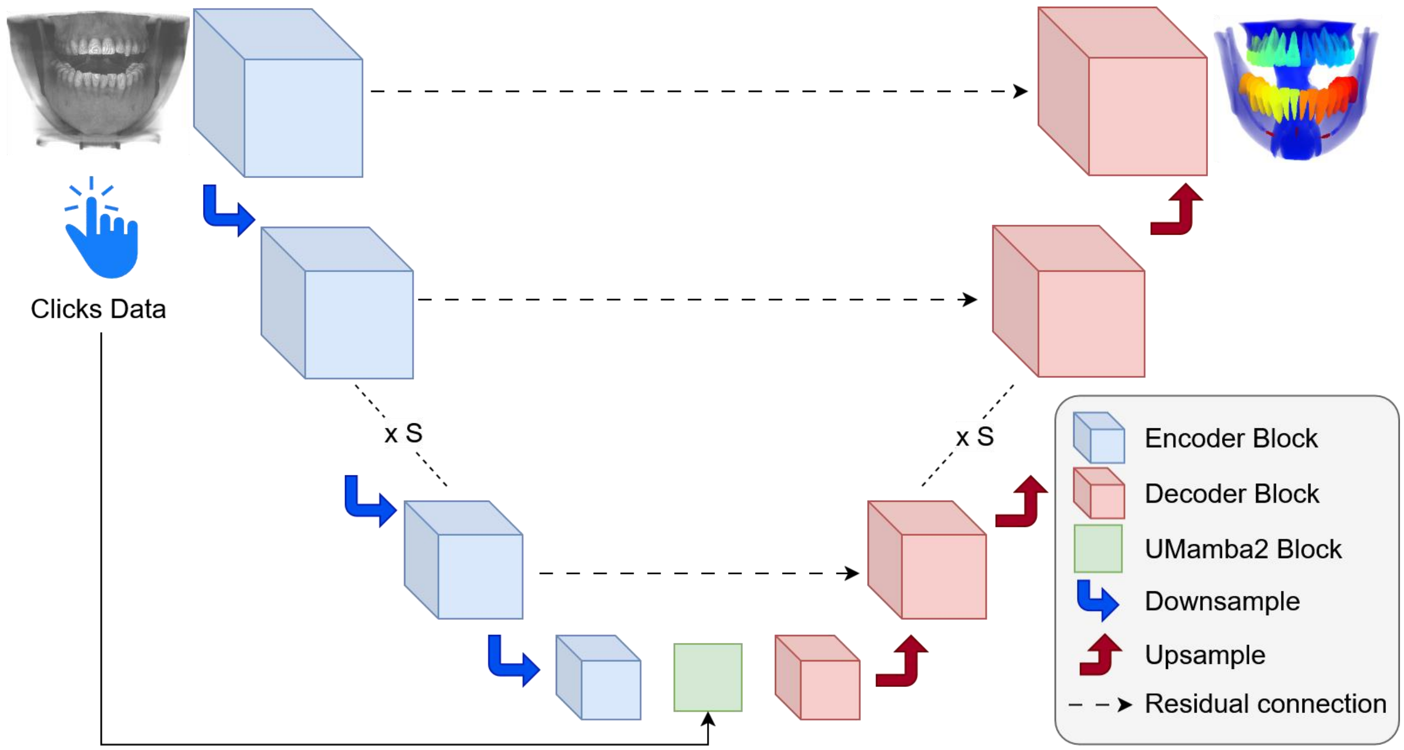
1. CNNs (U-Net, DeepLab)
    - Translational-invariant and parameter-efficient
    - Weak in capturing global features
2. Transformers (SETR, SwinTransformer)
    - Attention mechanism to capture global relationship
    - Computational resource-intensive
3. Hybrid CNN-Transformer (nnFormer, SwinUNETR)
    - Exploit the strength of both architectures

Related Works

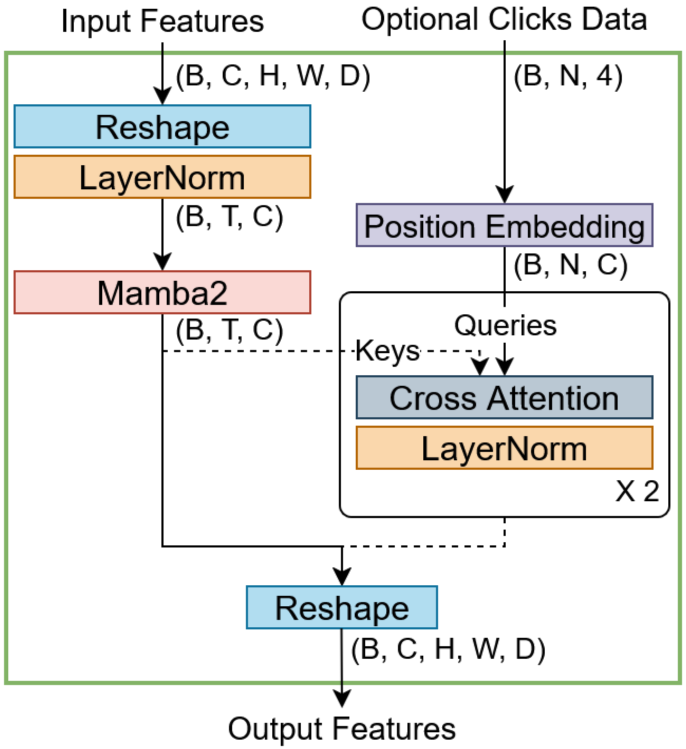
- U-Mamba
  - Utilize Mamba for semantic segmentation
  - Surpass transformers in accuracy and speed

- Mamba2
  - Enforce stronger constraints on the hidden space update matrix and move several projection layers earlier
  - Enable tensor and sequence parallelism within the same framework

U-Mamba2 Model



U-Mamba2 Block

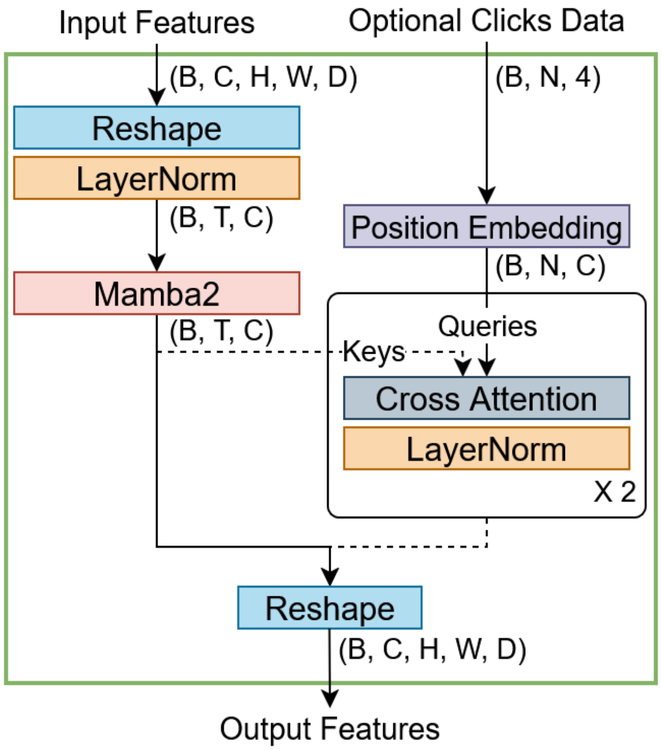


B: Batch size
C: No. of channels
N: No. of clicks
H, W, W: Spatial dim.
T = H x W x D

Interactive Clicks

- Similar to SAM2
  - Learnable Position Embedding embeds 3D coordinates and class labels
  - Two Cross-Attention blocks fuse the prompts and image features

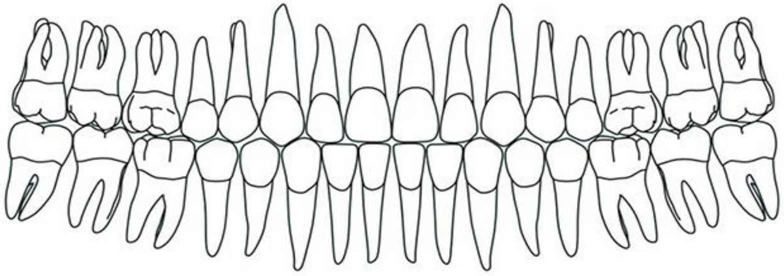
B: Batch size
C: No. of channels
N: No. of clicks
H, W, W: Spatial dim.
T = H x W x D

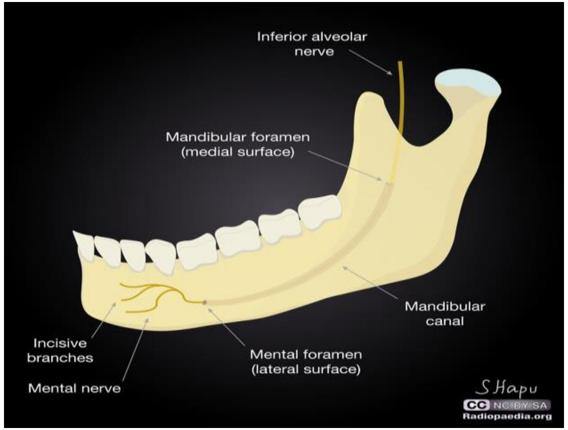


Incorporating Domain Knowledge

Related Anatomies

- Anatomies share similar shapes and properties
- Label smoothing -> guide the model in recognizing similar classes and their relationship

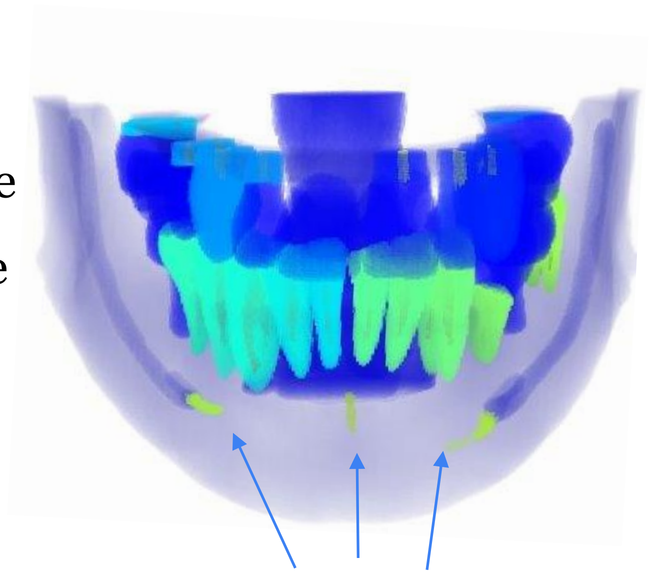




Incorporating Domain Knowledge

Anatomy Structural Size

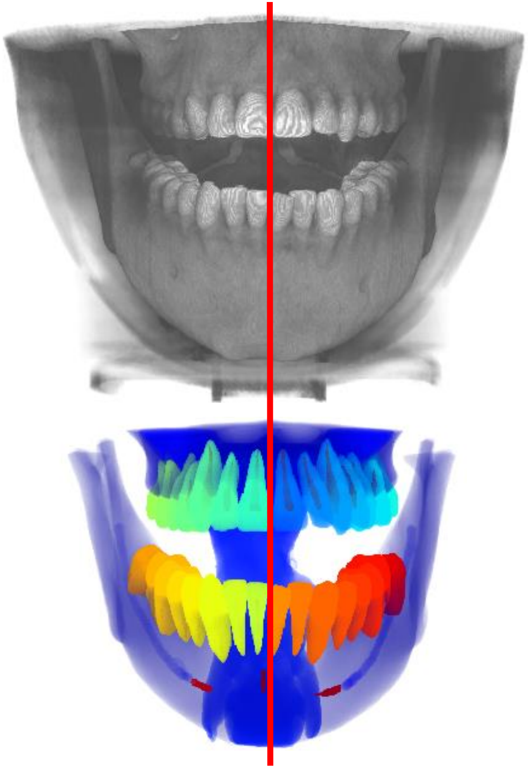
- Left and right incisive nerves, and lingual foramen houses thin structures in the mandible
- Assign class weights based on structure volume
- Prevent larger anatomies from dominating the overall loss



Incorporating Domain Knowledge

Left-Right Mirroring Augmentation

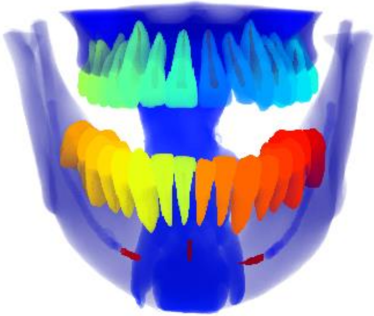
- Structural symmetry between L/R anatomies in the sagittal plane confuses even dentists
- Propose to swap the class labels of L/R counterparts when L/R mirroring is applied
- Increasing number of possible axes combination for mirroring augmentation: 3 -> 7

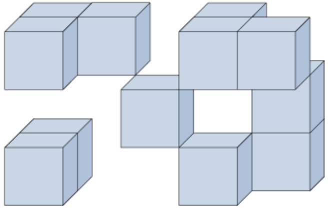


Incorporating Domain Knowledge

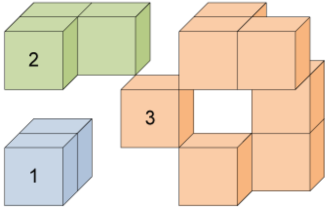
Post-processing

- Voxels of the same anatomy are connected and not separated blobs
- Incorporate this prior with post-processing to remove small disconnected segmentations
- Volume threshold computed with the training dataset





Region of interest

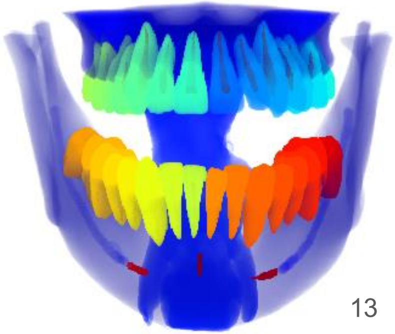


18-connected

Experiments

Dataset

- Task 1: Fast and accurate multi-structure segmentation (77 classes)
  - 32 tooth + pulps, jaws, sinuses, crowns, IAN, etc.
- Task 2: Segmentation of left and right IAN + interactive clicks
- 532 CBCT with shapes: (170, 272, 345) - (298, 512, 512)



Experiments

Baselines: SwinUNETR, nnU-Net ResE, U-Mamba

- 7 encoder-decoder stages
- Input patch size: 128 x 256 x 256
- Batch size: 1
- Sliding window inference with 0.5 tile size & no L/R mirroring in test-time augmentation (TTA)

Results

Task 1

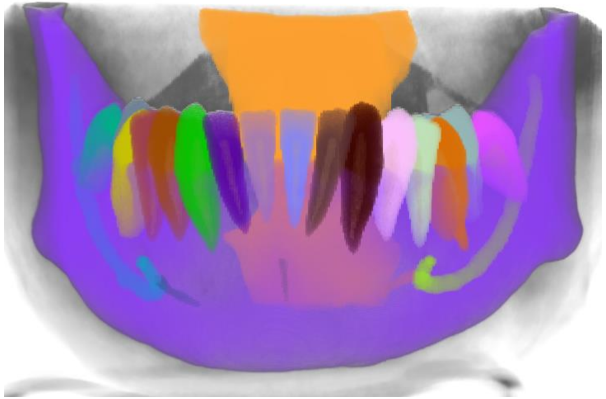
Model	Dice	HD95	Dice (PP)	HD95 (PP)	Time (s)
SwinUNETR	0.858	48.86	0.874	40.09	7.23
nnU-Net ResE	0.861	45.28	0.887	32.05	**6.20**
U-Mamba	0.865	42.06	0.896	25.88	6.98
U-Mamba2	**0.873**	**41.08**	**0.908**	**21.35**	6.81

Task 2

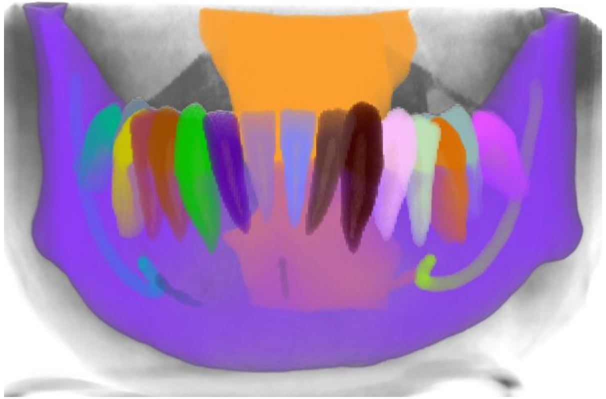
Model	Dice	HD95	Dice (PP)	HD95 (PP)	Time (s)
nnU-Net ResE	0.901	1.98	0.905	1.71	**5.06**
U-Mamba	0.903	1.65	**0.913**	1.58	5.88
U-Mamba2	**0.905**	**1.63**	**0.913**	**1.57**	5.70

Results (Best Case)

Ground Truth



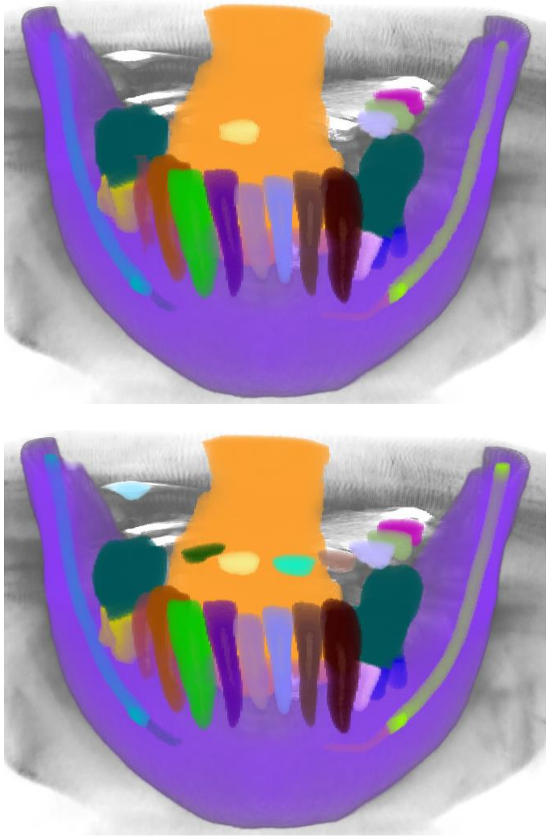
Prediction

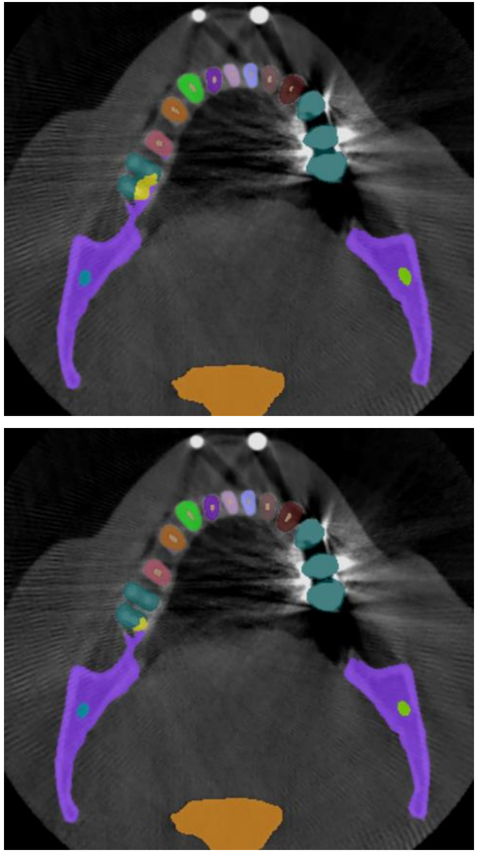


Results (Worst Case)

Ground Truth

Prediction





Results

Task 1 Winners - Fast Multi-Structure Segmentation in CBCT Volumes

Participant	Runtime(s)	Dice	Dice Rank	HD95	HD95 Rank	Rank
🥇 TAIR-Lab	40,58	0,84	2,4	38,17	3,0	3,1
🥈 sjtu_eiee_2-426lab	17,46	0,77	4,9	94,77	5,3	3,7
🥉 Black_Myth	90,04	0,85	2,0	33,23	2,3	3,8

Results

Task 2 Winners - Interactive Segmentation of the Inferior Alveolar Canal (IAC) in CBCT Volumes ¶

Participant	Runtime(s)	Dice Left IAC AUC	Dice Left IAC Final	Dice Right IAC AUC	Dice Right IAC Final	HD95 Left IAC AUC	HD95 Left IAC Final	HD95 Right IAC AUC	HD95 Right IAC Final	Rank
🥇 TAIR_Lab	100,64	4,34	0,87	4,31	0,86	11,29	2,26	10,2	2,04	1,66
🥈 BlackMyth	168,42	4,3	0,86	4,32	0,87	15,05	2,54	14,55	2,03	2,11
🥉 changkkk	16,09	3,79	0,76	3,83	0,77	201,75	40,35	131,42	26,28	3,44

Conclusion

- Presented a new architecture, U-Mamba2 for multi-anatomy CBCT segmentation
- Integrating U-Net with Mamba2 achieved higher efficiency and performance
- Incorporating domain knowledge of dental anatomy improved performance of CBCT segmentation

Thank you!

Codes

https://github.com/zhiqin1998/UMamba2