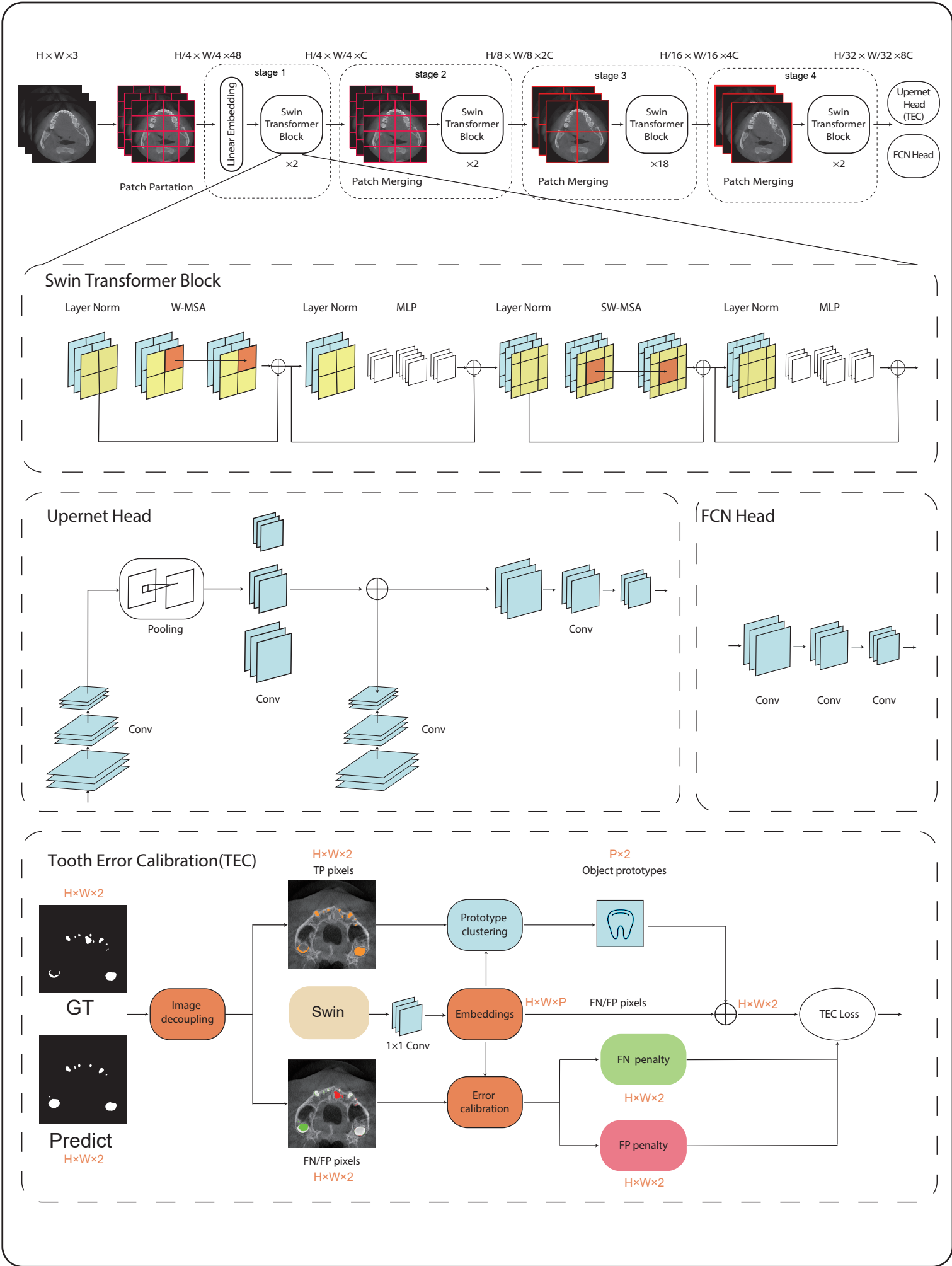


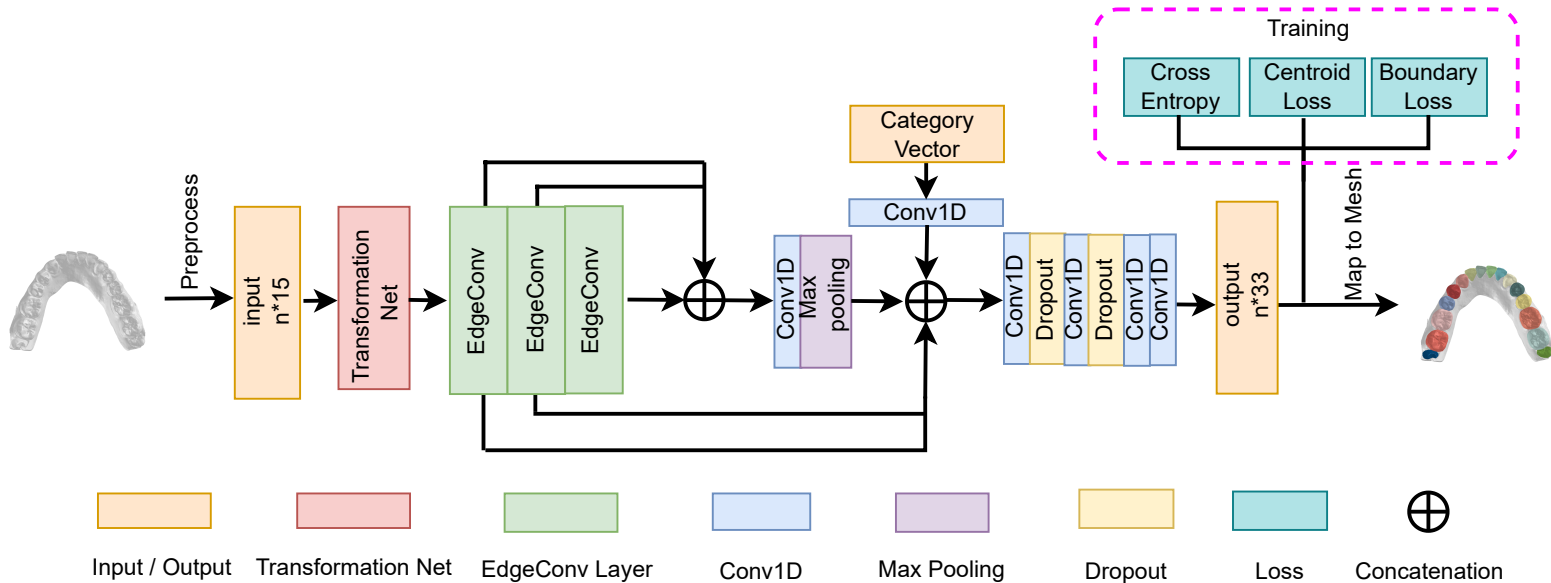
Supplementary Figure 1



Supplementary Figure 2

Layers	Output size (Output name)	TSTNet
Input	$h \times w \times 3$	
Patch Embedding	$h/4 \times w/4 \times 128$	$\left[\begin{array}{l} \text{Patch Partition} \rightarrow h/4 \times w/4 \times 48 \\ \text{Linear Embedding: } 4 \times 4, \text{ dim } 128 \text{ Conv2d, stride } 4 \rightarrow \text{LN} \end{array} \right]$
Swin Block	$h/4 \times w/4 \times 128$	$\left[\begin{array}{l} \text{LN} \rightarrow \text{window size } 7, \text{ dim } 128, \text{ head } 4 \text{ W-MSA} \rightarrow \text{Residual connect} \\ \text{LN} \rightarrow \text{MLP} \rightarrow \text{Residual connect} \\ \text{LN} \rightarrow \text{window size } 7, \text{ dim } 128, \text{ head } 4 \text{ SW-MSA} \rightarrow \text{Residual connect} \\ \text{LN} \rightarrow \text{MLP} \rightarrow \text{Residual connect} \end{array} \right]$
Swin_stage2	$h/8 \times w/8 \times 256$	$\left[\begin{array}{l} \text{Patch Merging} \left[\begin{array}{l} \text{group feature map} \rightarrow h/8 \times w/8 \times 4 \times 128 \\ \text{LN} \rightarrow \text{dim } 128 \times 2 \text{ linear} \rightarrow h/8 \times w/8 \times 256 \end{array} \right] \\ \text{window size } 7, \text{ dim } 256, \text{ head } 8 \text{ Swin Block} \end{array} \right]$
Swin_stage3	$h/16 \times w/16 \times 512$	$\left[\begin{array}{l} \text{Patch Merging} \left[\begin{array}{l} \text{group feature map} \rightarrow h/16 \times w/16 \times 4 \times 256 \\ \text{LN} \rightarrow \text{dim } 256 \times 2 \text{ linear} \rightarrow h/16 \times w/16 \times 512 \end{array} \right] \\ \text{window size } 7, \text{ dim } 512, \text{ head } 16 \text{ Swin Block} \end{array} \right] \times 9$
Swin_stage4	$h/32 \times w/32 \times 1024$ (pred seg)	$\left[\begin{array}{l} \text{Patch Merging} \left[\begin{array}{l} \text{group feature map} \rightarrow h/32 \times w/32 \times 4 \times 512 \\ \text{LN} \rightarrow \text{dim } 512 \times 2 \text{ linear} \rightarrow h/32 \times w/32 \times 1024 \end{array} \right] \\ \text{window size } 7, \text{ dim } 1024, \text{ head } 32 \text{ Swin Block} \end{array} \right]$
TEC [discard when infering]	(metric loss)	$\left[\begin{array}{l} \text{Compute TP, FN, FP(GT labels, pred seg)} \\ \text{Compute category prototype(TP pixel) by EMA} \\ \text{Compute cos similarity(pixel, category prototype)} \\ \text{Error calibration} \left[\begin{array}{l} \text{compute FP penalty term(FP pixel, anchor)} \\ \text{compute FN penalty term(FN pixel, anchor)} \end{array} \right] \\ \text{Compute metric loss(cos similaity, penalty, compensation)} \end{array} \right]$
FCN Head [auxiliary head, discard when infering]	$2 \times h \times w$	$\left[\begin{array}{l} 3 \times 3, \text{ dim } 256 \text{ Conv2d} \rightarrow \text{SyncBN} \\ 3 \times 3, \text{ dim } 256 \text{ Conv2d} \rightarrow \text{SyncBN} \\ \text{ratio } 0.1 \text{ Dropout2d} \rightarrow 1 \times 1, \text{ dim } 2 \text{ Conv2d} \end{array} \right]$
UperNet Head [decode head]	$2 \times h \times w$	$\left[\begin{array}{l} \text{head_input channels } [128, 256, 512, 1024] \\ \left[\begin{array}{l} \text{channels } 128 \rightarrow 1 \times 1, \text{ dim } 512 \text{ Conv2d} \rightarrow \text{SyncBN} \rightarrow \text{out_128} \\ \text{channels } 256 \rightarrow 1 \times 1, \text{ dim } 512 \text{ Conv2d} \rightarrow \text{SyncBN} \rightarrow \text{out_256} \\ \text{channels } 512 \rightarrow 1 \times 1, \text{ dim } 512 \text{ Conv2d} \rightarrow \text{SyncBN} \rightarrow \text{out_512} \end{array} \right] \\ \text{channels } 1024 \rightarrow \left[\begin{array}{l} \text{scale } 1 \text{ AdaptiveAvgPool} \rightarrow 1 \times 1, \text{ dim } 512 \text{ Conv2d} \rightarrow \text{SyncBN} \\ \text{scale } 2 \text{ AdaptiveAvgPool} \rightarrow 1 \times 1, \text{ dim } 512 \text{ Conv2d} \rightarrow \text{SyncBN} \\ \text{scale } 3 \text{ AdaptiveAvgPool} \rightarrow 1 \times 1, \text{ dim } 512 \text{ Conv2d} \rightarrow \text{SyncBN} \\ \text{scale } 6 \text{ AdaptiveAvgPool} \rightarrow 1 \times 1, \text{ dim } 512 \text{ Conv2d} \rightarrow \text{SyncBN} \end{array} \right] \\ \rightarrow \text{concat along channel dim} \rightarrow 3 \times 3, \text{ dim } 512 \text{ Conv2d} \rightarrow \text{SyncBN} \rightarrow \text{out_1024} \\ \left[\begin{array}{l} \text{out_128} \rightarrow 3 \times 3, \text{ dim } 512 \text{ Conv2d} \rightarrow \text{SyncBN} \\ \text{out_256} \rightarrow 3 \times 3, \text{ dim } 512 \text{ Conv2d} \rightarrow \text{SyncBN} \\ \text{out_512} \rightarrow 3 \times 3, \text{ dim } 512 \text{ Conv2d} \rightarrow \text{SyncBN} \end{array} \right] \\ \text{SyncBN} \leftarrow 3 \times 3, \text{ dim } 512 \text{ Conv2d} \leftarrow \text{concat along channel dim} \leftarrow \\ \rightarrow \text{ratio } 0.1 \text{ Dropout2d} \rightarrow 1 \times 1, \text{ dim } 2 \text{ Conv2d} \end{array} \right]$

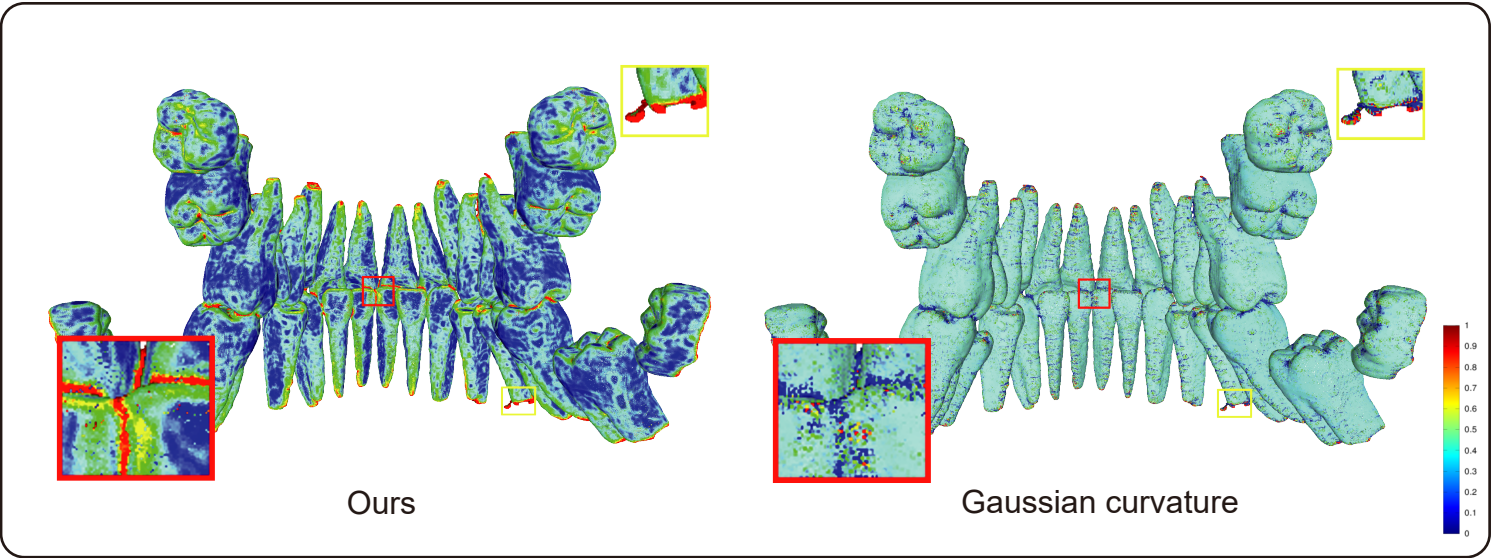
Supplementary Figure 3



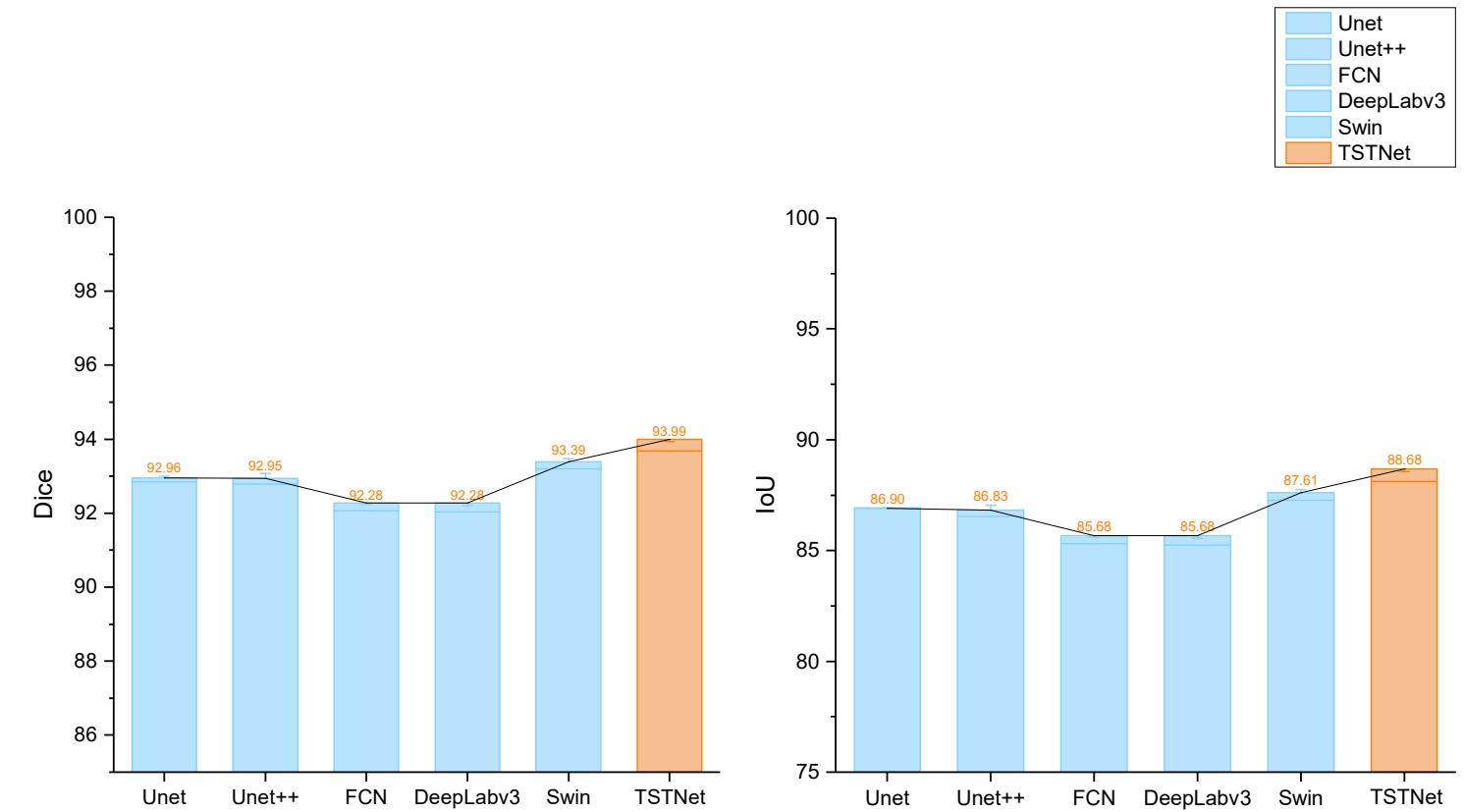
Supplementary Figure 4

Layers	Output Size (output name)	IOSNet
Input	$b \times n \times 15$	
Get_graph_feature	$b \times 30 \times n \times k$	$\text{knn}(k) \rightarrow b \times n \times k$ $\text{get_neighbor_feature} \rightarrow b \times n \times k \times 15$ $\text{offset_feature} = \text{neighbor_feature} - \text{center_point_feature}$ $\rightarrow b \times n \times k \times 15$ $\text{concat offset_feature and center_point_feature along channel dimension} \rightarrow b \times n \times k \times 30$
Transformation Net	$b \times n \times 15$	$1 \times 1, 128 \text{ Conv2d} \rightarrow \text{BatchNorm} \rightarrow \text{Mish activation}$ $1 \times 1, 256 \text{ Conv2d} \rightarrow \text{BatchNorm} \rightarrow \text{Mish activation}$ $\text{max pool along } k_neighbor \text{ dim} \rightarrow b \times 256 \times n$ $1 \times 1, 1024 \text{ Conv1d} \rightarrow \text{BatchNorm} \rightarrow \text{Mish activation}$ $\text{max pool along } n \text{ dim} \rightarrow b \times 1024$ $512\text{-d linear} \rightarrow \text{Mish activation}$ $256\text{-d linear} \rightarrow \text{Mish activation}$ $225\text{-d linear} \rightarrow \text{reshape as } b \times 15 \times 15 (\text{transform matrix})$ $\text{input} * \text{transform matrix}$
EdgeConv_1	$b \times 64 \times n$ (edgeconv1)	get_graph_feature $1 \times 1, 128 \text{ Conv2d} \rightarrow \text{BatchNorm} \rightarrow \text{Mish activation}$ $1 \times 1, 128 \text{ Conv2d} \rightarrow \text{BatchNorm} \rightarrow \text{Mish activation}$ $1 \times 1, 64 \text{ Conv2d} \rightarrow \text{BatchNorm} \rightarrow \text{Mish activation}$ $\text{max pool along } k_neighbor \text{ dim}$
EdgeConv_2	$b \times 64 \times n$ (edgeconv2)	get_graph_feature $1 \times 1, 64 \text{ Conv2d} \rightarrow \text{BatchNorm} \rightarrow \text{Mish activation}$ $1 \times 1, 64 \text{ Conv2d} \rightarrow \text{BatchNorm} \rightarrow \text{Mish activation}$ $1 \times 1, 64 \text{ Conv2d} \rightarrow \text{BatchNorm} \rightarrow \text{Mish activation}$ $\text{max pool along } k_neighbor \text{ dim}$
EdgeConv_3	$b \times 64 \times n$ (edgeconv3)	get_graph_feature $1 \times 1, 64 \text{ Conv2d} \rightarrow \text{BatchNorm} \rightarrow \text{Mish activation}$ $1 \times 1, 64 \text{ Conv2d} \rightarrow \text{BatchNorm} \rightarrow \text{Mish activation}$ $\text{max pool along } k_neighbor \text{ dim}$
Get_global_feature	$b \times 1024$ (global_feature)	$\text{concat (edgeconv1, edgeconv2, edgeconv3)}$ $1 \times 1, 1024 \text{ Conv1d} \rightarrow \text{BatchNorm} \rightarrow \text{Mish activation}$ $\text{max pool along } n \text{ dim}$
Category Vector	$b \times 2$	
Add_category_vector	$b \times 1088 \times n$ (new_global_fea)	$1 \times 1, 64 \text{ Conv1d} \rightarrow \text{Mish activation} \rightarrow \text{category_feature}$ $\text{concat(global_feature, category_feature)}$
Get_all_feature	$b \times 1280 \times n$	$\text{concat (new_global_fea, edgeconv1, edgeconv2, edgeconv3)}$
Segmentation head	$b \times 33 \times n$	$1 \times 1, 256 \text{ Conv1d} \rightarrow \text{BatchNorm} \rightarrow \text{Mish activation}$ dropout $1 \times 1, 256 \text{ Conv1d} \rightarrow \text{BatchNorm} \rightarrow \text{Mish activation}$ dropout $1 \times 1, 128 \text{ Conv1d} \rightarrow \text{BatchNorm} \rightarrow \text{Mish activation}$ $1 \times 1, 33 \text{ Conv1d}$

Supplementary Figure 5

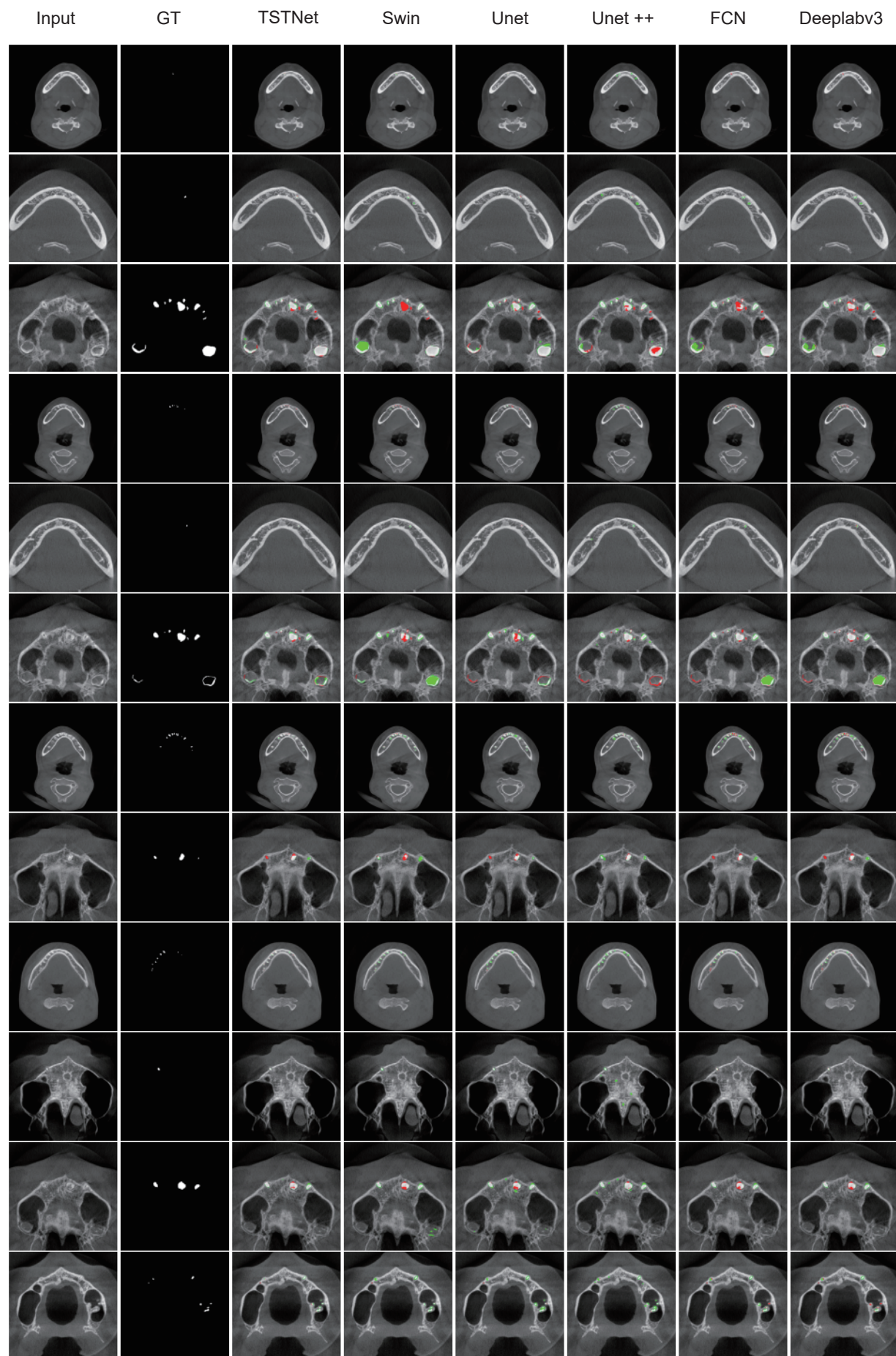


Supplementary Figure 6



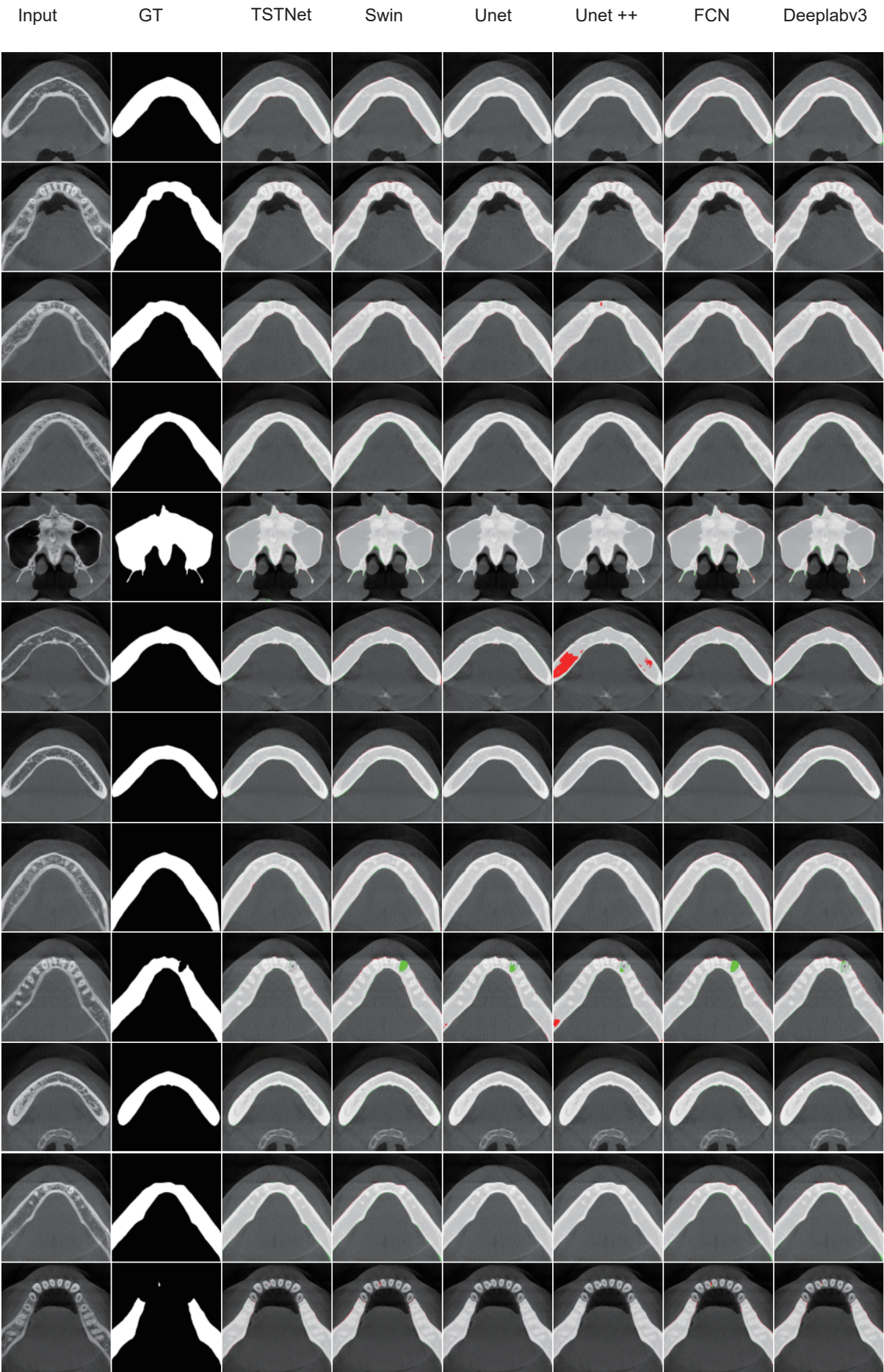
Supplementary Figure 7

CBCT Teeth segmentation visualization



Supplementary Figure 8

CBCT Jaw segmentation visualization



Supplementary Figure 9



Supplementary Figure 10

