TOOTH FAIRY3

Fast and **Interactive** Multi-class Segmentation in CBCT Volumes

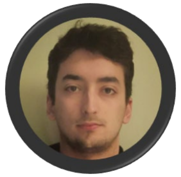
odin-workshops.org MICCAI2025

Organizers


Federico Bolelli
University of Modena and Reggio Emilia, Italy
federico.bolelli@unimore.it


Mattia Di Bartolomeo
University of Modena and Reggio Emilia, Italy
mattiadiba@hotmail.it


Natasha Hallberg
3Shape A/S, Denmark
Natasha.Hallberg@3shape.com


Luca Lumetti
University of Modena and Reggio Emilia, Italy
luca.lumetti@unimore.it


Kevin Marchesini
University of Modena and Reggio Emilia, Italy
kevin.marchesini@unimore.it


Alexandre Anesi
University of Modena and Reggio Emilia, Italy
alexandre.anesi@unimore.it


Zdravko Marinov
Karlsruhe Institute of Technology, Germany
zdravko.marinov@kit.edu


Vittorio Pipoli
University of Modena and Reggio Emilia, Italy
vittorio.pipoli@unimore.it


Rainer Stiefelhagen
Karlsruhe Institute of Technology, Germany
rainer stiefelhagen@kit edu


Shankeeth Vinayahalingam
Radboud University, Netherlands
shankeeth.vinayahalingam@radboudumc.nl


Torkan Gholamalizadeh
3Shape A/S, Denmark
torkan.gholamalizadeh@3shape.com

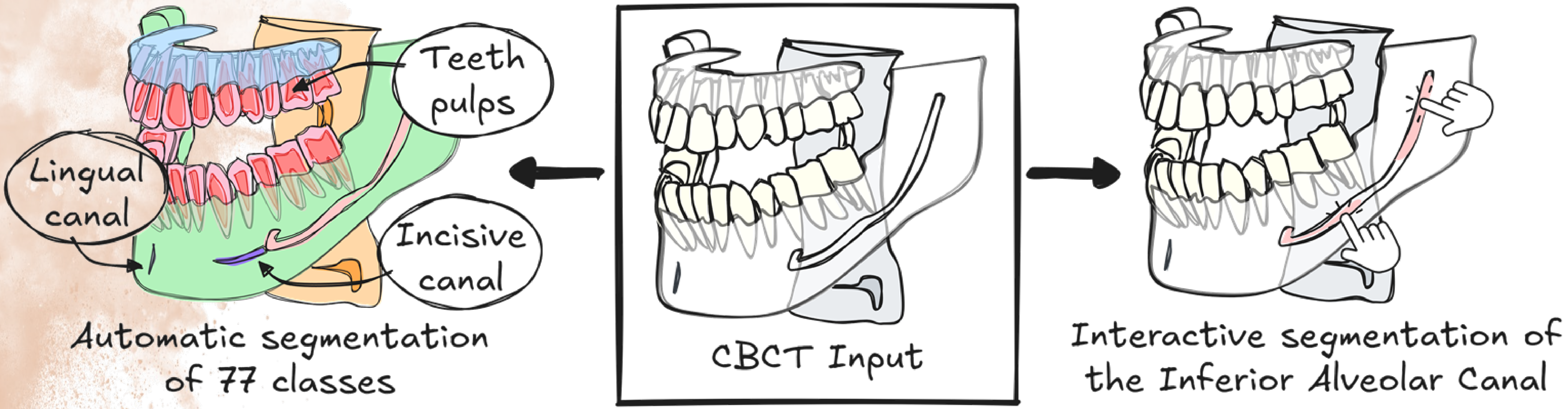

Costantino Grana
University of Modena and Reggio Emilia, Italy
costantino.grana@unimore.it


Niels van Nistelrooij
Radboud University, Netherlands
niels.vanNistelrooij@radboudumc.nl

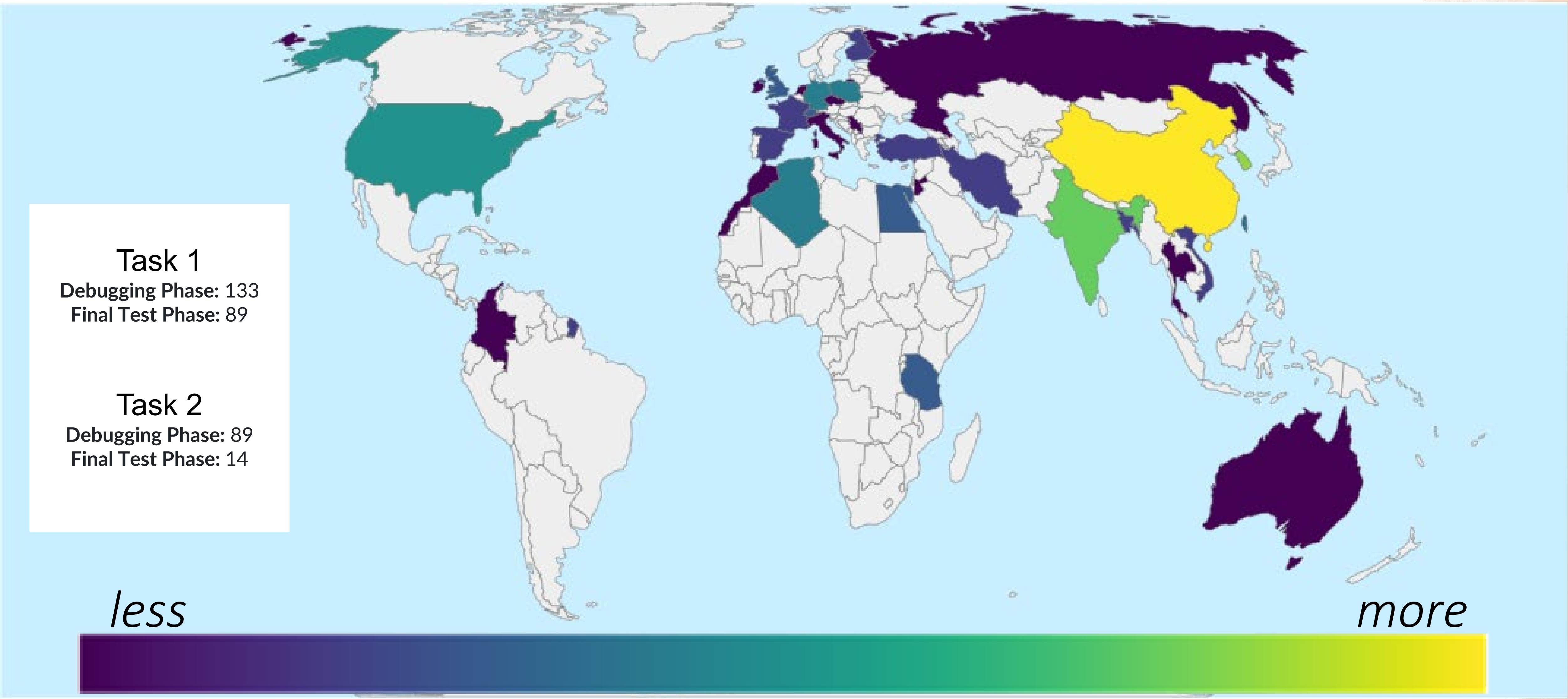

Laura Fenoy
3Shape A/S, Denmark
Laura.Fenoy@3shape.com

Challenge Tasks

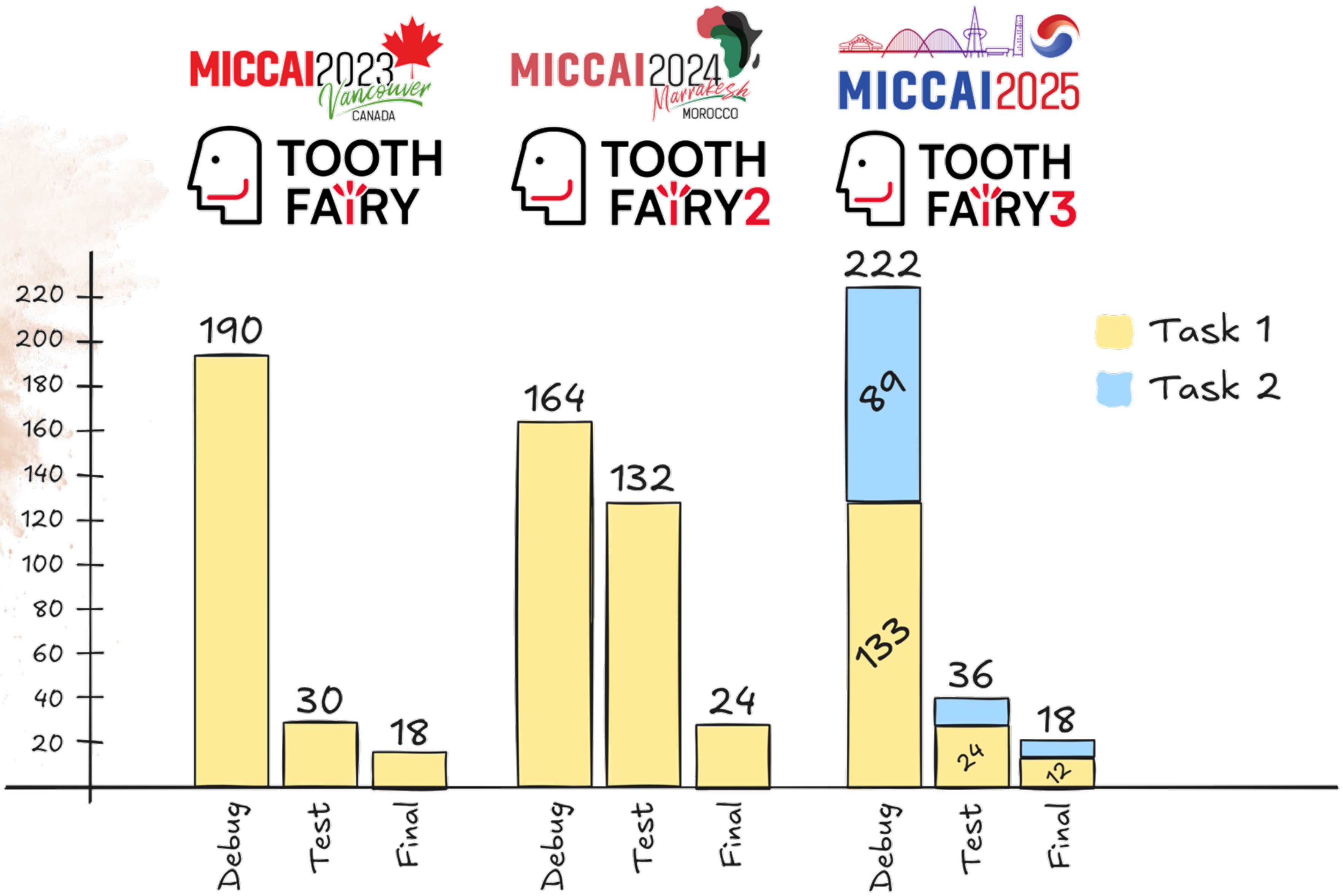
- **Task 1** - Fast Multi-Structure Segmentation in CBCT Volumes
- **Task 2** - Interactive Segmentation of the Inferior Alveolar Canal (IAC) in CBCT Volumes



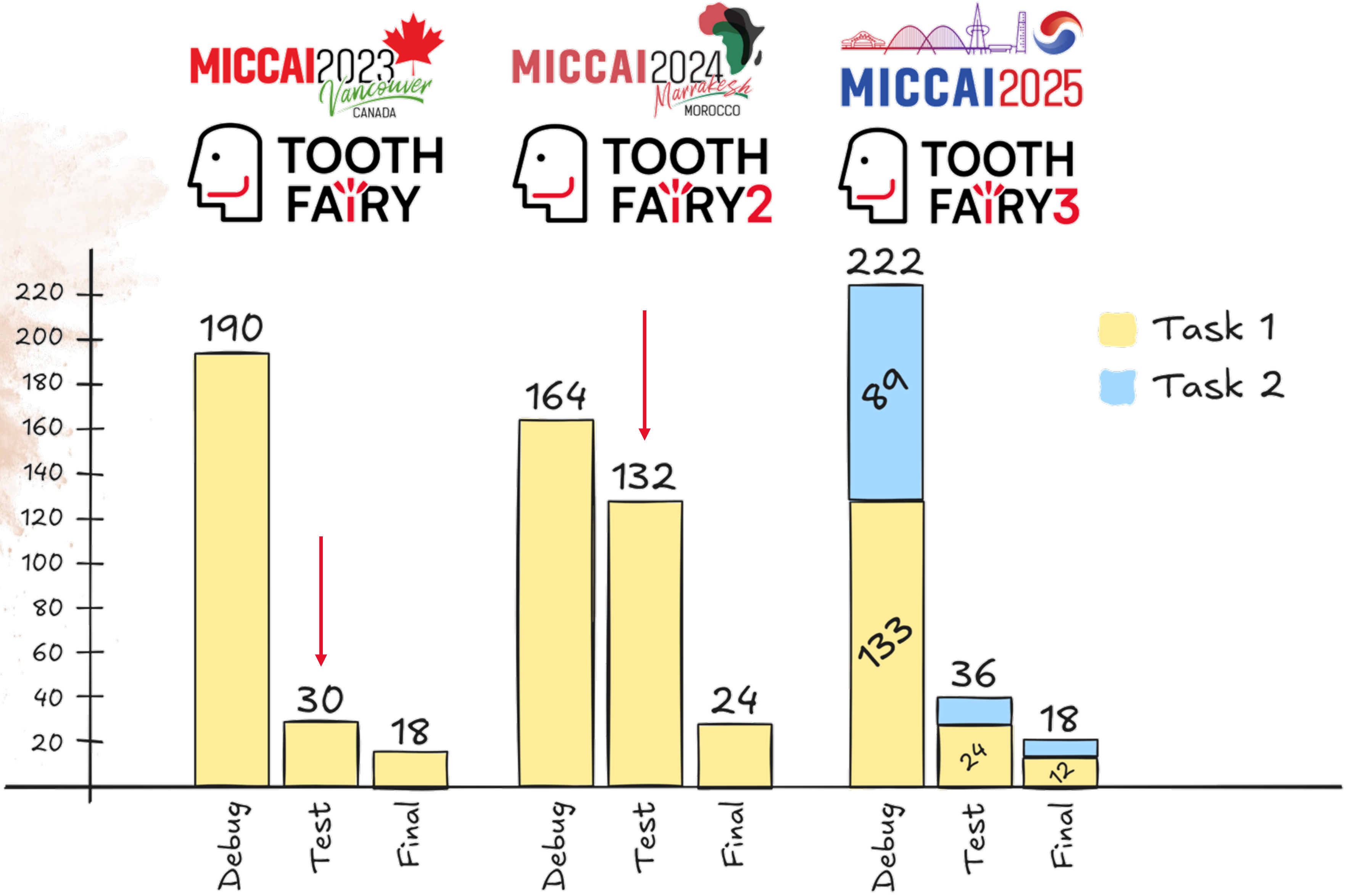
Submissions from all over the world!



But with a decreasing trend … 😞



But with a decreasing trend … 😞



Thanks to all the participants!

Contents lists available at ScienceDirect

Medical Image Analysis

journal homepage: www.elsevier.com/locate/media

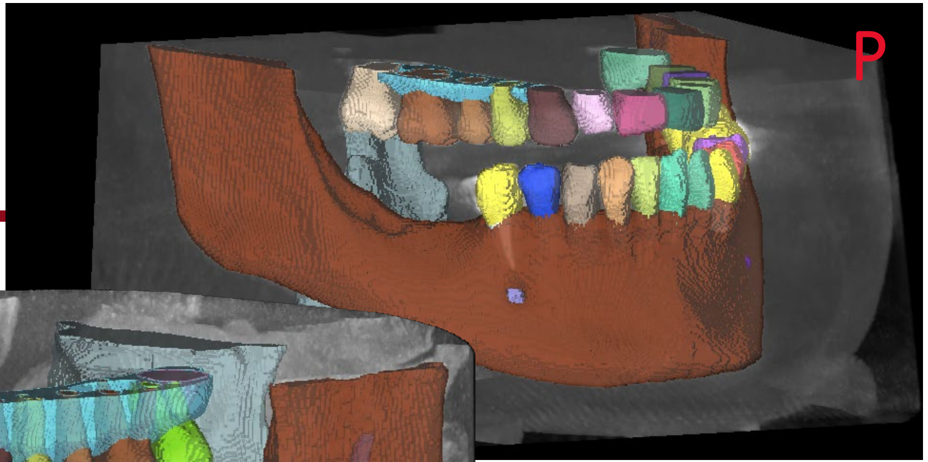
ToothFairy3: Fast and Interactive Multi-class Segmentation in CBCT Volumes

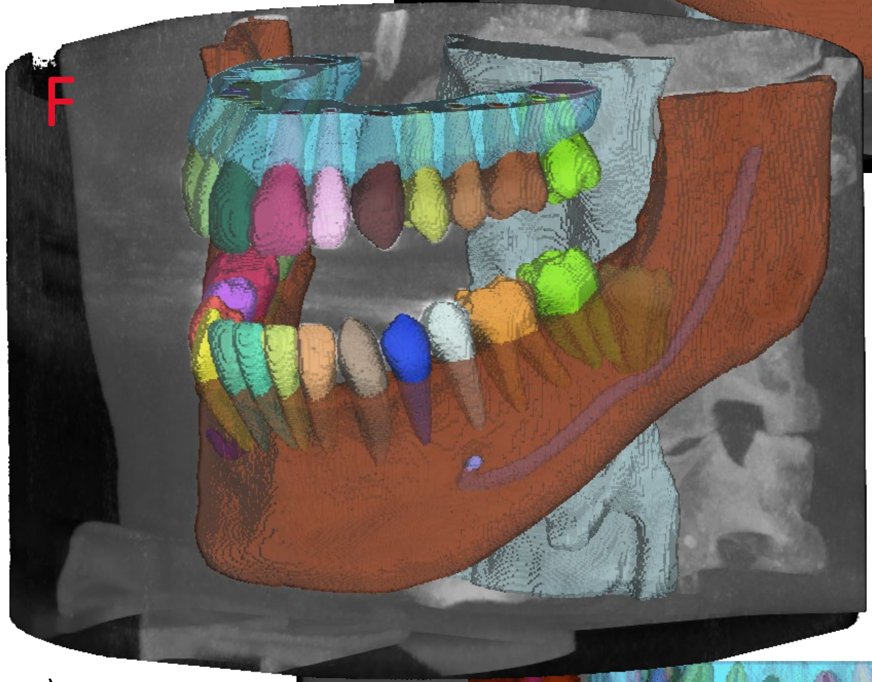
Federico Bolelli[a,1,2,*], Luca Lumetti[a,1,2], Zdravko Marinov[b,2], Niels van Nistelrooij[c,2], Shankeeth Vinayahalingam[c,2], Mattia Di Bartolomeo[a,2], Kevin Marchesini[a,2], Vittorio Pipoli[a], Ettore Candeloro[a], Paolo Minafra[d], Torkan Gholamalizadeh[e], Laura Montesdeoca Fenoy[e], Natasha Hallberg[e], Zhi Qin Tan[f], Yunpeng Li[f,g], Xiatian Zhu[g], Changkai Ji[h], LiSheng Wang[h], Yusheng Liu[h], Fan Xiao[i], Xinrui Huang[j], Xudong Wang[i], Tomasz Szczepański[k], Szymon Płotka[k], Dominic LaBella[l], Esther Yu[l], Jared Robbins[l], Sunwoo Pi[m], Junhyeon Sim[m], Seungha Noh[m], Qingyu Kuang[n,o], Yu Yao[p], Bo Wu[q], Li He[r], Juhan Lee[s], Youngsun Park[t], Alexandre Anesi[a,2], Rainer Stiefelhagen[b,2], Costantino Grana[a,2]

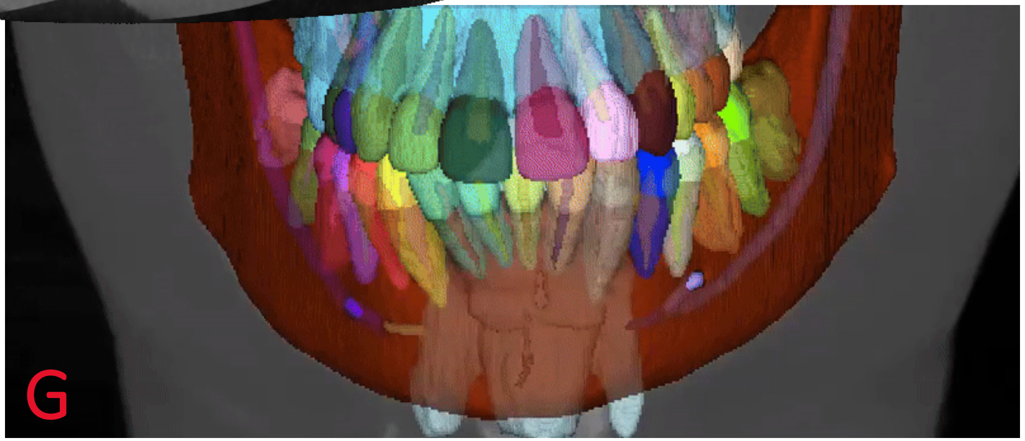
[a] *University of Modena and Reggio Emilia, Department of Engineering "Enzo Ferrari", Italy, Modena*
[b] *Karlsruhe Institute of Technology, Department of Computer Science, Germany, Karlsruhe*
[c] *Radboud University Medical Center, Department of Oral and Maxillofacial Surgery, the Netherlands, Nijmegen*
[d] *Affidea, Italy*
[e] *3Shape A/S, Italy*
[f] *King's College London, Centre for Oral Clinical & Translational Sciences, United Kingdom, London*
[g] *University of Surrey, Surrey Institute for People-Centred AI, United Kingdom, Guildford*
[h] *Shanghai Jiao Tong University, School of Automation and Intelligence, China, Shanghai*
[i] *Ninth People's Hospital Shanghai Jiao Tong University School of Medicine, Department of Oral and Craniomaxillofacial Surgery, China, Shanghai*
[j] *Shanghai Jiao Tong University, School of Electronic Information and Electrical Engineering, China, Shanghai*
[k] *Sano Centre for Computational Personalized Medicine, Medical Imaging and Robotics, Poland, Cracow*
[l] *Duke University Medical Center, Department of Radiation Oncology, USA, Durham*

Dataset

- **Extension** of the **ToothFairy2** Dataset:
  - 52 more volumes;
  - 35 more classes;
  - 3D annotations on all volumes;
- **532** CBCTs for **training**, **50 test** volumes from two external institutions;
- **77** labeled **classes** (42 on GC) **+ clicks**;
- **Challenges**:
  - Large **number** of **classes** (VRAM requirements);
  - Non-uniform **class distribution** and **missing classes**;
  - Considerable **difference in** label **sizes**;
  - **Varying field of view** (P, F and C cases);
  - **Inference time** as a major **constraints**.

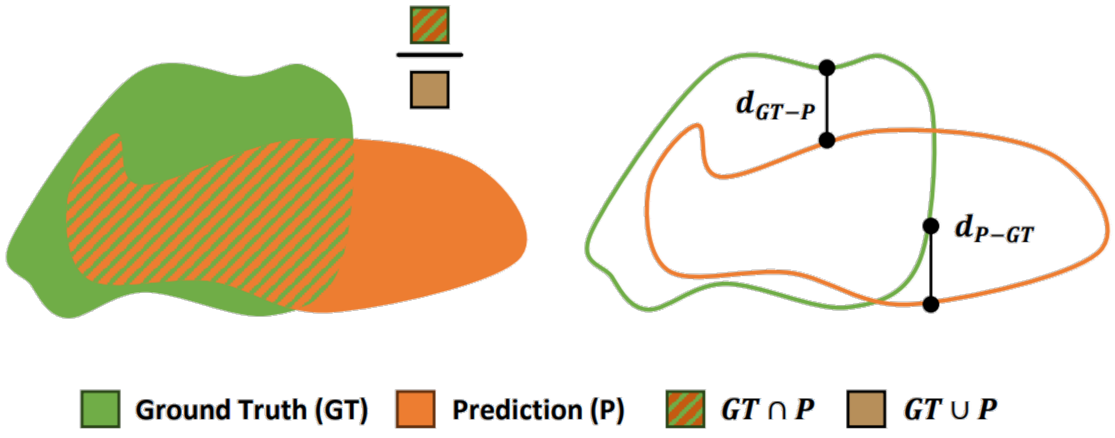




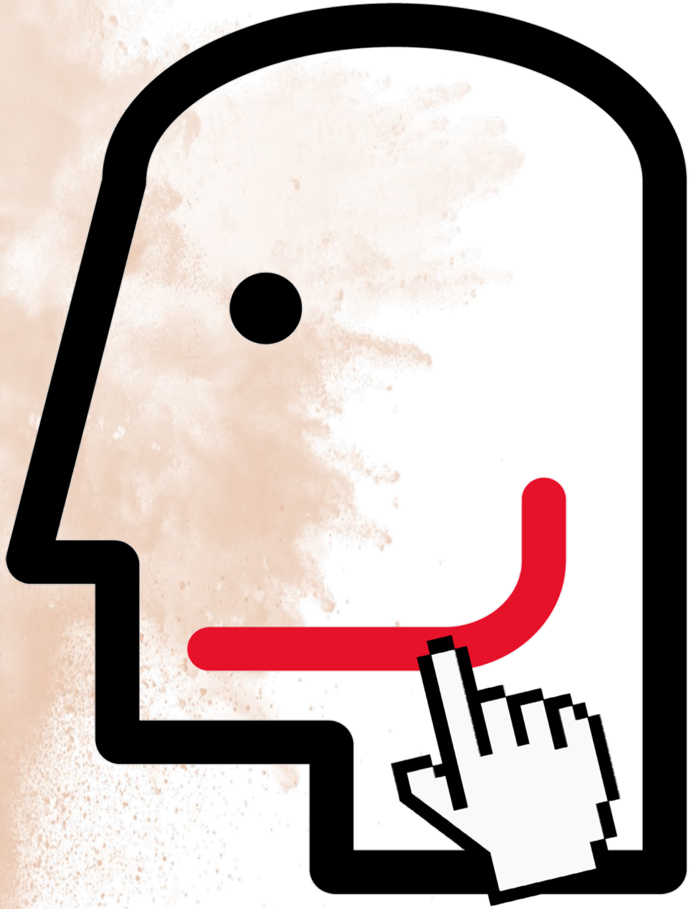


Task 1 - Metrics & Ranking

1. For each class and for each volume, calculate the **Dice score (DSC)** and the **HD95**. Measure also the **inference time** (Time) and the **maximum used memory** (Mem);
2. **Average** the DSC and the HD95 **for each class across all volumes** and compute the **average Time** across **all volumes**.
3. **Rank all** the DSC, HD95, Time, and maximum used memory **independently** (93 rankings);
4. **Average the rankings** obtained at point 3 for each DSC, HD95, and Time to produce the final rank. Average **Time is weighted as much as the number of classes included in the dataset** to balance its importance with respect to the other metrics;
5. If two or more final ranks obtained at point 4 are equal, **Mem will be used to break ties**;
6. If two or more ranks are still equal, it is a tie: the prize will be evenly split.



Task 2 - The Interactive Segmentation Initiative

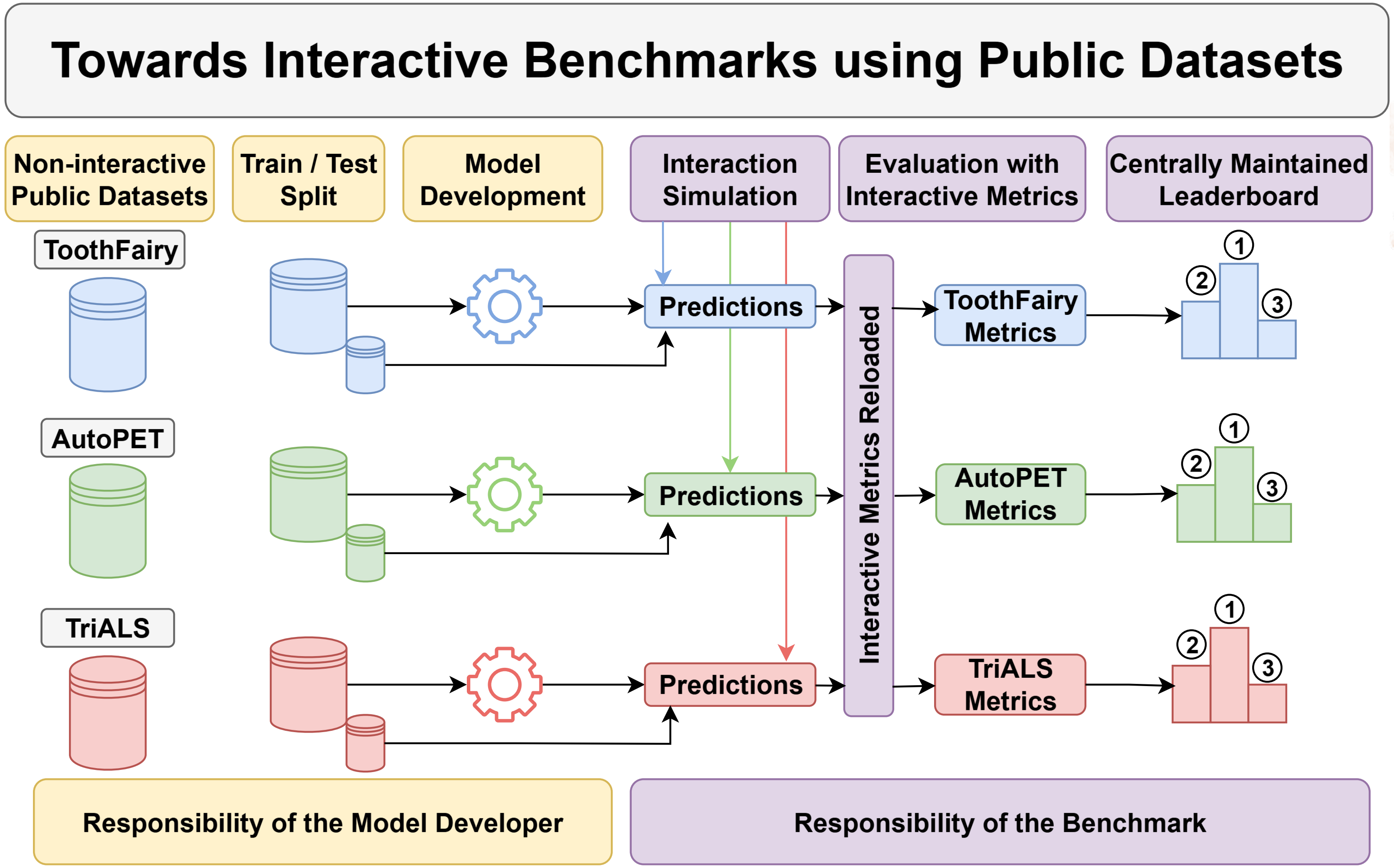


Task 2: **Interactive** Segmentation of Inferior Alveolar Canal:

- Novel task since 2025 (MICCAI + CVPR)
- Use human clicks to help the model

The Interactive Segmentation Initiative

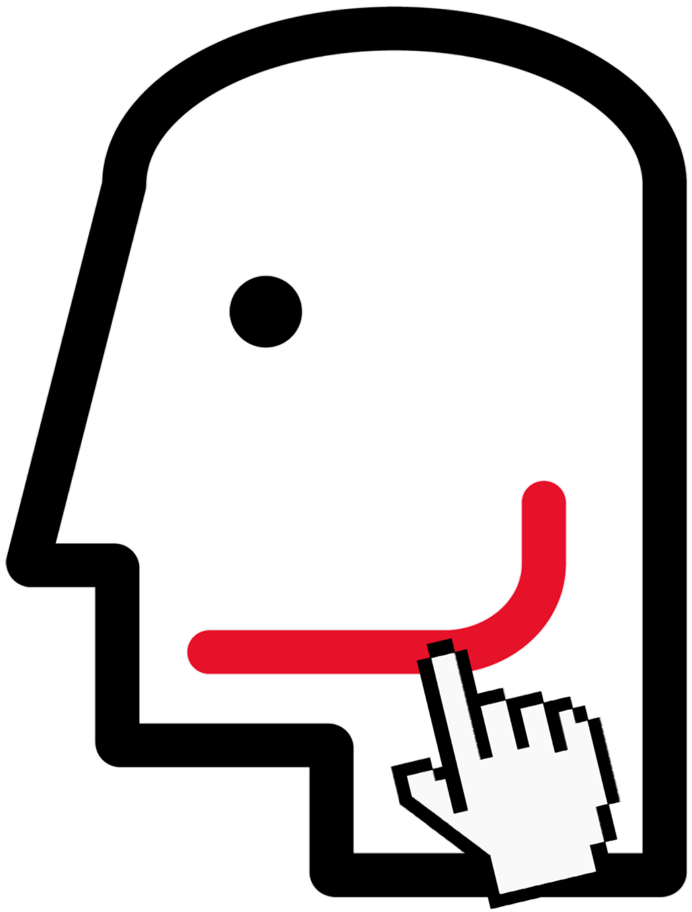
How it all started...



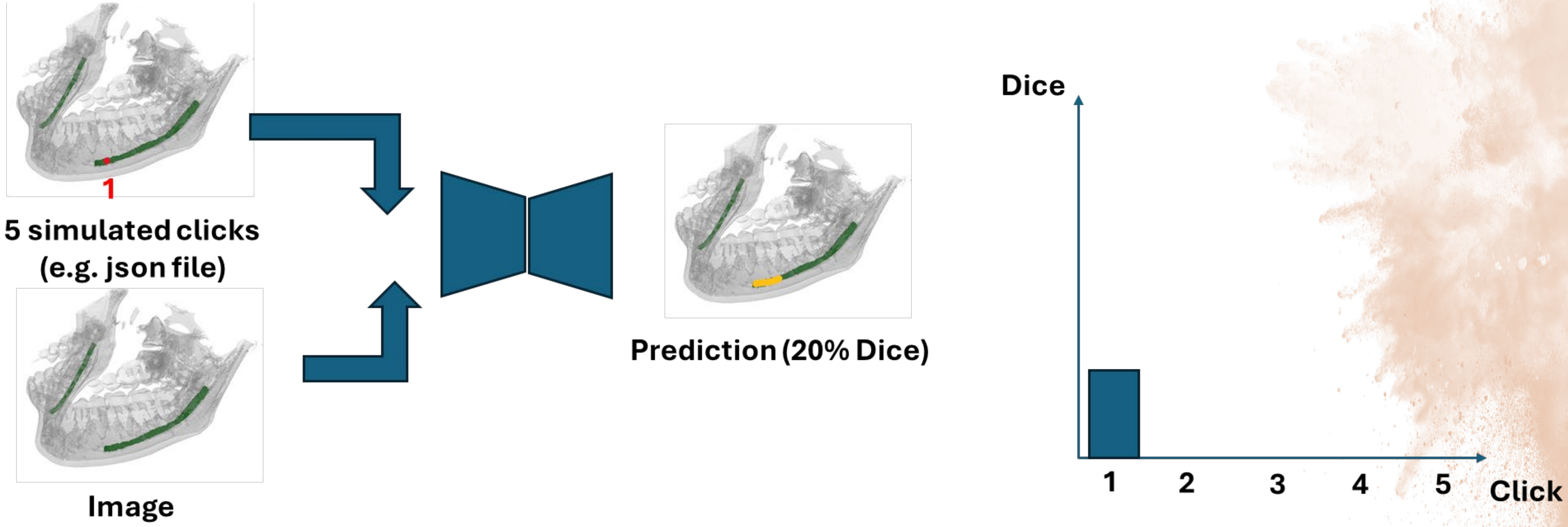
Why interactions?

Reasons for using interactive segmentation models:

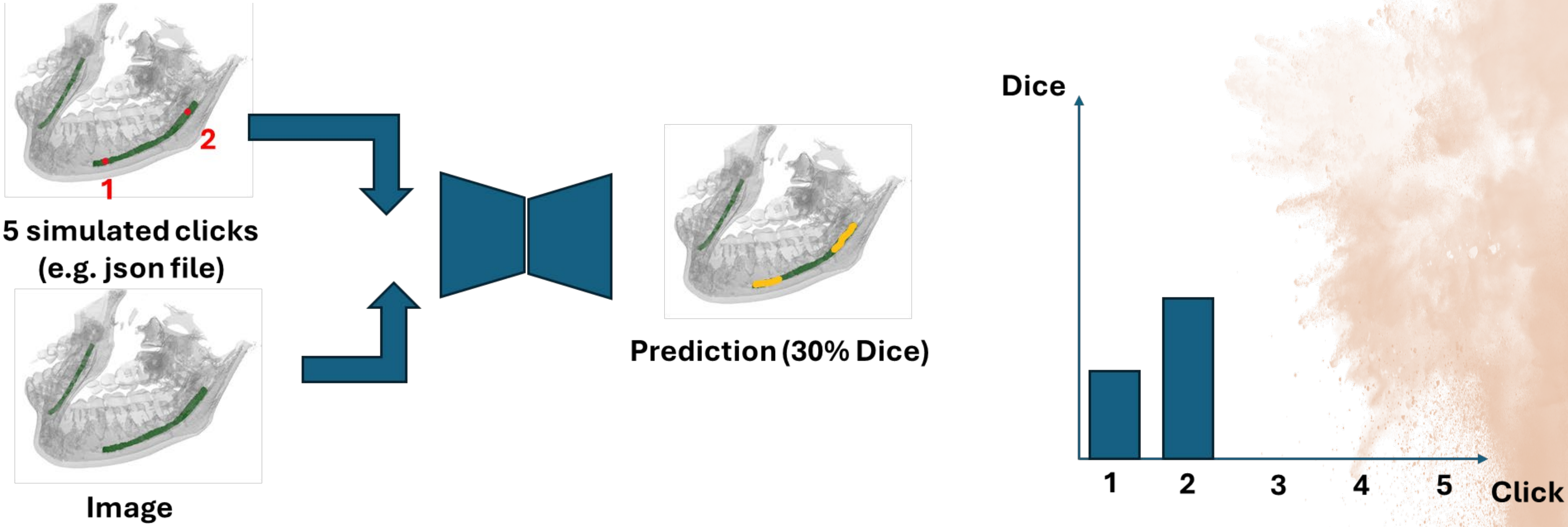
- Click-based annotation tools
- Bridge the gap for difficult tasks
  - Clicks encode expert knowledge
- Make models more trustworthy and reliable
  - Implicit quality control → User continuously corrects model's mistakes



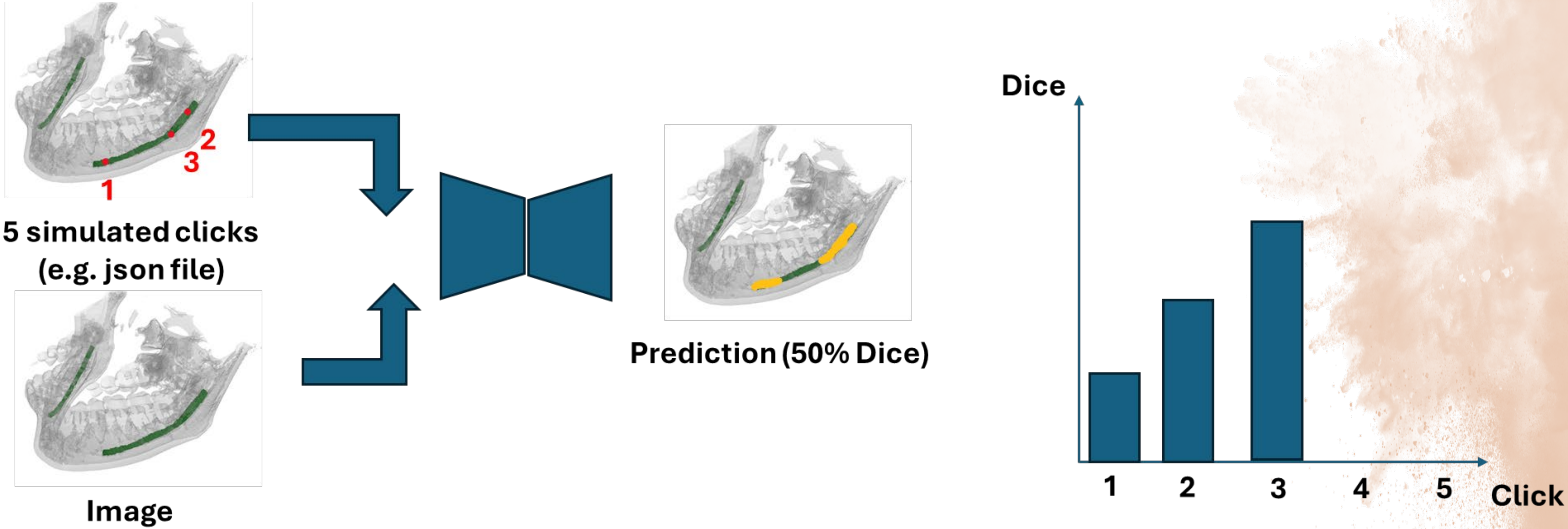
Why interactions?



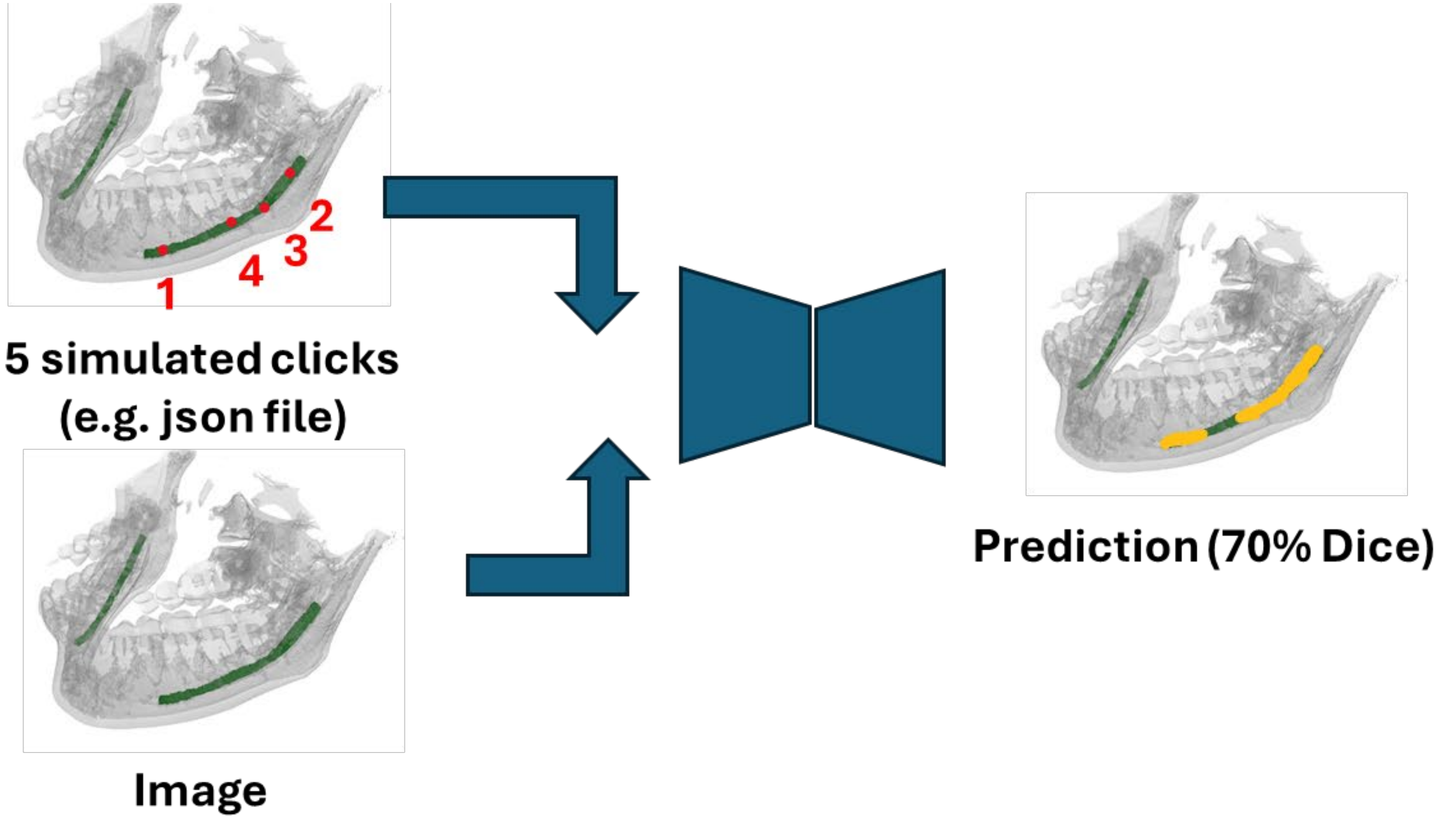
Why interactions?

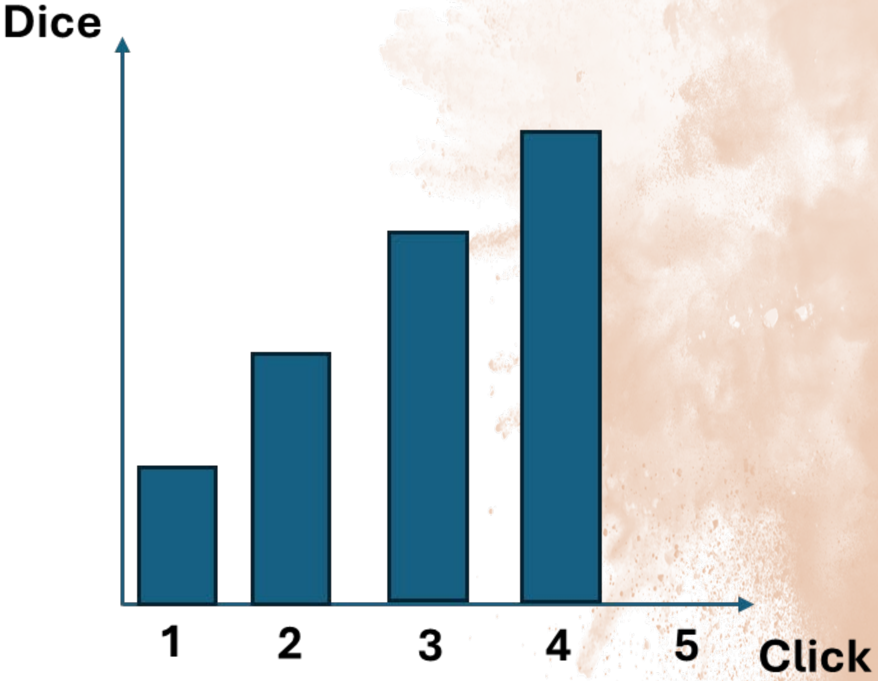


Why interactions?

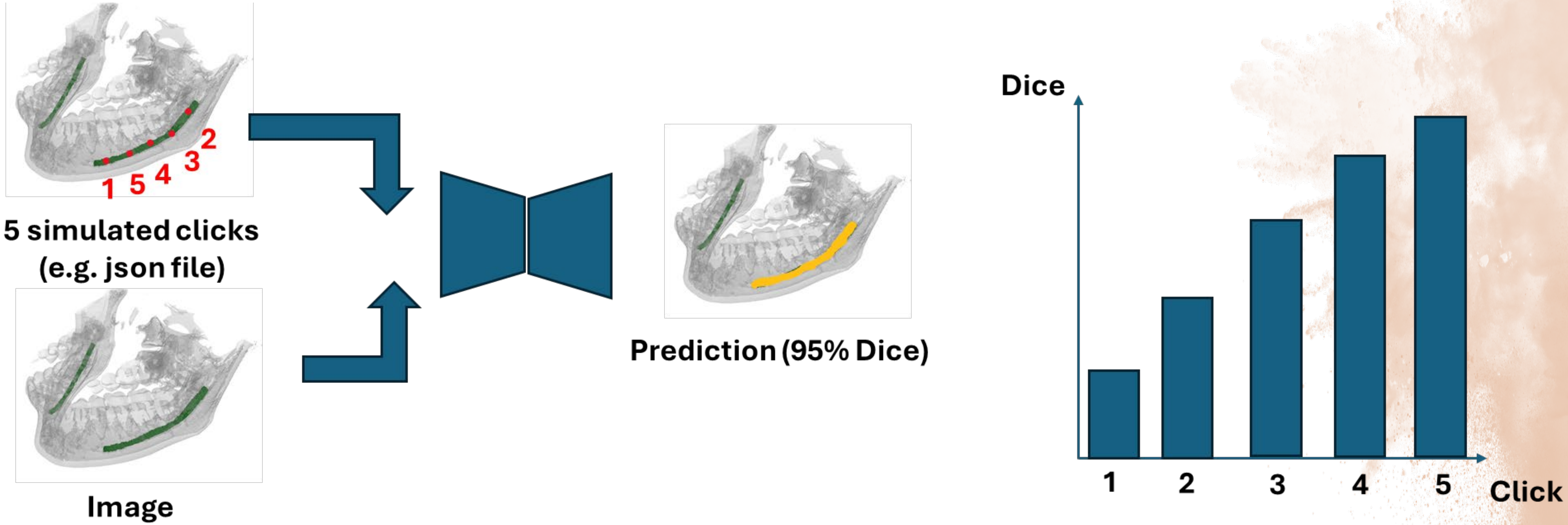


Why interactions?





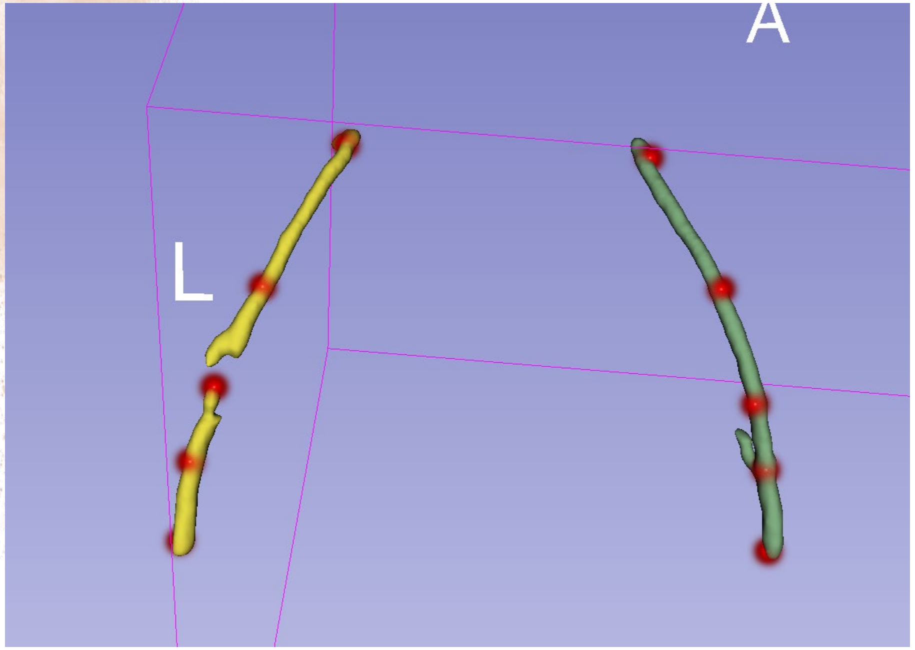
Why interactions?

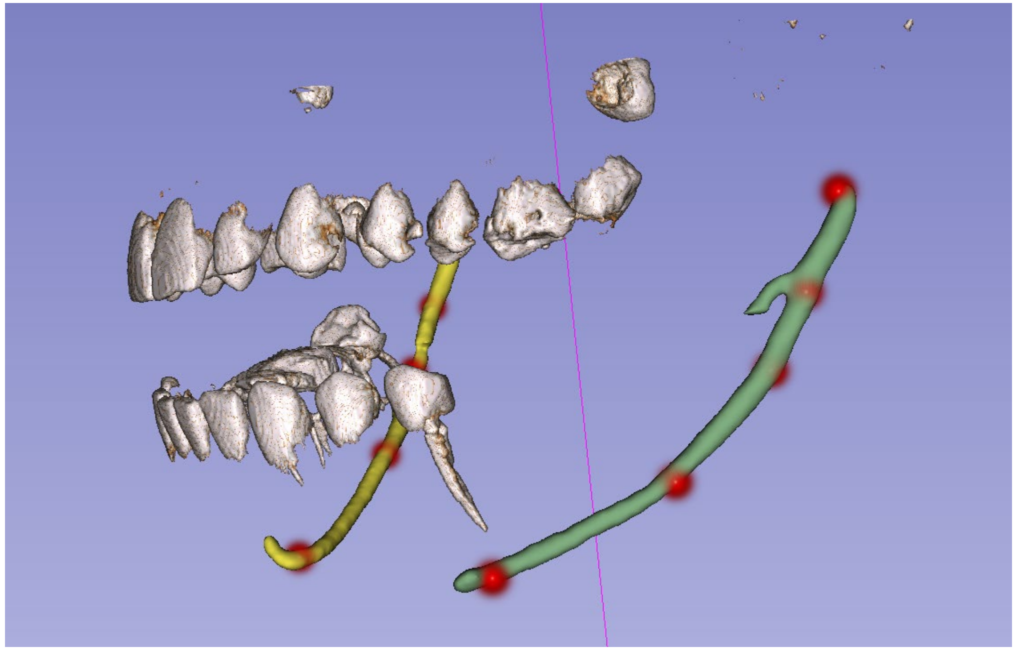


Clicks simulation

Simulation method is the same for all volumes

- **Aim:** method can later be used as an annotation protocol
- Approach:
  - First / last click is always 0-5 voxels from the first / last axial slice of the IAC
  - The other 3 clicks are placed (almost) uniformly across the middle axial slices
  - Each of the 5 clicks is placed in the center of the mask in the slice + small perturbation
- All participants receive the **same clicks → fair evaluation**

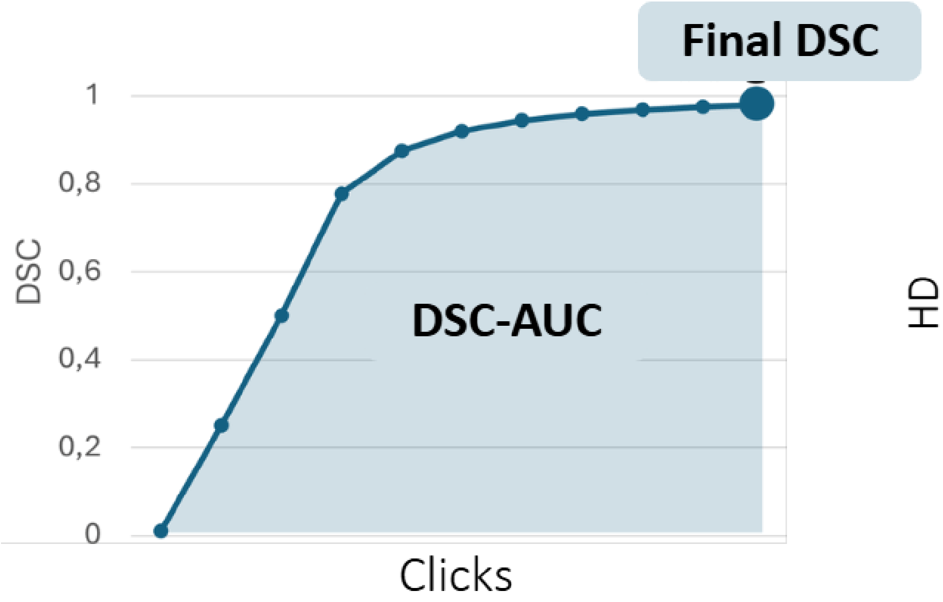


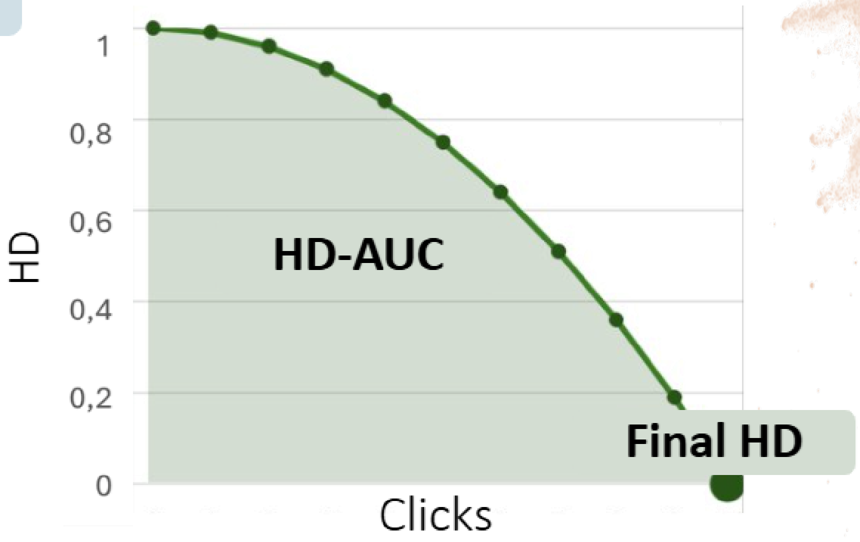


Interactive Metrics

Use same metrics as Task 1, but extend them to be interactive!

- Additionally:
  - Inference time
  - Maximum memory (only for tie breakers)





Task 2 - Ranking

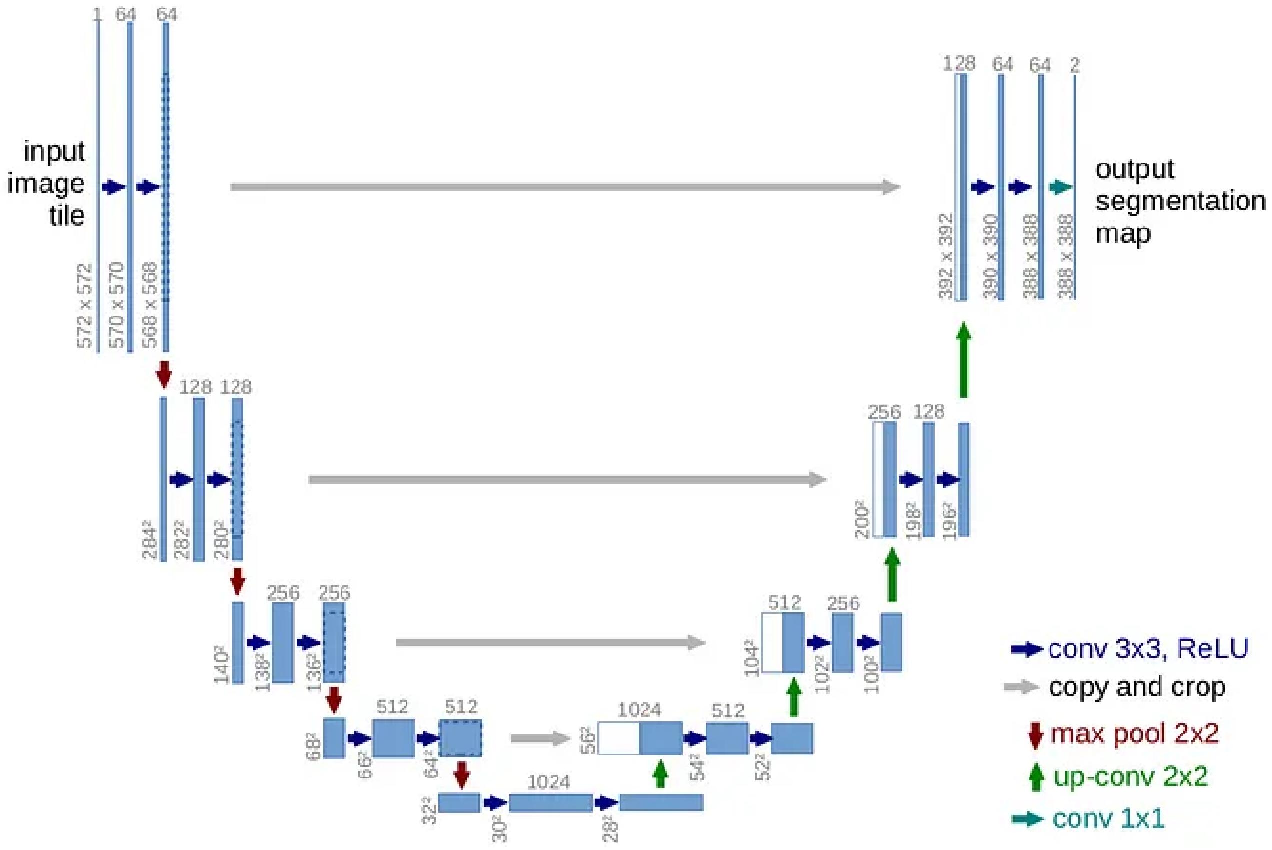
1. **For each branch** of the IAC (left and right) **and** for each **volume**, calculate the four metrics:
   a) DSC@FinalClick;
   b) HD95@FinalClick;
   c) DSC-to-Click AUC, for a fixed number of 5 interaction steps;
   d) HD95-to-Click AUC, for a fixed number of 5 interaction steps;

   The result is 4 metrics * 2 IAC branches, 8 scores in total. Measure also:

   e) average inference time of the five interaction steps (Time);
   f) and (f) maximum used memory (Mem), for all cases;

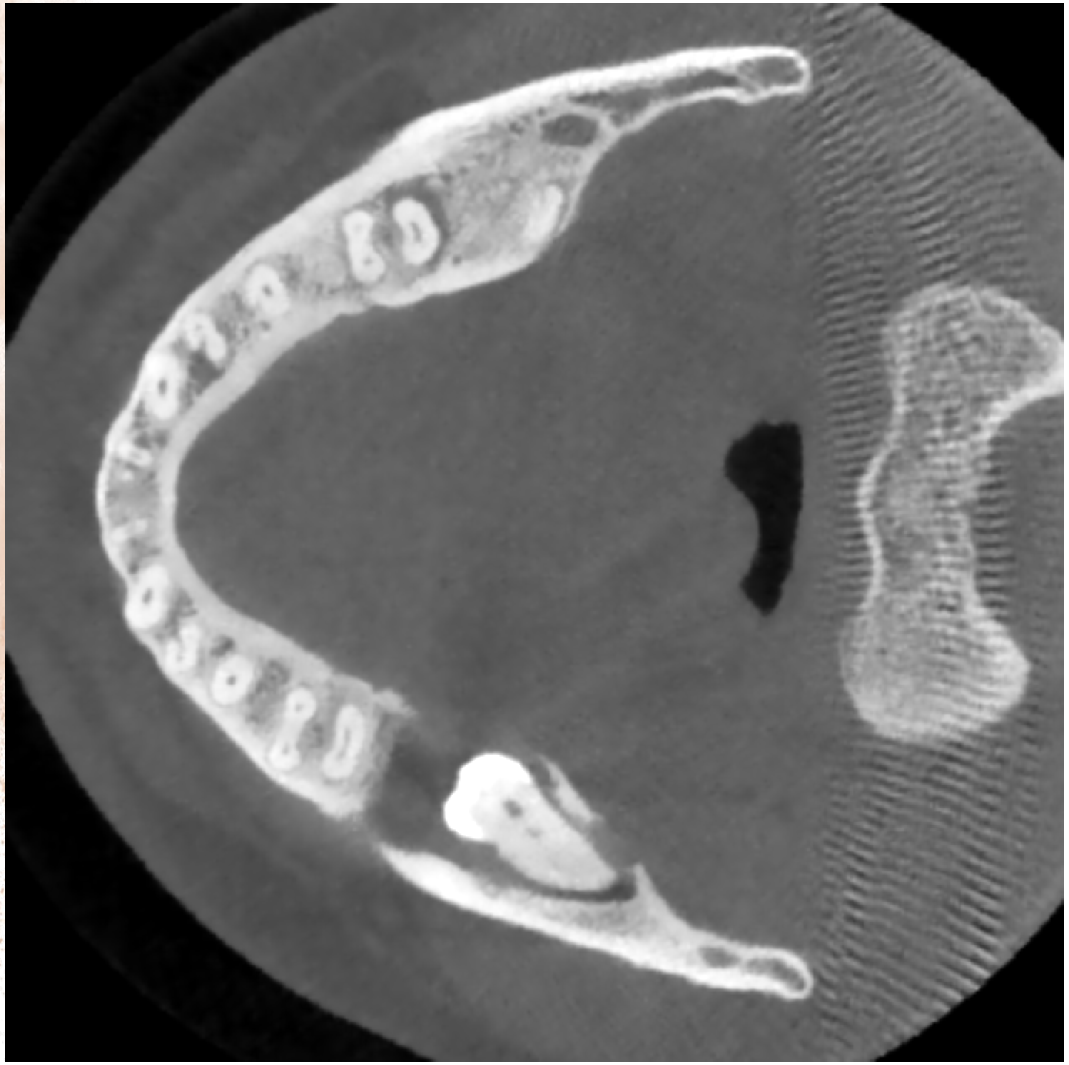
2. **Average DSC- and HD95-based** metrics **for each IAC** branch **across all volumes**.
3. **Rank** all ten metrics independently;
4. **Average the eight rankings** obtained **at point 3** for each DSC- and HD95-based metric **and the rankings obtained for Time** to produce the final rank (9 rankings in total);
5. If two or more final ranks obtained at point 4 are equal, use the **Mem ranking to break ties**;
6. If two or more ranks are still equal, it is a tie: the prize will be evenly split.

Submitted Algorithms



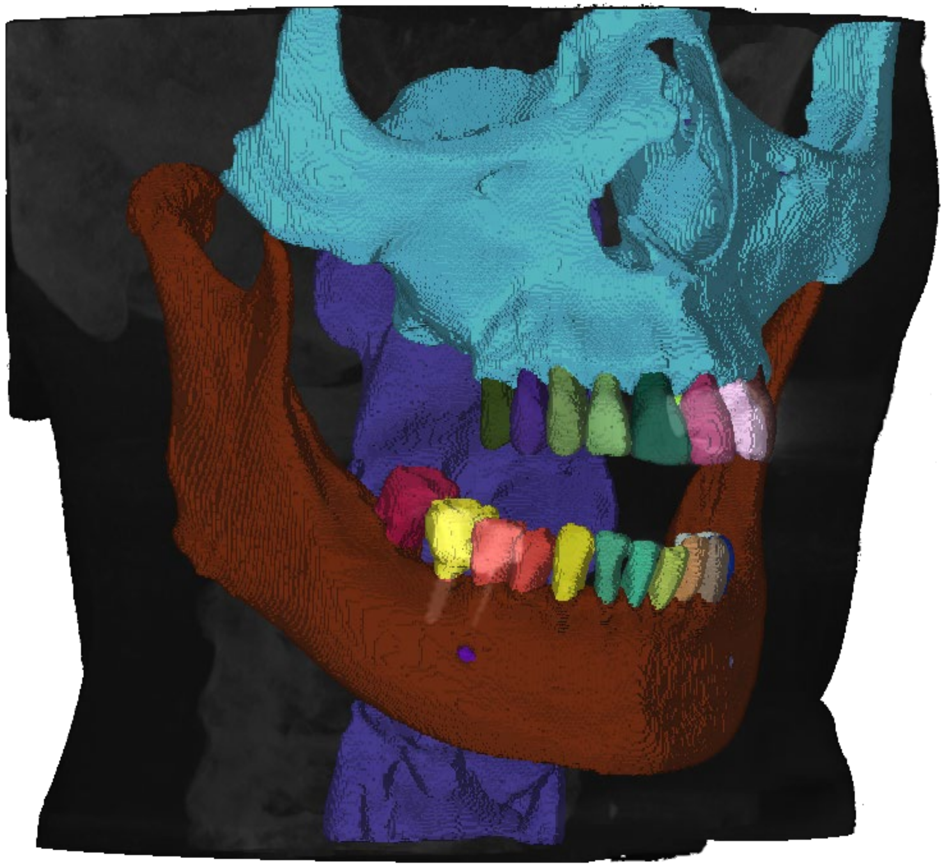
- Many participants used **Unet** with the **nnUnet** Framework;
- Approaches integrating **Mamba** placed in the top section of the final ranking;
- **3D SegResNet** has been employed as well;
- **nnInteractive** and **VISTA** for the interactive task;

Submitted Algorithms



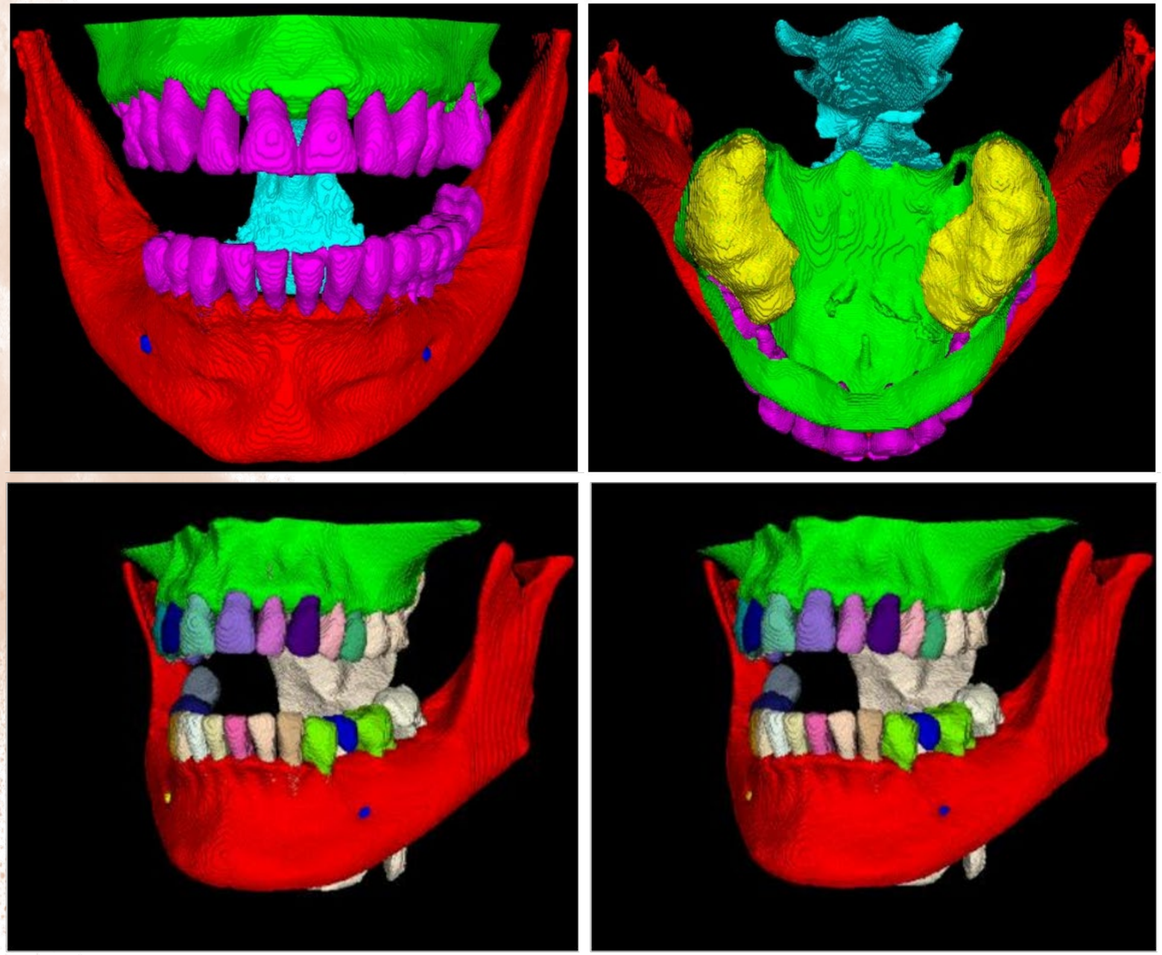
- Many participants used Unet with the nnUnet Framework;
- Moderate use of **augmentations** but almost no augmentation at test time;

Submitted Algorithms



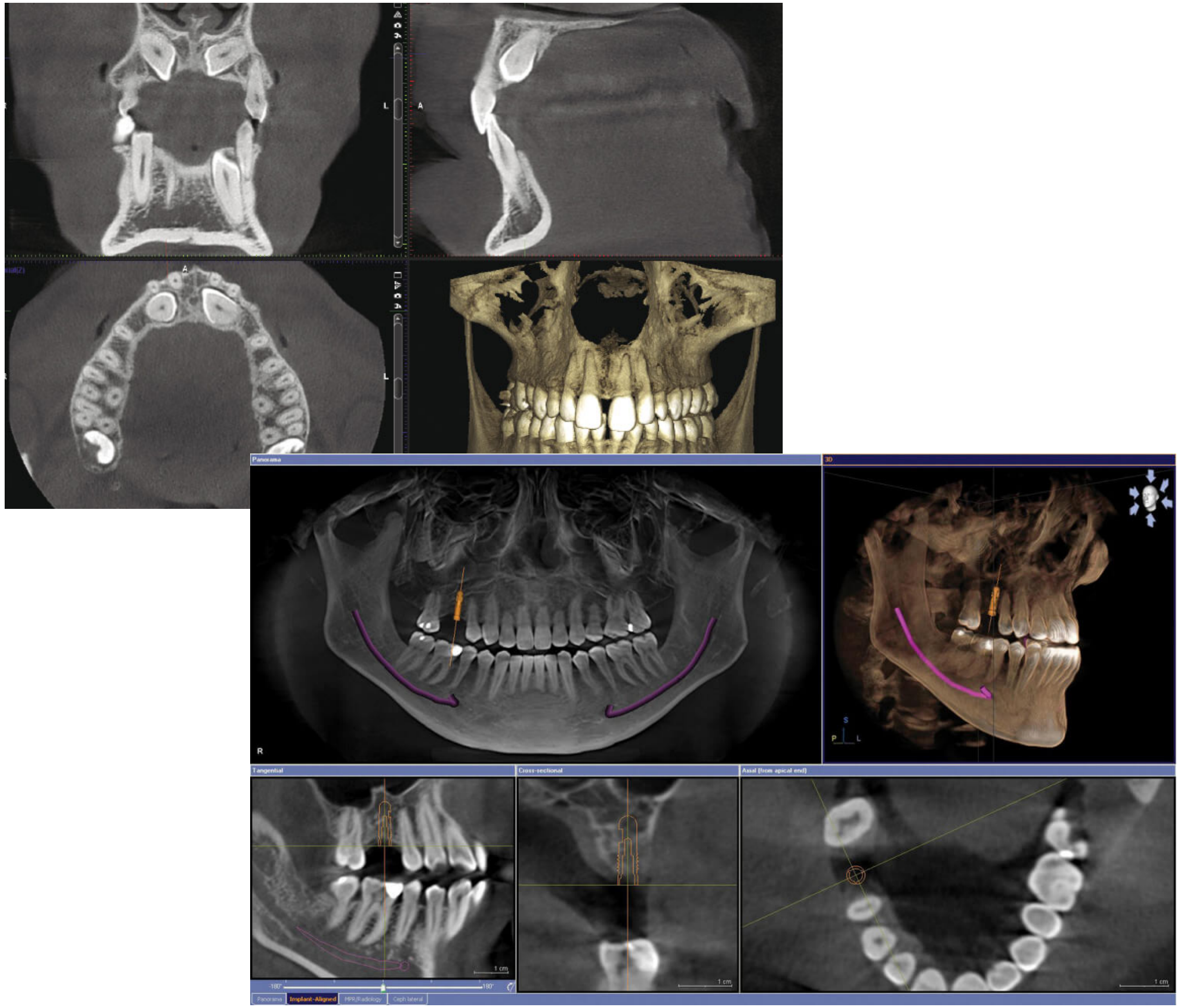
- Many participants used Unet with the nnUnet Framework;
- Moderate use of augmentations;
- **Some** have preprocessed the dataset by **removing training samples**, others have included **only completely labeled sets** in the training;

Submitted Algorithms



- Many participants used Unet with the nnUnet Framework;
- Moderate use of augmentations;
- Some have preprocessed the dataset by removing training samples, others have included only completely labeled sets in the training;
- Some employed a **multi-stage approach**;

Submitted Algorithms



- Many participants used Unet with the nnUnet Framework;
- Moderate use of augmentations;
- Some have preprocessed the dataset by removing training samples, others have included only completely labeled sets in the training;
- Some employed a multi-stage approach;
- As far as we know, **only** one employed **external data from the STSR challenge**;

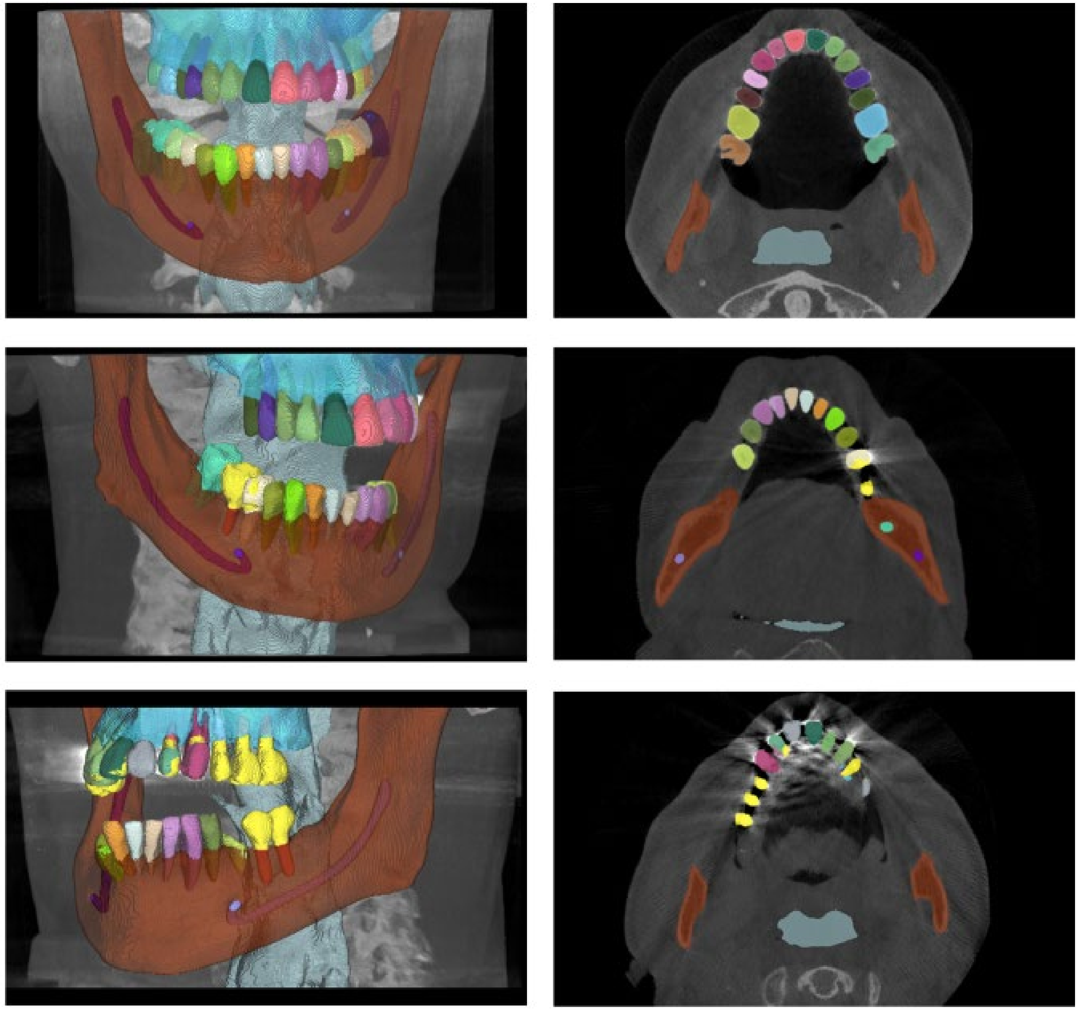
Submitted Algorithms

$$\text{Dice } Loss = 1 - \frac{2 \cdot \sum_i^N (p_i \cdot g_i)}{\sum_{i=1}^N (p_i + g_i)}$$

$$\text{CE } Loss = -\frac{1}{N} \sum_{i=1}^{N} [\, y_i \cdot \log(p_\mathrm{i}) + (1 - y_i) \cdot \log(1 - p_i)]$$

- Many participants used Unet with the nnUnet Framework;
- Moderate use of augmentations;
- Some have preprocessed the dataset by removing training samples, others have included only completely labeled sets in the training;
- Some employed a multi-stage approach;
- As far as we know, only one employed external data from the STSR challenge;
- Most sticked on a **combination** of **Dice** and **Cross-entroy** loss;

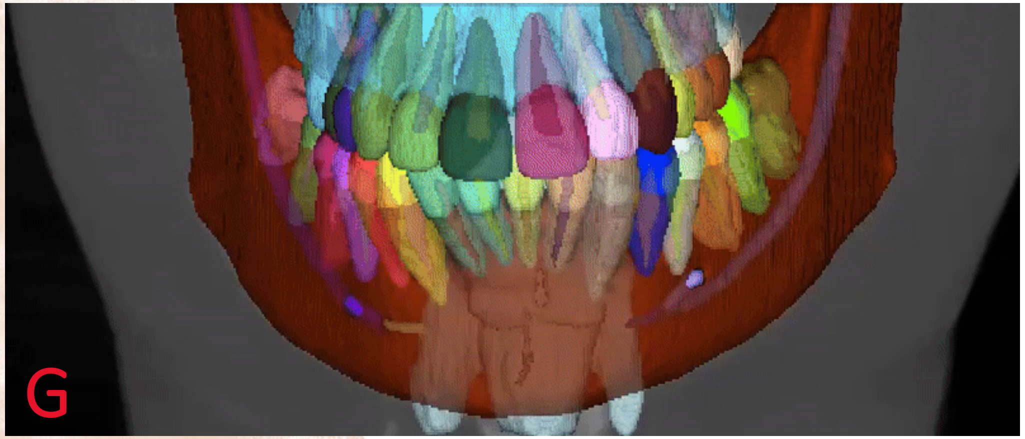
Submitted Algorithms

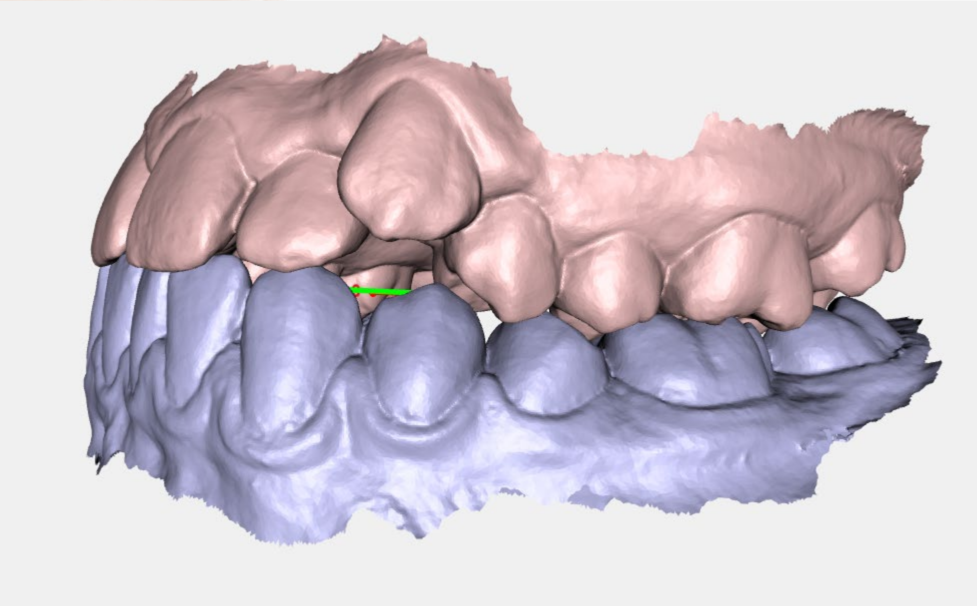


- Most of the participants used Unet with the nnUnet Framework;
- Moderate use of augmentations;
- Some have preprocessed the dataset by removing training samples;
- Some employed a multi-stage approach;
- Some have preprocessed the dataset by removing training samples, others have included only completely labeled sets in the training;
- Some employed a multi-stage approach;
- As far as we know, only one employed external data from the STSR challenge;
- Many have designed solutions to **filter output predictions**;
- The used of **ensemble** has been reduced;

TOOTH FAIRY4

MICCAI 2026





Sponsor

see
through

Task 1 – Top Scorers!

Task 1 – Top Scorers!

帆 萧

ff741333

(Black_Myth)

Task 1 – Top Scorers!

Changkai Ji
Changkkk
(sjtu_eiee_2-426lab)

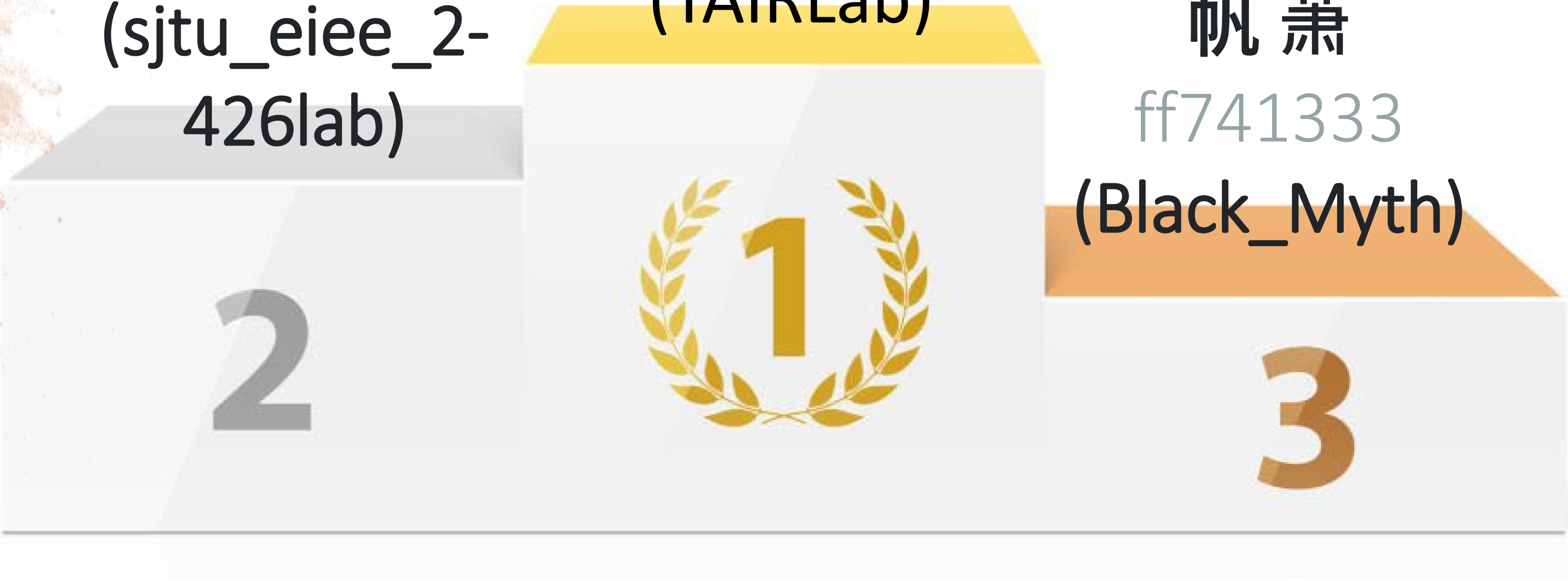
帆 萧
ff741333
(Black_Myth)

Task 1 – Top Scorers!

Changkai Ji
Changkkk
(sjtu_eiee_2-426lab)

Zhi Qin Tan
thuengitmoi
(TAIRLab)

帆 萧
ff741333
(Black_Myth)



Final Results

TAIR-Lab	**Runtime(s)**	**Dice**	**Dice Rank**	**HD95**	**HD95 Rank**	**Rank**
Aggregated	40,58	0,84	2,4	38,17	3,0	3,1
sjtu_eiee_2-426lab	**Runtime(s)**	**Dice**	**Dice Rank**	**HD95**	**HD95 Rank**	**Rank**
Aggregated	17,46	0,77	4,9	94,77	5,3	3,7
Black_Myth	**Runtime(s)**	**Dice**	**Dice Rank**	**HD95**	**HD95 Rank**	**Rank**
Aggregated	90,04	0,85	2,0	33,23	2,3	3,8
SMIR	**Runtime(s)**	**Dice**	**Dice Rank**	**HD95**	**HD95 Rank**	**Rank**
Aggregated	30,30	0,76	4,9	96,40	5,3	4,1
DLaBella29	**Runtime(s)**	**Dice**	**Dice Rank**	**HD95**	**HD95 Rank**	**Rank**
Aggregated	74,38	0,78	5,0	43,20	3,4	4,5
KiRyum_Prince	**Runtime(s)**	**Dice**	**Dice Rank**	**HD95**	**HD95 Rank**	**Rank**
Aggregated	97,68	0,82	3,2	50,32	3,7	5,0
gagaha	**Runtime(s)**	**Dice**	**Dice Rank**	**HD95**	**HD95 Rank**	**Rank**
Aggregated	38,42	0,58	6,9	125,10	6,7	5,5
medlab	**Runtime(s)**	**Dice**	**Dice Rank**	**HD95**	**HD95 Rank**	**Rank**
Aggregated	79,54	0,47	7,3	138,68	7,0	6,8

Task 2 – Top Scorers!

Task 2 – Top Scorers!

Changkai Ji
Changkkk
(sjtu_eiee_2-426lab)

Task 2 – Top Scorers!

帆萧
ff741333
(Black_Myth)

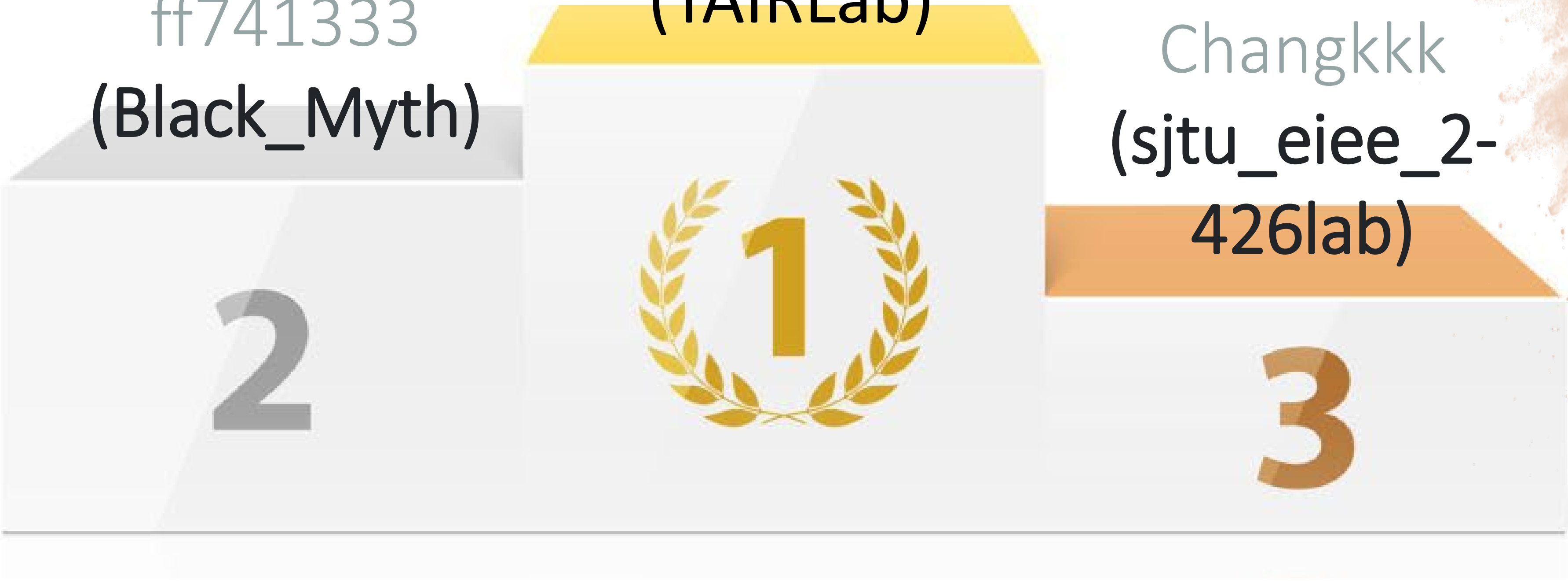
Changkai Ji
Changkkk
(sjtu_eiee_2-426lab)

Task 2 – Top Scorers!

帆 萧
ff741333
(Black_Myth)

Zhi Qin Tan
thuengitmoi
(TAIRLab)

Changkai Ji
Changkkk
(sjtu_eiee_2-426lab)



Final Results

TAIR Lab	**Runtime(s)**	**Dice_L IAC_AUC**	**Dice_L IAC_Final**	**Dice_R IAC_AUC**	**Dice_R IAC_Final**	**HD95_L IAC_AUC**	**HD95_L IAC_Final**	**HD95_R IAC_AUC**	**HD95_R IAC_Final**	**Rank**
Aggregated	100,64	4,34	0,87	4,31	0,86	11,29	2,26	10,2	2,04	
Rank	4	1	1	2	2	1	1	1	2	1.66
BlackMyth	**Runtime(s)**	**Dice_L IAC_AUC**	**Dice_L IAC_Final**	**Dice_R IAC_AUC**	**Dice_R IAC_Final**	**HD95_L IAC_AUC**	**HD95_L IAC_Final**	**HD95_R IAC_AUC**	**HD95_R IAC_Final**	**Rank**
Aggregated	168,42	4,3	0,86	4,32	0,87	15,05	2,54	14,55	2,03	
Rank	6	2	2	1	1	2	2	2	1	2.11
changkkk	Runtime(s)	Dice_L IAC_AUC	Dice_L IAC_Final	Dice_R IAC_AUC	Dice_R IAC_Final	HD95_L IAC_AUC	HD95_L IAC_Final	HD95_R IAC_AUC	HD95_R IAC_Final	Rank
Aggregated	16,09	3,79	0,76	3,83	0,77	201,75	40,35	131,42	26,28	
Rank	1	3	3	3	3	4	5	4	5	3.44
DLaBella	**Runtime(s)**	**Dice_L IAC_AUC**	**Dice_L IAC_Final**	**Dice_R IAC_AUC**	**Dice_R IAC_Final**	**HD95_L IAC_AUC**	**HD95_L IAC_Final**	**HD95_R IAC_AUC**	**HD95_R IAC_Final**	**Rank**
Aggregated	152,02	2,98	0,75	2,97	0,74	19,27	4,82	17,82	4,45	
Rank	5	5	4	5	4	3	3	3	3	3.88
gagaha	**Runtime(s)**	**Dice_L IAC_AUC**	**Dice_L IAC_Final**	**Dice_R IAC_AUC**	**Dice_R IAC_Final**	**HD95_L IAC_AUC**	**HD95_L IAC_Final**	**HD95_R IAC_AUC**	**HD95_R IAC_Final**	**Rank**
Aggregated	26,09	2,64	0,68	2,63	0,7	585,8	11,93	588,75	9,63	
Rank	2	6	5	6	5	6	4	6	4	4.88
ATTIAC	**Runtime(s)**	**Dice_L IAC_AUC**	**Dice_L IAC_Final**	**Dice_R IAC_AUC**	**Dice_R IAC_Final**	**HD95_L IAC_AUC**	**HD95_L IAC_Final**	**HD95_R IAC_AUC**	**HD95_R IAC_Final**	**Rank**
Aggregated	83,41	3,18	0,64	3,18	0,64	536,33	107,25	580,63	116,13	
Rank	3	4	6	4	6	5	6	5	6	5