ACA: Post-hoc Adaptive Logit Alignment for 3D Medical Image Segmentation

Nanyu Dong[1], Qi Chen[1], Johan Verjans[1], and Zhibin Liao[2,1⋆]

[1] Australian Institute for Machine Learning, Adelaide University, Australia
[2] School of Computer Science and Information Technology, Adelaide University, Australia {nanyu.dong, qi.chen04, johan.verjans, zhibin.liao}@adelaide.edu.au

Abstract. Despite the rapid advancement of 3D medical image segmentation foundation models trained on large-scale datasets, generalization to unseen data remains challenging. We observe that performance degradation is closely associated with localized voxel-level confidence distribution shifts between training and testing data. Existing Unsupervised Domain Adaptation (UDA) methods attempt to address sample-level domain discrepancies; however, they typically rely on computationally expensive gradient-based updates, which may lead to catastrophic forgetting and degradation of pretrained representations. To address these limitations, we propose a simple yet efficient post-hoc method termed **Adaptive Cliff Alignment (ACA)** to reduce voxel-level confidence distribution gaps in logit space. Our approach analyzes the shape of the predicted logit distributions and performs isotonic-regression-based quantile matching to align training and testing distributions. Importantly, the proposed alignment is confined to the model-uncertain region of the logit space rather than applied globally, enabling targeted correction while preserving confident predictions. Extensive experiments across multiple 3D medical image segmentation benchmarks demonstrate that ACA effectively mitigates train–test prediction distribution shifts and consistently improves segmentation performance in a train-free manner. Code is available at `https://github.com/MelindaDong/aca-segmentation`.

Keywords: Post-Hoc Distribution Alignment · 3D Medical Image Segmentation.

1 Introduction

The rapid development of medical image segmentation foundation models, trained on large-scale diverse datasets, has established a new paradigm in medical image analysis [21,11]. However, their performance can still decline substantially when applied to unseen test data, indicating persistent generalization gaps. And this gap is pervasive across semantic tasks (*e.g.*, segmenting liver *vs*. heart), and the magnitude of performance degradation varies substantially.

⋆ Corresponding author

To address the generalization gaps, unsupervised Domain Adaptation (UDA) aims to mitigate performance degradation caused by distributional shifts between labeled source domains and unlabeled target domains. Traditional UDA strategies typically achieve domain alignment through image-level translation using generative adversarial networks [20]. At the feature-level, alignment is commonly achieved either through adversarial discriminators [3], or by minimizing statistical discrepancies, such as Correlation Alignment (CORAL) and Maximum Mean Discrepancy (MMD) [4,17]. However, these approaches generally need gradient-based optimization, architectural modifications, or the integration of auxiliary trainable modules.

In the era of large-scale pretrained foundation models, retraining-based adaptation becomes increasingly impractical. It is not only computationally prohibitive but also risks *catastrophic forgetting*, which degrades the model's original capabilities. These challenges are further amplified in 3D medical imaging segmentation; unlike standard classification tasks, the high dimensionality of 3D dense prediction makes traditional training-based adaptation excessively heavy or entirely infeasible in resource-constrained deployment scenarios.

Motivated by these observations, we propose a novel post-hoc adaptation strategy that operates directly in the model's output logit space by aligning logit distributions within the uncertain region, which we term the "cliff". Our approach enables effective logit distribution alignment, improving test-time performance of large segmentation foundation models while preserving their representational capacity and computational efficiency.

Our main contributions are summarized as follows:

- We propose Adaptive Cliff Alignment (ACA), a novel post-hoc, train-free, and label-free method that aligns voxel-level logit distributions within the uncertain prediction range (*i.e.*, the cliff), improving the performance of foundation segmentation models without requiring lengthy model fine-tuning.
- We demonstrate that, in 3D medical image segmentation, mitigating train-test voxel-level shift is more effective than correcting sample-wise shifts. Moreover, aligning extremely confident logits offers limited generalization benefits, as anatomical variability introduces distributional inconsistencies.
- Extensive experiments on multiple 3D medical image segmentation benchmarks demonstrate that ACA consistently improves performance across diverse medical tasks.

2 Related Work

Unsupervised Domain Adaptation. Unsupervised Domain Adaptation (UDA) [10] aims to mitigate performance degradation caused by domain shifts. UDA aligns distributions at the image level via GANs [20] or at the feature level using adversarial discriminators [3] and statistical divergence metrics like MMD [4] or CORAL [17]. However, it generally requires computationally expensive backward passes, which risks corrupting the carefully learned representations of foundation models. Moment matching is another classic training-free approach.

For instance, AdaBN [8] aligns domains by matching mean and variance statistics across batches. Yet, the massive memory footprint of 3D volumes in medical image segmentation inherently restricts batch sizes, rendering batch-dependent methods inapplicable.

Post-hoc Calibration and Logit Alignment. Post-hoc calibration and logit adjustment techniques both seek to improve model reliability by mapping raw predictions to real-world empirical probabilities or priors. Classic calibration approaches include temperature scaling [15] and optimal threshold tuning [9,2], which utilize a held-out validation set to find an optimal global rescaling. Non-parametric approaches include histogram binning, quantile matching [13] for cumulative distribution function alignment, and isotonic regression [18,14] for learning monotonic mappings to minimize squared error under monotonicity constraints. Similarly, Long-tail learning [12] aims to align the class prior to the real-world probability to train robust models [19]. However, these methods inherently require a labeled target-domain validation set [5,16], limiting their practicality in clinical settings where such annotations are scarce or unavailable.

3 Methodology

3.1 Problem Formulation

To improve the generalizability of pre-trained foundation models and operate without an available extra validation set, we propose a post-hoc method that calibrates the test predictions by exploring the relationship between the training and test sets' segmentation logit distributions.

Let $\mathcal{X}_{\text{train}}$ and $\mathcal{X}_{\text{test}}$ denote the training and test sets, respectively, and a segmentation foundation model F can produce a logit prediction map $\mathbf{z}$ for an input image $\mathbf{x}$, *i.e.*, $\mathbf{z} = F(\mathbf{x})$. We first compute the set of training sample logit maps $\mathcal{Z}_{\text{train}} = \{\mathbf{z}|\mathbf{z} = F(\mathbf{x}), \forall \mathbf{x} \in \mathcal{X}_{\text{train}}\}$ and test set logits $\mathcal{Z}_{\text{test}}$ in the same manner. Here, we decompose the multi-class segmentation task into multiple binary segmentation tasks, following the paradigm of class-prompted or text-driven segmentation foundation models [21,6]. Our goal is to match the test set logit distribution to the training set counterpart within an adaptively selected logit range for the prompted class.

3.2 Adaptive Cliff Alignment

Range filtering. In 3D medical image segmentation, boundary region information contributes the most to the segmentation performance, yet it accounts for only a small fraction of the total volume. Furthermore, the dimensionality of $\mathbf{z}_{\text{train/test}}$ (see Eq. (1)) is exceptionally high, as it scales with both the number of samples and the total voxel count.

To mitigate memory constraints and reduce computational complexity, we implement a coarse range-filtering step with predefined bounds of $[-5, 10]$. These hyperparameters are chosen to exclude highly confident background voxels (foreground confidence < 0.006) and highly confident foreground voxels (foreground

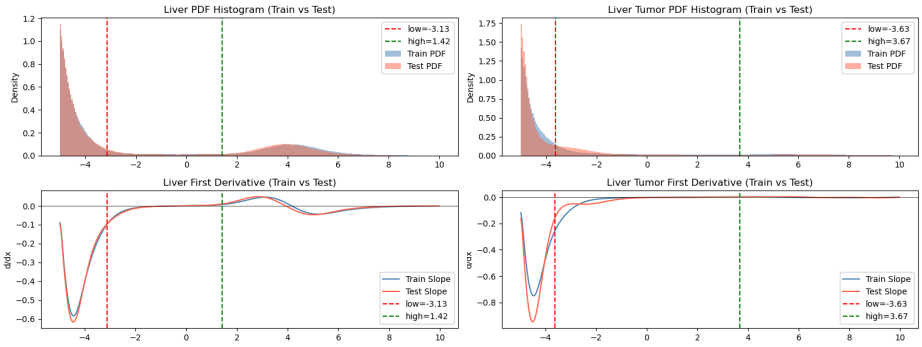

Fig. 1. Two types of logit distributions illustrated using liver and liver tumor in Liver task from the MSD dataset [1]. Left: *two peaks*, a highly confident background, and a less confident foreground distribution. Right: *one peak*, strong background dominance followed by a long tail foreground distribution. Top: probability density distribution; Bottom: slope of the probability density.

confidence > 0.9999), while preserving an appropriately sized logit range for identifying and aligning uncertain regions that typically correspond to object boundaries.

Specifically, we concatenate the flattened logits and apply the filtering at the same time ($\mathbf{z}_{\text{test}}$ constructed analogously):

$$\mathbf{z}_{\text{train}} = \text{cat}\left(\{z \in \text{flatten}(\mathbf{z}) \mid -5 \leq z < 10, \mathbf{z} \in \mathcal{Z}_{\text{train}}\}\right), \tag{1}$$

Logit distribution consistency. The ratio $\rho_{\text{train}} = \frac{\sum \mathbb{1}(\mathbf{z}_{\text{train}} > 0)}{|\mathbf{z}_{\text{train}}|}$ provides an empirical estimate of the average proportion of foreground voxels within the volume. This ratio reflects a stable relationship between foreground and background regions under the default segmentation decision threshold corresponding to a foreground probability of 0.5. Here $\mathbb{1}(\cdot)$ denotes the indicator function. Ideally, under the same probability threshold, the corresponding ratio on the test set, ρ_{test}, should match ρ_{train} if the independent and identically distributed (i.i.d.) assumption holds. More generally, this statistical relationship can be extended to any logit interval with sufficient support in its probability density:

$$\frac{\sum \mathbb{1}(l \leq \mathbf{z}_{\text{train}} < u)}{|\mathbf{z}_{\text{train}}|} \approx \frac{\sum \mathbb{1}(l \leq \mathbf{z}_{\text{test}} < u)}{|\mathbf{z}_{\text{test}}|}, \tag{2}$$

where l and u denote the lower and upper bounds of the logit range.

Given that Eq. (2) performs a binarized count over logit values within a specified range, we can instead enforce a more fine-grained density alignment between $\mathbf{z}_{\text{train}}$ and $\mathbf{z}_{\text{test}}$ within the same bounds:

$$P(l \leq \mathbf{z}_{\text{train}} < u) \approx P(l \leq \mathbf{z}_{\text{test}} < u), \tag{3}$$

where $P(l \leq z < u) = \int_l^u f(z)\, dz$, and $f(\cdot)$ denotes the probability density function.

Adaptive "cliff" alignment. In segmentation tasks, overfitting often manifests as reduced predictive confidence near uncertain object boundaries in test samples. This behavior can be attributed to the memorization of training-specific patterns and the potential learning of spurious, speckle-like noise. While calibrating the prediction threshold is a common approach, it typically requires the availability of a dedicated validation set.

In the absence of such a validation set, we identify the uncertainty region by searching outward from the default decision threshold $z = 0$ to detect the onset of a density rise. Specifically, we first compute a smoothed histogram of logit values using a Gaussian filter ($\sigma = 8$) over 300 bins to suppress high-frequency noise. To automatically determine the boundaries of the uncertainty region, we compute the slope (first derivative) of the smoothed probability density and perform a bidirectional search from $z = 0$. The uncertainty region is defined by the two onset points at which the density begins to increase markedly, where the gradient magnitude exceeds 15% of the maximum gradient on that side, indicating a transition from uncertain to confident predictions. These two onset points define the lower and upper bounds, l and u, in Eq. (3) for both $\mathbf{z}_{\text{train}}$ and $\mathbf{z}_{\text{test}}$. By enforcing Eq. (3) within the derived bounds, we align the level of prediction uncertainty around segmentation boundaries in the test set with the corresponding statistics observed in the training set. Fig. 1 illustrates two commonly observed types of the $\mathbf{z}_{\text{train}}$ and $\mathbf{z}_{\text{test}}$ value distributions, visualized using histogram-based probability density estimates (top row) and their corresponding slopes at each location (bottom row), together with the adaptively selected logit ranges. Observing that segmentation prediction logits often exhibit a sharp cliff-like structure when they overfit on the training set, we refer to the proposed method as *Adaptive Cliff Alignment* (ACA).

Isotonic Quantile Mapping. Within the identified uncertain logit range $[l, u]$, we align $\mathbf{z}_{\text{test}}$ to $\mathbf{z}_{\text{train}}$ distribution using an isotonic regression-based quantile mapping method.

Logits inherently encode relative confidence through their ordering; therefore, a monotonic mapping is required to preserve the semantic consistency implied by the predicted logit rankings. Isotonic regression yields a piecewise linear mapping with greater expressive capacity than standard linear regression, while strictly enforcing monotonicity.

We compute empirical quantiles at $N = 1000$ equally spaced levels $p_i = \frac{i}{N}$ for $i = \{1, \ldots, N\}$ from $\mathbf{z}_{\text{train}}$ and $\mathbf{z}_{\text{test}}$. This generates two quantile sequences: $\mathcal{Q}_{\text{train}} = \{q_i^{\text{train}}\}_{i=1}^{N}$ and $\mathcal{Q}_{\text{test}} = \{q_i^{\text{test}}\}_{i=1}^{N}$, where $q_i^{\text{train}} = \int_{l_i}^{u_i} f(\mathbf{z}_{\text{train}})$ represents the logit whose empirical CDF equals p_i; l_i and u_i denote the lower and upper bounds of each quantile. We then fit an isotonic regression model $I : \mathcal{Q}^{\text{test}} \to \mathcal{Q}^{\text{train}}$ to learn a monotonically non-decreasing mapping from the test quantiles to the training quantiles.

During inference, for any test logit $z \in [l, u]$, the calibrated logit is obtained via the learned mapping $z' = I(z)$ and used as the adjusted prediction logit. The final foreground prediction is then produced by applying the default 0.5

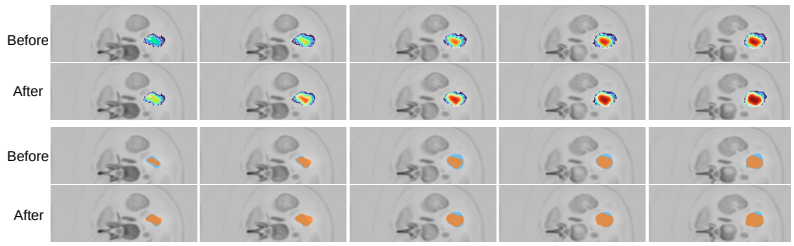

Fig. 2. An example of the proposed *Adaptive Cliff Alignment* (ACA) on a *pancreas tumor* segmentation task from the MSD benchmark [1]. The top two rows show the predicted logits visualized as heatmaps within the identified uncertainty range, before and after applying ACA. The bottom two rows display the corresponding binarized segmentation masks, where blue denotes the ground-truth annotation and orange indicates the predicted mask.

probability threshold. We illustrate two examples of logit shifts before and after alignment in Fig. 2.

4 Experiments

4.1 Dataset and Comparison Methods

Datasets. To evaluate our method's adaptability, we utilize the Medical Segmentation Decathlon (MSD) benchmark [1], a comprehensive collection of datasets designed to test algorithmic generalizability across diverse 3D medical imaging tasks. The evaluated datasets and their specific segmentation tasks include: Heart (left heart atrium), Liver (liver, liver tumor), Hippocampus (anterior and posterior hippocampus), Prostate (transition and peripheral zones), Lung (lung tumor), Pancreas (pancreas, pancreas tumor), Hepatic Vessel (liver vessel, liver tumor), Spleen (spleen), and Colon (colon cancer). For datasets with multiple tasks, the mean result is reported.

Comparison Methods. We adopt SAT [21] as our primary baseline as it uniquely supports non-interactive, text-driven 3D segmentation. Unlike SAM-based models [11], which require interactive prompting, or VISTA3D [6] and STU-Net [7], which are constrained by fixed class mappings and 'segment-all' architectures, SAT offers a highly flexible and scalable framework for large-vocabulary segmentation.

A sensitivity analysis of our hyperparameters has been conducted, which shows stable performance under reasonable variations; details are provided in the code repository.

To rigorously evaluate our approach, we compare it against several established distribution alignment and calibration techniques, configured under a unified post-hoc evaluation setting.

(1) **Logit Adjustment** [12] shifts test-time logits using foreground class priors estimated from the training split, where a scalar offset controlling the

Table 1. Dice Similarity Coefficient (DSC) (%) comparison. Best performance in each row is boldfaced. BL: Baseline; LA: Long tail Logit Adjustment; OT: Optimal Thresholding; MM: Moment Matching; FQM: Full-range Quantile Mapping; ACA: Proposed Adaptive Cliff Alignment.

Dataset [1]	**BL** [21]	**LA** [12]	**OT** [9]	**MM**	**FQM**	***ACA***
Cardiac	92.60	**92.64**	92.58	**92.64**	92.32	92.44
Liver	76.64	76.27	76.40	76.90	75.25	**77.87**
Hippocampus	86.81	87.12	87.12	**87.16**	87.14	**87.16**
Prostate	73.64	73.44	73.64	73.70	70.71	**73.90**
Lung	63.01	62.13	63.04	63.79	64.47	**64.62**
Pancreas	59.75	59.01	59.74	59.87	61.69	**62.26**
Hepatic Vessels	**63.19**	62.47	63.18	62.85	61.80	63.06
Spleen	94.26	94.26	94.18	94.18	82.31	**94.27**
Colon	35.51	33.99	34.61	36.54	**41.40**	40.68
Average	71.71	71.26	71.61	71.96	70.79	**72.92**

Table 2. Normalized Surface Distance (NSD) (%) comparison. Best performance in each row is boldfaced. BL: Baseline; LA: Long tail Logit Adjustment; OT: Optimal Thresholding; MM: Moment Matching; FQM: Full-range Quantile Mapping; ACA: Proposed Adaptive Cliff Alignment.

Dataset [1]	**BL** [21]	**LA** [12]	**OT** [9]	**MM**	**FQM**	***ACA***
Cardiac	83.19	**83.23**	83.08	83.20	81.83	82.59
Liver	63.11	62.87	62.80	63.19	56.37	**63.72**
Hippocampus	95.96	96.15	96.14	96.18	96.20	**96.21**
Prostate	**55.52**	55.35	**55.52**	55.39	49.27	55.50
Lung	53.00	51.57	53.08	54.44	55.43	**55.56**
Pancreas	47.13	46.39	47.13	47.24	48.32	**49.10**
Hepatic Vessels	**56.08**	55.18	56.07	55.65	54.50	56.03
Spleen	85.85	**85.98**	85.19	85.20	32.70	85.97
Colon	29.99	28.39	29.10	30.95	**34.49**	34.16
Average	63.31	62.79	63.12	63.49	56.57	**64.31**

magnitude of the shift is selected via grid search. Since a dedicated validation set is not available, we randomly split 20% of the training set as a tuning set.

(2) **Optimal Thresholding** [9] bypasses the standard zero-logit boundary by empirically determining the best decision threshold on the same tuning set.

(3) **Moment Mapping** To ensure compatibility with the training-free setting, we implement a simple moment matching strategy inspired by AdaBN [8], aligning the mean and standard deviation of source and target logits.

(4) **Full-range Quantile Mapping** [18] is adapted as an ablation baseline, directly isotonic regression between distributions to evaluate the impact of quantile alignment without our proposed adaptive cliff selection.

4.2 Quantitative Results Analysis

We evaluated the proposed method against several train-free, post-hoc refinement techniques using the Dice Similarity Coefficient (DSC) for regional overlap and Normalized Surface Distance (NSD) for boundary alignment evaluation.

As presented in Table 1 and Table 2, our method achieves the highest average performance across all 9 datasets, yielding a 72.92% average DSC and 64.31% average NSD, outperforming the baseline by 1.21% and 1.00%, respectively. In particular, our method shows consistent improvements on tasks involving tumor targets, such as Pancreas (+2.51% DSC) and Colon (+5.17% DSC) compared to the baseline.

Notably, supervised calibration methods such as Logit Adjustment (LA) and Optimal Thresholding (OT) do not consistently outperform the baseline. LA aligns foreground class probabilities between training and validation labels to mitigate class imbalance, but this correction does not necessarily improve segmentation accuracy. OT assumes that the validation set closely resembles the test distribution, an assumption that may not hold in real-world scenarios. Moreover, both approaches operate at a global semantic scale and fail to resolve voxel-level distribution discrepancies that are critical for precise segmentation.

Moment Matching (MM) yields modest improvements over the baseline, supporting the presence of a distributional gap between training and test predictions despite the commonly assumed i.i.d. sampling assumption. By aligning only the mean and standard deviation, MM performs a coarse global adjustment of the logit distribution. Although relatively robust to background dominance and extreme outliers, it lacks the capacity to correct fine-grained distribution discrepancies.

As a sanity check, full-range quantile mapping (FQM) without adaptive cliff selection degrades performance. While FQM aligns the entire logit spectrum, it is dominated by background volume and foreground variability, biasing the mapping toward extreme confident logits. For example, in the Spleen task, large target variability causes fluctuations in foreground logit ratios, leading to a marked performance drop.

Finally, our method aligns the test logit distribution within the prediction cliff to the corresponding training distribution, yielding consistent gains in both DSC and NSD. This improvement arises because foundation models retain relative voxel ordering near segmentation boundaries despite potential overfitting. By calibrating logits in the uncertain region, ACA enhances confidence along the prediction cliff, achieving precise sub-distributional alignment and improved generalization.

5 Conclusion

In conclusion, we present Adaptive Cliff Alignment (ACA), a post-hoc, train-free method that targets voxel-level confidence distribution shifts in 3D medical image segmentation. By applying isotonic regression–based quantile matching within the model-uncertain logit region, ACA effectively reduces train–test

distribution gaps while preserving confident predictions. Extensive experiments show that ACA consistently improves segmentation performance across diverse benchmarks, offering a simple and efficient solution for enhancing the generalization of pretrained segmentation models.

Future work will explore richer utilization of the logit space for detecting hard or out-of-distribution cases and providing deeper insights into model behavior and data characteristics, potentially extending ACA beyond alignment toward more comprehensive model understanding.

Acknowledgments. This work was supported by the Centre for Augmented Reasoning, an initiative by the Department of Education, Australian Government.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Antonelli, M., Reinke, A., Bakas, S., Farahani, K., Kopp-Schneider, A., Landman, B.A., Litjens, G., Menze, B., Ronneberger, O., Summers, R.M., et al.: The medical segmentation decathlon. Nature communications **13**(1), 4128 (2022)
2. Ding, Z., Han, X., Liu, P., Niethammer, M.: Local temperature scaling for probability calibration (2021), https://arxiv.org/abs/2008.05105
3. Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., March, M., Lempitsky, V.: Domain-adversarial training of neural networks. Journal of machine learning research **17**(59), 1–35 (2016)
4. Gretton, A., Borgwardt, K.M., Rasch, M.J., Schölkopf, B., Smola, A.: A kernel two-sample test. The journal of machine learning research **13**(1), 723–773 (2012)
5. Guo, C., Pleiss, G., Sun, Y., Weinberger, K.Q.: On calibration of modern neural networks (2017), https://arxiv.org/abs/1706.04599
6. He, Y., Guo, P., Tang, Y., Myronenko, A., Nath, V., Xu, Z., Yang, D., Zhao, C., Simon, B., Belue, M., Harmon, S., Turkbey, B., Xu, D., Li, W.: Vista3d: A unified segmentation foundation model for 3d medical imaging (2024), https://arxiv.org/abs/2406.05285
7. Huang, Z., Wang, H., Deng, Z., Ye, J., Su, Y., Sun, H., He, J., Gu, Y., Gu, L., Zhang, S., Qiao, Y.: Stu-net: Scalable and transferable medical image segmentation models empowered by large-scale supervised pre-training (2023), https://arxiv.org/abs/2304.06716
8. Li, Y., Wang, N., Shi, J., Liu, J., Hou, X.: Revisiting batch normalization for practical domain adaptation. Pattern Recognition **74**, 236–245 (2018)
9. Lipton, Z.C., Elkan, C., Naryanaswamy, B.: Optimal thresholding of classifiers to maximize f1 measure. In: Joint European Conference on Machine Learning and Knowledge Discovery in Databases. pp. 225–239. Springer (2014)
10. Liu, X., Yoo, C., Xing, F., Oh, H., Fakhri, G.E., Kang, J.W., Woo, J.: Deep unsupervised domain adaptation: A review of recent advances and perspectives. arXiv preprint arXiv:2208.07422 (2022)
11. Ma, J., He, Y., Li, F., Han, L., You, C., Wang, B.: Segment anything in medical images. Nature Communications **15**(1) (Jan 2024). https://doi.org/10.1038/s41467-024-44824-z, http://dx.doi.org/10.1038/s41467-024-44824-z

12. Menon, A.K., Jayasumana, S., Rawat, A.S., Jain, H., Veit, A., Kumar, S.: Long-tail learning via logit adjustment (2021), https://arxiv.org/abs/2007.07314
13. Naeini, M., Cooper, G., Hauskrecht, M.: Obtaining well calibrated probabilities using bayesian binning. aaai, 2901 (2015)
14. Niculescu-Mizil, A., Caruana, R.: Predicting good probabilities with supervised learning. In: Proceedings of the 22nd international conference on Machine learning. pp. 625–632 (2005)
15. Platt, J., et al.: Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. Advances in large margin classifiers **10**(3), 61–74 (1999)
16. Rousseau, A.J., Becker, T., Appeltans, S., Blaschko, M., Valkenborg, D.: Post hoc calibration of medical segmentation models. Discover Applied Sciences **7**(3), 180 (2025)
17. Sun, B., Feng, J., Saenko, K.: Correlation alignment for unsupervised domain adaptation (2016), https://arxiv.org/abs/1612.01939
18. Zadrozny, B., Elkan, C.: Obtaining calibrated probability estimates from decision trees and naive bayesian classifiers. In: Icml. vol. 1, p. 2001 (2001)
19. Zhang, C., Almpanidis, G., Fan, G., Deng, B., Zhang, Y., Liu, J., Kamel, A., Soda, P., Gama, J.: A systematic review on long-tailed learning. IEEE Transactions on Neural Networks and Learning Systems (2025)
20. Zhang, X., Wu, Y., Angelini, E., Li, A., Guo, J., Rasmussen, J.M., O'Connor, T.G., Wadhwa, P.D., Jackowski, A.P., Li, H., et al.: Mapseg: Unified unsupervised domain adaptation for heterogeneous medical image segmentation based on 3d masked autoencoding and pseudo-labeling. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5851–5862 (2024)
21. Zhao, Z., Zhang, Y., Wu, C., Zhang, X., Zhou, X., Zhang, Y., Wang, Y., Xie, W.: Large-vocabulary segmentation for medical images with text prompts (2025), https://arxiv.org/abs/2312.17183