From Prediction to Prompt: Leveraging nnU-Net Outputs to Guide SAM for Active Learning in 3D Dental Segmentation

Nicolas Martin[1,2][0000−0002−2788−1042], Jean-Pierre Chevallet[2][0000−0002−5945−9444], and Philippe Mulhem[2][0000−0002−3245−6462]

[1] PEEKTORIA, Grenoble, France
nicolas.martin@peektoria.com
[2] Univ. Grenoble Alpes, CNRS, Grenoble INP**, LIG, Grenoble, France
{jean-pierre.chevallet,philippe.mulhem}@univ-grenoble-alpes.fr

Abstract. To enhance annotation efficiency in 3D dental Cone Beam Computed Tomography (CBCT) image segmentation, this paper explores an active learning (AL) approach that leverages nnU-Net predictions to generate prompts for a specialized 3D Segment Anything Model (SAM). The objective is to minimize the annotation burden without relying on prompts during the inference phase. First, our experiments showed that AL offers similar segmentation performance with less than 20% of the original annotations. Second, random selection offers similar results than more complex sampling method with less more computing demand. Third, the predictions of nnU-Net on unannotated images provided effective prompts for the SAM model specialized in 3D medical images (i.e., SAM-Med3D). Combining these two approaches reduced the required amount of manual annotation by up to 50%. This paper paves the way for more easily obtaining new annotated datasets in the dental domain while simultaneously training a segmentation model, by leveraging SAM-like models.

Keywords: Active Learning · nnU-Net · Segment Anything · Segmentation · 3D dental CBCT.

1 Introduction

Organ segmentation is a highly active research area within computer vision for medical imaging. In the dental domain, the widespread adoption of imaging technologies like Cone Beam Computed Tomography (CBCT) and panoramic X-rays in clinical settings has underscored the critical need for automated solutions to effectively leverage this information. Precisely segmenting anatomical structures (e.g., teeth) is often an essential step for robust computer-aided detection systems [19]. Despite dental issues affecting a significant global population, dedicated computer vision tools for dentistry remain less developed, largely due

** Institute of Engineering Univ. Grenoble Alpes

to a scarcity of annotated datasets outside the scope of recent MICCAI challenges (e.g., ToothFairy [2], 3DTeethSeg [1]). As highlighted by these challenges, accurately segmenting dental organs, particularly in 3D images, presents a major difficultly, and currently often relies on adaptations of the nnU-Net model [10] specialized for dental datasets (e.g., [11, 31]).

On the other hand, inspired by the success of large language models (LLMs), which are pre-trained using self-supervised learning (SSL) on very large datasets and fine-tuned to follow instructions (prompt-based models) [24], the Segment Anything Model (SAM) [15] has been proposed. The initial SAM model [15] have been trained on approximately one billion image-mask pairs. This attention-based model is designed to be applied to any image, aiming to address nearly any segmentation task. Despite this initial assertion, these models are unable to correctly segment specific image types, such as medical images [9], necessitating fine-tuning (e.g., MedSAM [21, 22], SAM-Med3D [30]). Furthermore, such SAM-like models heavily rely on "prompts" (e.g., bounding boxes, points), which serve as strong indicators for defining the image region to be segmented [15]. In practice, in daily clinical routine, the introduction of SAM is barely impossible, as it requires a precise bounding box or multiple points to perform accurate segmentation [16]. Consequently, it remains essential to train segmentation models on annotated data. This paper investigates the integration of Active Learning (AL) with SAM-like models to reduce the expert annotation burden in 3D dental segmentation tasks.

2 Related work

Prior studies on active learning (AL) have shown that not all data points are equally informative [25]. Their annotations can significantly influence both the training process and the final performance of the model [28]. Selecting the most informative images should be more beneficial to model performance than random selection of images [32]. This assumption has led to the development of numerous AL methods designed to select the most informative samples for annotation [25, 28].

In the dental domain, obtaining images for diagnostic or archival purposes has become standard practice, leading to the availability of large datasets [33]. However, these datasets are rarely annotated [5]. Thus, selecting the most informative images using AL methods presents a valuable opportunity to significantly alleviate the annotation workload for experts, thereby promoting the creation of more efficient medical tools based on deep learning algorithms: see [3] for a review of AL for medical images. In 3D dental domain, Huang et al. [8] and Jung et al. [14] showed that AL can improve the segmentation performance.

In the context of 2D medical images, Li et al. [18] explored the combination of nnU-Net and a generic SAM model. SAM predictions are directly integrated into the nnU-Net architecture as an external module to enhance segmentation performance. Stock el al. [29] investigated the integration of nnU-Net with SAM for 3D images. However, due to computational constraints, their approach is

applied in a 2D slice-by-slice manner. On the other hand, interactive annotation relying on SAM-like models have been proposed: Isensee et al. [12] trained nnU-Net model on 120+ 3D datasets to produce segmentation masks using prompts.

In this paper, we explore the integration of active learning with promptable segmentation models (e.g., SAM-like models). To the best of our knowledge, no prior study has investigated the combination of nnU-Net and SAM for 3D dental image segmentation within an active learning framework.

3 Method

This paper investigates two key aspects: (1) the impact of various AL sampling strategies on 3D image segmentation performance and (2) the performance of SAM-like models (i.e., SAM-Med3D [30]) when integrated with nnU-Net-derived prompts during AL training.

3.1 Datasets

The dataset ToothFairy2 [2] have been used in the following experiments. It is composed of 480 Cone Beam Computed Tomography (CBCT) with 42 classes. To reduce computational complexity and focus our analysis, the original anatomical classes were re-categorized into the following 6 broader classes for segmentation:

- Background
- Jawbones: Lower and Upper
- Inferior Alveolar Canal (IAC): Left and Right
- Sinus: Left and Right
- Pharynx
- Teeth (32 classes originally)

Due to their sparse representation in the dataset, the Bridge, Crown, Implant, and NA classes were excluded from segmentation and assigned to the background.

3.2 Metrics

The segmentation performance was evaluated using the Dice Similarity Coefficient (DSC in %). For a given image i and a specific target class C, let Sg_i^C represent the set of pixels assigned to class C in the ground truth segmentation, and Sa_i^C denote the corresponding set of pixels predicted by the automatic segmentation model. The $Dice$ score for class C on image i quantifies the overlap between these two segmentations and is defined by equation (1):

$$Dice(Sg_i^C, Sa_i^C) = \frac{2|Sg_i^C \cap Sa_i^C|}{|Sg_i^C| + |Sa_i^C|} \tag{1}$$

DSC ranges from 0 to 1, where 1 indicates perfect agreement between the predicted and ground truth segmentations for that specific class. The overall

performance is typically reported as the mean of these per-image, per-class Dice scores averaged across all relevant classes and images in the dataset.

To evaluate the effectiveness of SAM-Med3D [30] in facilitating annotation, we calculated the Symmetric Difference (SD). This metric quantifies the total volume of discrepancy between two segmentations, representing the exact voxels an expert would need to adjust (either add or remove) to align a prediction with the ground truth. It is defined as the sum of false positives (FP) and false negatives (FN), as shown in Equation (2):

$$SD(A, B) = FP + FN \tag{2}$$

This metric is normalized (Normalized Symmetric Difference – NSD) per class by the union of predicted and the ground truth for the corresponding voxels. That ensures a fair comparison between classes with large regions (e.g., jawbones) and those with small regions (e.g., IAC). NSD ranges from 0 to 100, where 0 indicates perfect masks not requiring any modification.

To account for differences in organ size across classes (e.g., large regions such as Jawbones versus small regions such as the Sinus), SD was normalized by the union of predicted and ground-truth voxels, resulting in the Normalized Symmetric Difference (NSD). NSD ranges from 0 to 100, where 0 indicates perfectly overlapping masks that require no modification.

3.3 Active Learning sampling methods

Two AL sampling methods have been evaluated: Naive sampling (random selection) and Least confidence sampling. The random sampling consists into randomly select N images at each AL round. The least confidence [17] approach involves selecting the images for which the model is the least confident. The least confidence score for a single pixel is defined in Equation (3):

$$Uncertainty_{LeastConfidence}(\hat{y}) = |1 - \hat{y}| \tag{3}$$

where $\hat{y}$ is the predicted value for pixel y of an input image. The uncertainty score for an entire image is obtained by averaging the individual pixel uncertainty scores across all considered classes.

3.4 Workflow

During the AL process (see Fig. 1), round 0 corresponds to the cold-start and consists of the following: (1) N images are randomly selected for annotation, (2) a data fingerprint is generated and used to prepare the dataset for nnU-Net, and (3) the model is trained.

The following steps are performed in each subsequent AL round:

1. the informativeness of each unlabeled image is computed using previously trained model,
2. the most informative images are selected,

3. these images are annotated and incorporated into the set of images labeled in previous AL rounds,
4. the images are prepared for nnU-Net. Following the approach of [7], a fixed data fingerprint (generated in round 0) is reused across iterations to accelerate data preparation,
5. a new model is fine-tuned, and
6. the model is evaluated, with the best checkpoint always used to make predictions at each AL round.

This AL process is repeated until the annotation budget is exhausted.

Concerning the SAM predictions, the following steps are performed (see Fig. 2):

1. Predictions are generated using the nnU-Net model.
2. Prompts (i.e., simulated clicks on relevant areas corresponding to classes) are generated based on these predictions.
3. The images and prompts are fed into SAM-Med3D to produce 3D segmentations.

3.5 Network architecture

The segmentation is performed using the nnU-Net model [10]. It builds upon the successful U-Net architecture [26] and offers a self-configuring approach that minimizes the need for manual parameter tuning. nnU-Net has consistently demonstrated high performance across various medical datasets [10] and becomes the default model for medical image segmentation [13,27]. Concerning prompt-based models for segmentation, the SAM-Med3D model [30] has been used. This model has been specialized for 3D medical images and adapted to handle click-based prompts.

3.6 Hyper-parameters

Concerning nnU-Net [10], the default parameters were used, with three exceptions. To reduce computational demands and mitigate overfitting, since AL involves significantly fewer annotated examples than standard training, the number of iterations per epoch was limited to 100. Additionally, the number of epochs per AL round was limited to 50. Lastly, only the 3D low-resolution configuration of nnU-Net was used.

Concerning the AL part, prospective comparison of AL methods (i.e., actually asking an expert to annotate the selected images) is problematic, since image selection influences subsequent selections and, consequently, the results. To enable a fair comparison, the AL process was simulated using the fully annotated dataset. The following parameters was used:

- Number of AL rounds: 10
- Number of images selected at each AL round: 5

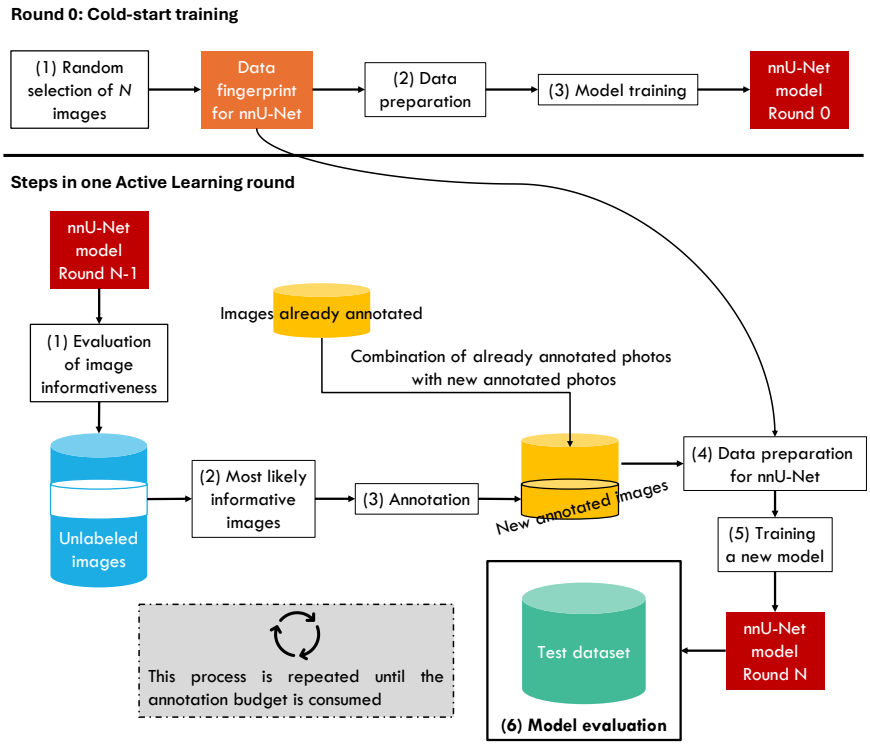

Fig. 1. Active Learning (AL) workflow. Round 0: A predefined number of images are randomly selected to generate the nnU-Net data fingerprint and train the initial model. Steps in a single AL round: (1) evaluate the informativeness of each image in the unlabeled pool using the current model, (2) select the most informative images, (3) annotate the selected images, (4) prepare the images for nnU-Net using the existing data fingerprint, (5) fine-tune the model with both previously and newly annotated images, and (6) evaluate the updated model. Steps 1–6 in are repeated until the annotation budget is exhausted.

- Cold start (round 0): 5% of annotated data (20 images) have been randomly selected images and used to initialize model training
- At each AL round, in accordance with the survey of Budd et al. [3], the model was finetuned using all available annotated data (previously + newly annotated images), from prior best checkpoint at the previous round.

For the evaluation, 15% of the dataset (72 images) was randomly sampled to form the test dataset. For a fair comparison, a nnU-Net model was also trained on the fully annotated dataset for the same number of iterations (50,000) as used in the 10 AL iterations (called "Internal Test" in Table 1).

Concerning the SAM-Med3D model [30], the default parameters were used.

Experiments were performed on a system with an NVIDIA A6000 GPU (48 GB VRAM), Intel Xeon Silver 4208 CPU (16 cores), and 128 GB RAM. The

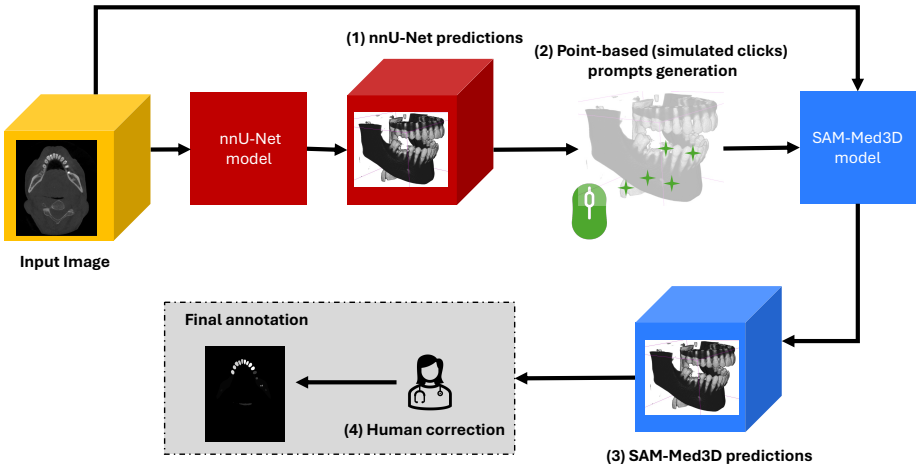

Fig. 2. Overview of 3D-assisted annotation using nnU-Net and a SAM-like model. The process consists of four steps: (1) generating pixel-wise predictions with nnU-Net, (2) creating point-based prompts (simulated clicks) from these predictions, (3) producing pixel-wise predictions with SAM-Med3D using these prompts, and (4) performing human corrections on the generated masks to obtain the final annotation.

code used for the experiments is publicly available at `https://github.com/martinicmrim/sam_nnunet`.

4 Results

4.1 Active Learning sampling methods on segmentation performance

The comparison between Active Learning (AL) sampling methods is depicted in Table 1. The performance of the AL sampling methods was evaluated using the model weights from the final AL round (i.e., round 10). We also report the performance obtained using the fully annotated dataset ("Internal Test"), as well as the performance of random sampling AL with the 3D full-resolution configuration of nnU-Net.

The AL methods demonstrated performance comparable to training on the full dataset, utilizing less than 20% of the original training data, with the exception of "Sinus" segmentation. Similar segmentation performance is observed between the random selection method and least confidence selection, although training time is 5 time longer.

Figure 3 depicts the qualitative evaluation of segmentation across AL rounds.

4.2 Evaluation of SAM-Med3D masks with prompts derived from nnU-Net predictions

Table 1. Mean Dice Score (in %) at the last AL round (round 10) and training time on grouped ToothFairy2 classes according to Active Learning sampling method

	Average	Jawbones	IAC	Sinus	Pharynx	Teeth	Training time	Data used
Method	DSC	DSC	DSC	DSC	DSC	DSC	Hours	%
Full dataset (FD) [2]	70.92	90.31	71.34	64.81	95.66	73.17	NA	100
Random Samp. AL	74.33	98.5	88.38	0	96.73	88.38	5	18
Least Conf. Samp. AL	73.7	98	86.02	0	96.82	87.67	27	18
Internal Test FD	74.33	98.15	88.38	0.0	96.74	88.37	5	100
Random Samp. AL - Full resolution	72.62	97.89	85.42	0.0	95.16	84.60	7	100

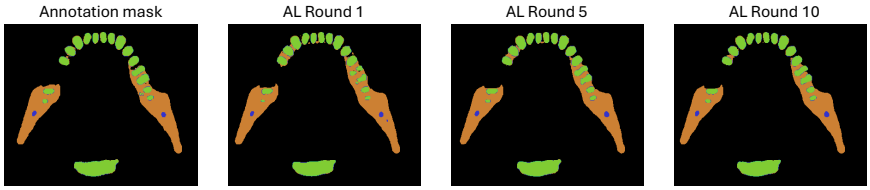

Fig. 3. Qualitative visualization of predictions for image 58 (ToothFairy dataset) at AL rounds 1, 5, and 10, compared with the annotation mask (axial slice S: 43.8 mm, 3D Slicer).

To evaluate the potential of SAM-Med3D [30] in facilitating the annotation of 3D dental images, we simulated an additional Active Learning (AL) iteration. The objective was to assess, if SAM-Med3D were deployed at the step 3 of the AL process, how much annotation effort could be reduced through the combination of nnU-Net and SAM-Med3D. Specifically, the quality of the masks generated by SAM-Med3D from prompts derived from nnU-Net predictions was evaluated. The procedure was as follows, based on the last AL iteration (with random sampling method):

1. Randomly select 5 images,
2. Generate 3D predictions using the most recently trained nnU-Net model,
3. Generate point-based prompts (i.e., simulated clicks) for each predicted class,
4. Use SAM-Med3D with the prompts and input images to produce 3D annotation masks,
5. Evaluate the quality of the generated 3D annotation masks.

The influence of the number of prompts per class (i.e., 1, 5, and 10 clicks per class) on the quality of the masks was also evaluated. The quality of the generated masks was quantitatively assessed using the Normalized Symmetric Difference (NSD), with Table 2 reporting the percentage of voxels requiring expert annotation or correction based on the combination of nnU-Net and SAM-Med3D.

Table 2. Evaluation of SAM-Med3D performance (Normalized Symmetric Difference, in %) with varying numbers of prompts per class.

Number of Prompts	Average	Jawbones	IAC	Sinus	Pharynx	Teeth
1 click	62.61	88.92	37.94	98.60	98.97	0
5 clicks	50.75	76.59	37.58	98.12	97.38	0
10 clicks	51.52	69.34	35.38	97.56	97.30	0

5 Discussion

Concerning AL, estimating informativeness at each AL round is computationally expensive. In this paper, we focus exclusively on the least confidence method to compare to random selection. While other strategies, such as entropy or Monte Carlo (MC) dropout, may improve the performance, they come with significantly higher computational costs. For example, MC dropout requires multiple forward passes per image, substantially increasing the overall runtime. The choice of cold-start images may also influence the outcomes, as noted in [20]. Moreover, consistent with findings in other medical domains (e.g., [6,23]), random selection has shown performance comparable to more complex selection strategies such as least confidence.

The preliminary results on the annotation using Med-SAM3D show that with 5 simulated point-based prompt (i.e., simulated clicks) from nnU-net prediction allows to reduce the number of pixels to annotate or verify to up to 50%. Other SAM models exploiting other type of prompt (e.g., [21]) could be explored to improve this pre-annotation.

Moreover, contrary to 2D image segmentation, where training U-Net-like models can be very fast and require fewer iterations, 3D image training demands significantly more computational time. Adding the use of SAM generate also a lot of time between each AL round. An asynchronous iteration need to be considered to limit the waiting for the experts during the annotation. Moreover, even if random is very hard to beat to selection the images to annotate, other sampling methods could be considered in the future. The trade-off between gain in term of quality in selection and the computing power required as well as computing time to select the images seems to be an essential criteria to develop new methods.

It is important to note that the selected test dataset may not be entirely representative of the underlying distribution of the full dataset. Furthermore, the chosen class grouping strategy appears to have significantly impacted segmentation performance. On one hand, this grouping led to increased performance for classes with a large pixel representation in the images (e.g., Teeth), as the aggregation of pixels likely facilitated model training. On the other hand, it severely degraded performance for the "Sinus" class, which became largely undetected. This degradation could be attributed to the increased class imbalance introduced by the grouping, which disproportionately affects minority or less complex classes such as "Sinus".

This paper presents a preliminary work on the combination of traditional segmentation models (nnU-net [10]) and prompt-based segmentation models (SAM-Med3D [30]) to facilitate data annotation and model training in the 3D dental domain. In future studies, other dental datasets (e.g., 3DTeethSeg [1]) will be considered. Moreover, nnU-Net is a complex model due to its automated configuration capabilities, which accelerate model setup. Other models, such as TransUNet (e.g., [4]), could also be considered in future work, especially to evaluate other AL sampling methods.

Acknowledgments. This work has been supported by MIAI@Grenoble Alpes (ANR-19-P3IA-0003). This work benefited from state aid managed by the National Research Agency under France 2030 bearing the reference ANR-23-IACL-0006.

Disclosure of Interests. Philippe Mulhem and Jean-Pierre Chevallet have no competing interests to declare that are relevant to the content of this article. Nicolas Martin owns stock in PEEKTORIA.

References

1. Ben-Hamadou, A., Smaoui, O., Rekik, A., Pujades, S., Boyer, E., Lim, H., Kim, M., Lee, M., Chung, M., Shin, Y.G., Leclercq, M., Cevidanes, L., Prieto, J.C., Zhuang, S., Wei, G., Cui, Z., Zhou, Y., Dascalu, T., Ibragimov, B., Yong, T.H., Ahn, H.G., Kim, W., Han, J.H., Choi, B., Nistelrooij, N.v., Kempers, S., Vinayahalingam, S., Strippoli, J., Thollot, A., Setbon, H., Trosset, C., Ladroit, E.: 3DTeethSeg'22: 3D Teeth Scan Segmentation and Labeling Challenge (May 2023). https://doi.org/10.48550/arXiv.2305.18277, arXiv:2305.18277 [cs]
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 (CVPR). pp. 5238–5248 (June 2025)
3. Budd, S., Robinson, E.C., Kainz, B.: A survey on active learning and human-in-the-loop deep learning for medical image analysis. Medical Image Analysis **71**, 102062 (Jul 2021). https://doi.org/10.1016/j.media.2021.102062
4. Chen, J., Mei, J., Li, X., Lu, Y., Yu, Q., Wei, Q., Luo, X., Xie, Y., Adeli, E., Wang, Y., Lungren, M.P., Zhang, S., Xing, L., Lu, L., Yuille, A., Zhou, Y.: TransUNet: Rethinking the U-Net architecture design for medical image segmentation through the lens of transformers. Medical Image Analysis **97**, 103280 (Oct 2024). https://doi.org/10.1016/j.media.2024.103280
5. Dao, L., Ly, N.Q.: A Comprehensive Study on Medical Image Segmentation using Deep Neural Networks. International Journal of Advanced Computer Science and Applications **14**(3) (2023). https://doi.org/10.14569/IJACSA.2023.0140319
6. Ekner, A.B., Lowes, M.M., Paulsen, R.R., Kofoed, K.F., Johansen, A.O., Sørensen, K.A., Sundgaard, J.V.: Active Learning with nnUNet for Coronary Artery Lumen Segmentation Using a Centerline Prior. In: Petersen, J., Dahl, V.A. (eds.) Image Analysis. pp. 227–239. Springer Nature Switzerland, Cham (2025). https://doi.org/10.1007/978-3-031-95918-9_16
7. Föllmer, B., Schulze, K., Wald, C., Stober, S., Samek, W., Dewey, M.: Active Learning with the nnUNet and Sample Selection with Uncertainty-Aware Submodular Mutual Information Measure. In: Proceedings of The 7nd International Conference on Medical Imaging with Deep Learning. pp. 480–503. PMLR (Dec 2024), https://proceedings.mlr.press/v250/follmer24a.html, iSSN: 2640-3498
8. Huang, J., Farpour, N., Yang, B.J., Mupparapu, M., Lure, F., Li, J., Yan, H., Setzer, F.C.: Uncertainty-based Active Learning by Bayesian U-Net for Multi-label Cone-beam CT Segmentation. Journal of Endodontics **50**(2), 220–228 (Feb 2024). https://doi.org/10.1016/j.joen.2023.11.002
9. Huang, Y., Yang, X., Liu, L., Zhou, H., Chang, A., Zhou, X., Chen, R., Yu, J., Chen, J., Chen, C., Liu, S., Chi, H., Hu, X., Yue, K., Li, L., Grau, V., Fan, D.P., Dong, F., Ni, D.: Segment anything model for medical images? Medical Image Analysis **92**, 103061 (2024). https://doi.org/https://doi.org/10.1016/j.media.2023.103061
10. Isensee, F., Jaeger, P.F., Kohl, S.A.A., Petersen, J., Maier-Hein, K.H.: nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nature Methods **18**(2), 203–211 (Feb 2021). https://doi.org/10.1038/s41592-020-01008-z, publisher: Nature Publishing Group
11. Isensee, F., Kirchhoff, Y., Kraemer, L., Rokuss, M., Ulrich, C., Maier-Hein, K.H.: Scaling nnU-Net for CBCT Segmentation (Dec 2024). https://doi.org/10.48550/arXiv.2411.17213, arXiv:2411.17213 [cs]

12. Isensee, F., Rokuss, M., Krämer, L., Dinkelacker, S., Ravindran, A., Stritzke, F., Hamm, B., Wald, T., Langenberg, M., Ulrich, C., Deissler, J., Floca, R., Maier-Hein, K.: nnInteractive: Redefining 3D Promptable Segmentation (Mar 2025). https://doi.org/10.48550/arXiv.2503.08373, arXiv:2503.08373 [cs]
13. Isensee, F., Wald, T., Ulrich, C., Baumgartner, M., Roy, S., Maier-Hein, K., Jäger, P.F.: nnU-Net Revisited: A Call for Rigorous Validation in 3D Medical Image Segmentation. In: Linguraru, M.G., Dou, Q., Feragen, A., Giannarou, S., Glocker, B., Lekadir, K., Schnabel, J.A. (eds.) Medical Image Computing and Computer Assisted Intervention – MICCAI 2024. pp. 488–498. Springer Nature Switzerland, Cham (2024). https://doi.org/10.1007/978-3-031-72114-4_47
14. Jung, S.K., Lim, H.K., Lee, S., Cho, Y., Song, I.S.: Deep Active Learning for Automatic Segmentation of Maxillary Sinus Lesions Using a Convolutional Neural Network. Diagnostics **11**(4), 688 (Apr 2021). https://doi.org/10.3390/diagnostics11040688
15. Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.Y., Dollár, P., Girshick, R.: Segment anything. In: 2023 IEEE/CVF International Conference on Computer Vision (ICCV). pp. 3992–4003 (2023). https://doi.org/10.1109/ICCV51070.2023.00371
16. Li, B., Yuan, Y., Tan, W.: Optimization of MedSAM model based on bounding box adaptive perturbation algorithm (Mar 2025). https://doi.org/10.48550/arXiv.2503.19700, arXiv:2503.19700 [cs]
17. Li, X., Xia, M., Jiao, J., Zhou, S., Chang, C., Wang, Y., Guo, Y.: HAL-IA: A Hybrid Active Learning framework using Interactive Annotation for medical image segmentation. Medical Image Analysis **88**, 102862 (Aug 2023). https://doi.org/10.1016/j.media.2023.102862
18. Li, Y., Jing, B., Li, Z., Wang, J., Zhang, Y.: Plug-and-play segment anything model improves nnUNet performance. Medical Physics **52**(2), 899–912 (Feb 2025). https://doi.org/10.1002/mp.17481
19. Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., van der Laak, J.A.W.M., van Ginneken, B., Sánchez, C.I.: A Survey on Deep Learning in Medical Image Analysis. Medical Image Analysis **42**, 60–88 (Dec 2017). https://doi.org/10.1016/j.media.2017.07.005, http://arxiv.org/abs/1702.05747, arXiv: 1702.05747
20. Liu, H., Li, H., Yao, X., Fan, Y., Hu, D., Dawant, B., Nath, V., Xu, Z., Oguz, I.: COLosSAL: A Benchmark for Cold-start Active Learning for 3D Medical Image Segmentation (Jul 2023). https://doi.org/10.48550/arXiv.2307.12004, http://arxiv.org/abs/2307.12004, arXiv:2307.12004 [cs]
21. Ma, J., He, Y., Li, F., Han, L., You, C., Wang, B.: Segment anything in medical images. Nature Communications **15**(1), 1–9 (December 2024). https://doi.org/10.1038/s41467-024-44824-z
22. Ma, J., Yang, Z., Kim, S., Chen, B., Baharoon, M., Fallahpour, A., Asakereh, R., Lyu, H., Wang, B.: MedSAM2: Segment Anything in 3D Medical Images and Videos (Apr 2025). https://doi.org/10.48550/arXiv.2504.03600, http://arxiv.org/abs/2504.03600, arXiv:2504.03600 [eess]
23. Martin, N., Chevallet, J.P., Mulhem, P., Quénot, G.: Combining Image and Region Uncertainty-Based Active Learning for Melanoma Segmentation. In: 2024 International Conference on Content-Based Multimedia Indexing (CBMI). pp. 1–7. IEEE, Reykjavik, Iceland (Sep 2024). https://doi.org/10.1109/CBMI62980.2024.10859208

24. Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C.L., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., Ray, A., Schulman, J., Hilton, J., Kelton, F., Miller, L., Simens, M., Askell, A., Welinder, P., Christiano, P., Leike, J., Lowe, R.: Training language models to follow instructions with human feedback (Mar 2022). https://doi.org/10.48550/arXiv.2203.02155, arXiv:2203.02155 [cs]
25. Ren, P., Xiao, Y., Chang, X., Huang, P.Y., Li, Z., Gupta, B.B., Chen, X., Wang, X.: A Survey of Deep Active Learning. ACM Comput. Surv. **54**(9), 180:1–180:40 (Oct 2021). https://doi.org/10.1145/3472291
26. Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional Networks for Biomedical Image Segmentation. arXiv:1505.04597 [cs] (May 2015), http://arxiv.org/abs/1505.04597, arXiv: 1505.04597
27. Russell, E., Boyd, A., Finlay, D., Trindade, L.: Machine learning-based anatomical segmentation: A systematic review of methodologies and applications. Open J Clin Med Images **4**(1), 1187 (2024)
28. Settles, B.: Active Learning Literature Survey. Computer Sciences Technical Report 1648, University of Wisconsin–Madison (2009), http://axon.cs.byu.edu/~martinez/classes/778/Papers/settles.activelearning.pdf
29. Stock, R., Kirchhoff, Y., Rokuss, M.R., Ravindran, A., Maier-Hein, K.: Segment Anything in Medical Images with nnUNet. In: Ma, J., Zhou, Y., Wang, B. (eds.) Medical Image Segmentation Foundation Models. CVPR 2024 Challenge: Segment Anything in Medical Images on Laptop. pp. 167–179. Springer Nature Switzerland, Cham (2025). https://doi.org/10.1007/978-3-031-81854-7_11
30. Wang, H., Guo, S., Ye, J., Deng, Z., Cheng, J., Li, T., Chen, J., Su, Y., Huang, Z., Shen, Y., Fu, B., Zhang, S., He, J., Qiao, Y.: SAM-Med3D: Towards General-purpose Segmentation Models for Volumetric Medical Images (Sep 2024). https://doi.org/10.48550/arXiv.2310.15161, arXiv:2310.15161 [cs]
31. Wang, Y., Zhang, Y., Chen, X., Wang, S., Qian, D., Ye, F., Xu, F., Zhang, H., Zhang, Q., Wu, C., Li, Y., Cui, W., Luo, S., Wang, C., Li, T., Liu, Y., Feng, X., Zhou, H., Liu, D., Wang, Q., Lin, Z., Song, W., Li, Y., Wang, B., Wang, C., Chen, Q., Li, M.: STS MICCAI 2023 Challenge: Grand challenge on 2D and 3D semi-supervised tooth segmentation (Jul 2024). https://doi.org/10.48550/arXiv.2407.13246, arXiv:2407.13246 [cs]
32. Yoo, D., Kweon, I.S.: Learning Loss for Active Learning. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 93–102. IEEE, Long Beach, CA, USA (Jun 2019). https://doi.org/10.1109/CVPR.2019.00018
33. Zeng, X., Wen, L., Xu, Y., Ji, C.: Generating diagnostic report for medical image by high-middle-level visual information incorporation on double deep learning models. Computer methods and programs in biomedicine **197**, 105700 (2020)