UCBound-Net: Uncertainty-Guided Boundary-Aware Continual Learning for Domain-Incremental Ultrasound Segmentation

Anonymized Authors

Anonymized Affiliations
email@anonymized.com

Abstract. Continual learning in clinical imaging faces a dual challenge: a model must assimilate knowledge from new anatomical domains while faithfully retaining representations learned from prior tasks—a problem known as catastrophic forgetting. Existing mitigation strategies, including regularization and knowledge distillation, treat all spatial regions equally, ignoring the fact that prediction uncertainty is strongly correlated with the propensity for forgetting. We introduce **UCBound-Net**, a continual segmentation framework that exploits Monte Carlo (MC) Dropout uncertainty as a spatial proxy for forgetting risk. Our method contributes three synergistic components: (i) *uncertainty-weighted boundary distillation*, which amplifies the knowledge-transfer signal at high-entropy regions of the frozen teacher; (ii) *uncertainty-calibration regularization*, which explicitly penalizes overconfident erroneous predictions; and (iii) *uncertainty-guided exemplar selection*, a memory buffer that preferentially stores samples whose boundary regions exhibit the highest predictive entropy. Evaluated on a sequential domain-incremental benchmark comprising breast ultrasound (BUSI, Task 1) followed by thyroid ultrasound (TN3K, Task 2), UCBound-Net reduces backward transfer by **43.3%** relative to naive fine-tuning ($\Delta_{\mathrm{BWT}}=-0.098$ vs. -0.173) while achieving an average Dice Similarity Coefficient (DSC) of **0.755** across both tasks—surpassing both baselines without requiring task-boundary supervision. Our ablation study confirms that each component contributes independently to forgetting mitigation, yielding a clear design pathway for uncertainty-aware clinical continual learning.

Keywords: Continual learning · Catastrophic forgetting · Uncertainty quantification · Medical image segmentation · Ultrasound imaging · Knowledge distillation

1 Introduction

Medical imaging AI systems are typically trained under a closed-world assumption: a fixed dataset, a fixed domain, and a fixed deployment target. In clinical practice, however, imaging conditions evolve continuously—new scanner hardware is installed, acquisition protocols are updated, and diagnostic focus shifts across

anatomical regions [18]. When a deployed model is fine-tuned on a new data stream without protective mechanisms, it suffers catastrophic forgetting [1]: the rapid, irreversible erosion of performance on previously learned tasks. This is especially dangerous in safety-critical contexts such as ultrasound segmentation, where a model may be sequentially adapted across breast, thyroid, or abdominal imaging without access to stored patient data due to privacy constraints.

Existing continual learning (CL) approaches fall into three families. *Regularization methods* [2,3] penalize updates to weights identified as important for prior tasks. *Distillation methods* [4,5] preserve knowledge by enforcing consistency between new and old model outputs. *Replay methods* [6,7] store a subset of previous data or learned features for interleaved rehearsal. Despite their individual merits, none of these approaches accounts for *spatial heterogeneity* in forgetting risk: within a single medical image, high-uncertainty boundary pixels are far more vulnerable to forgetting than low-uncertainty interior regions, yet all pixels are treated identically during distillation or replay.

A critical observation motivates our work: *regions where the model is uncertain about its prediction are precisely the regions most prone to catastrophic forgetting.* Boundary regions in ultrasound images, which exhibit speckle noise and low contrast, yield high predictive entropy under MC Dropout. These same regions are most easily overwritten by gradient updates during new-task learning. We hypothesize that selectively amplifying the distillation and replay signals at high-uncertainty spatial locations will yield substantially better forgetting resistance.

We make the following contributions:

1. **Uncertainty-weighted boundary distillation:** a novel Knowledge Distillation (KD) loss in which the pixel-wise transfer weight is derived from the teacher model's MC Dropout entropy map, concentrating the distillation gradient on high-risk boundary regions.
2. **Uncertainty-calibration regularizer:** an auxiliary loss that encourages the student's uncertainty map to be consistent with its own prediction errors, improving calibration across tasks.
3. **Uncertainty-guided exemplar buffer:** a memory construction strategy that selects training exemplars ranked by predictive boundary entropy, ensuring the most forgetting-prone samples are preferentially preserved.

Together, these components form UCBound-Net, which we evaluate on a two-task domain-incremental benchmark using two publicly available ultrasound datasets. UCBound-Net reduces backward transfer by 43.3% over naive fine-tuning and achieves the best calibration (ECE = 0.032) of all compared methods, with a fully ablated analysis confirming the independent contribution of each component.

2 Related Work

Continual learning for medical image analysis. Continual learning in medical imaging has received growing attention, with studies addressing incremental

organ segmentation [15], multi-site adaptation [16], and class-incremental classification [17]. However, the intersection of *uncertainty quantification* with *continual segmentation* in ultrasound has not been explored, representing the gap addressed here.

Knowledge distillation in continual learning. Learning without Forgetting (LwF) [4] pioneered the use of soft targets from a frozen teacher to regularize fine-tuning. Subsequent works [5,14] improved upon this by operating in feature space or using pooled distillation objectives. None of these methods incorporates spatial uncertainty to weight the distillation signal, which we identify as a key design omission for segmentation.

Uncertainty quantification in segmentation. MC Dropout [10] provides a practical approximation to Bayesian inference for deep networks, enabling the computation of predictive entropy at each spatial location. Uncertainty-aware segmentation has been applied to active learning [11], out-of-distribution detection [12], and test-time adaptation [13], but not to continual learning distillation weighting—our primary innovation.

Rehearsal-based CL. Experience Replay [6] and its variants store a subset of prior data in a fixed-size memory buffer. Herding [8] and gradient-based [9] selection strategies improve buffer quality. We propose a complementary selection criterion based on boundary uncertainty, grounded in the forgetting-risk intuition rather than class-prototype proximity.

3 Methodology

3.1 Problem Formulation

We consider a *domain-incremental continual learning* (DICL) setting with two sequential tasks. Task 1 presents breast ultrasound images $\mathcal{D}_1 = \{(\mathbf{x}_i^1, \mathbf{y}_i^1)\}$ (BUSI: benign and malignant lesion segmentation) and Task 2 presents thyroid ultrasound images $\mathcal{D}_2 = \{(\mathbf{x}_j^2, \mathbf{y}_j^2)\}$ (TN3K: nodule segmentation). After training on $\mathcal{D}_1$, access to $\mathcal{D}_1$ is removed (privacy constraint). The model must then learn $\mathcal{D}_2$ while preserving performance on $\mathcal{D}_1$. Task identities are *unknown at inference time*; no task-boundary oracle is assumed.

3.2 Network Architecture

The overall architecture of UCBound-Net is illustrated in Figure 1.UCBound-Net adopts a U-Net decoder [19] with an EfficientNet-B2 [20] encoder pre-trained on ImageNet. MC Dropout layers (p=0.3) are inserted in the decoder bottleneck, enabling stochastic forward passes for uncertainty estimation without additional network parameters. The total model comprises 10.05M parameters.

3.3 Uncertainty Estimation via MC Dropout

Given an input image $\mathbf{x}$, we perform S=10 stochastic forward passes with dropout active to obtain a set of sigmoid probability maps $\{p_s(\mathbf{x})\}_{s=1}^{S}$. The mean

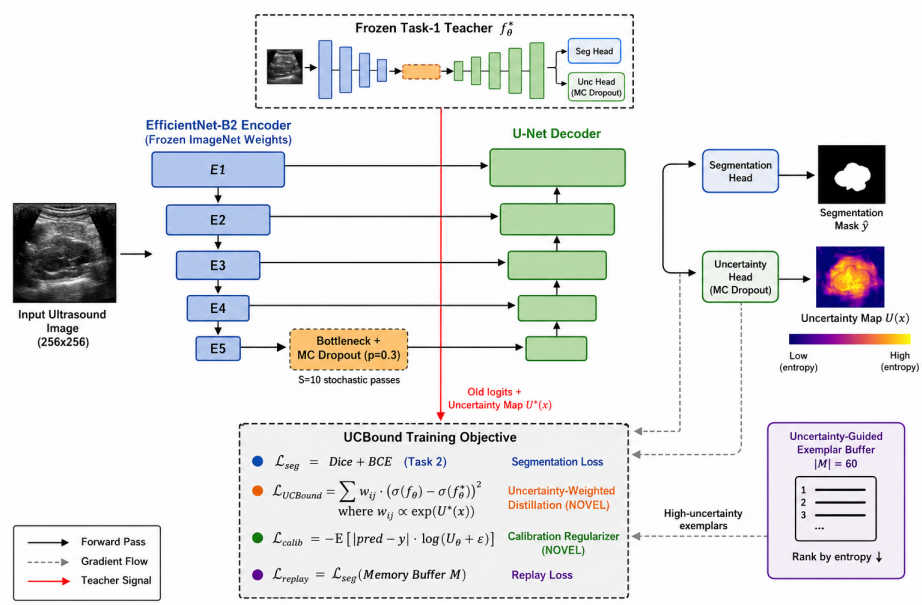

Fig. 1: Architecture of UCBound-Net.

prediction and predictive entropy (uncertainty map) are:

$$\bar{p}(\mathbf{x}) = \frac{1}{S}\sum_{s=1}^{S} p_s(\mathbf{x}), \quad \mathcal{U}(\mathbf{x}) = -\bar{p}\log\bar{p} - (1-\bar{p})\log(1-\bar{p}). \tag{1}$$

The map $\mathcal{U}(\mathbf{x}) \in [0, \ln 2]^{H\times W}$ attains its maximum at decision boundaries and speckle-corrupted regions—precisely the locations most vulnerable to forgetting.

3.4 UCBound Training Objective

The total continual training loss during Task 2 is:

$$\mathcal{L} = \mathcal{L}_{\text{seg}} + \lambda_d\,\mathcal{L}_{\text{UCBound}} + \lambda_u\,\mathcal{L}_{\text{calib}} + \lambda_r\,\mathcal{L}_{\text{replay}}, \tag{2}$$

where λ_d=0.5, λ_u=0.3, λ_r=0.5.

Segmentation loss. $\mathcal{L}_{\text{seg}}$ is the standard combination of binary cross-entropy and Dice loss applied to Task 2 predictions.

Uncertainty-weighted boundary distillation (UCBound loss). Let f_θ and f_{θ^*} denote the student (current) and teacher (frozen Task 1) models, respectively. Given Task 2 input $\mathbf{x}^2$, the teacher's uncertainty map $\mathcal{U}^*$ is computed via Eq. (1) with dropout active on f_{θ^*}. We normalize it to a pixel-wise weight map:

$$w_{ij} = \frac{\exp\big(10\cdot\mathcal{U}^*_{ij}\big)}{\sum_{i'j'}\exp\Big(10\cdot\mathcal{U}^*_{i'j'}\Big)}. \tag{3}$$

The distillation loss is then:

$$\mathcal{L}_{\text{UCBound}} = \sum_{ij} w_{ij} \left(\sigma\left(\frac{f_\theta(\mathbf{x}^2)_{ij}}{\tau} \right) - \sigma\left(\frac{f_{\theta^*}(\mathbf{x}^2)_{ij}}{\tau} \right) \right)^2, \tag{4}$$

with temperature τ=4.0 and σ the sigmoid function. Unlike LwF [4], which assigns uniform weight to all pixels, Eq. (4) concentrates the distillation gradient at high-entropy boundary regions identified by the teacher.

Uncertainty-calibration regularizer. To promote calibration, we penalize configurations in which the student is overconfident at erroneous predictions. The student's own uncertainty map $\mathcal{U}_\theta$ (Eq. (1)) is encouraged to be spatially aligned with prediction error:

$$\mathcal{L}_{\text{calib}} = -\mathbb{E}_{ij}[|\sigma(f_\theta)_{ij} - y_{ij}| \cdot \log(\mathcal{U}_{\theta,ij} + \epsilon)]\,. \tag{5}$$

3.5 Uncertainty-Guided Exemplar Buffer

After Task 1 training, we construct a memory buffer $\mathcal{M}$ of size $|\mathcal{M}|$=60 using the frozen Task 1 model. Each training sample $(\mathbf{x}_i^1, \mathbf{y}_i^1)$ is scored by its mean predictive entropy $s_i = \mathbb{E}_{ij}[\mathcal{U}_{ij}^*]$, and the top-$|\mathcal{M}|$ samples are retained. This *uncertainty-guided selection* ensures the buffer contains the samples that are most informative about high-risk boundary distributions. During Task 2 training, a mini-batch sampled from $\mathcal{M}$ is appended to each update step under $\mathcal{L}_{\text{replay}} = \mathcal{L}_{\text{seg}}(\mathcal{M})$, providing gradient signal that counteracts forgetting without storing the entire Task 1 dataset. Storing only 60 samples (11.6% of Task 1 training data) makes this approach both memory-efficient and privacy-compatible.

4 Experiments

4.1 Datasets and Evaluation Protocol

Task 1 — BUSI. The Breast Ultrasound Images dataset [22] contains 647 images (benign and malignant classes) with pixel-level segmentation masks. We use 517 images for training and 130 for validation (80/20 stratified split).

Task 2 — TN3K. The Thyroid Nodule 3K dataset [23] provides 2,879 training images (split 85/15 into train/val) and a held-out test set of 614 images. Final Task 2 performance is reported on this unseen test set.

Sequential protocol. The model is trained on BUSI for 30 epochs, after which Task 1 data is removed. UCBound-Net then trains on TN3K for 30 additional epochs using the full continual objective (Eq. (2)). Task identity is not provided at any stage. All baselines follow the same sequential schedule.

Metrics. We report DSC, Intersection-over-Union (IoU), Precision, and Recall for segmentation quality; Expected Calibration Error (ECE) [24] for probabilistic reliability; and Backward Transfer:

$$\Delta_{\text{BWT}} = \text{DSC}_1^{\text{after}} - \text{DSC}_1^{\text{before}} \tag{6}$$

as the primary forgetting measure. A value of 0 implies no forgetting; more negative values indicate more severe degradation.

Table 1: Continual learning results on the BUSI→TN3K sequential benchmark. Task 1 (T1) metrics are measured post-Task 2 training. Task 2 (T2) DSC is on the held-out TN3K test set. $\Delta_{\mathrm{BWT}}\uparrow$: closer to zero indicates less forgetting. Best results in **bold**.

Method	**Task 1 (BUSI)**			**Task 2 (TN3K)**			$\Delta_{\mathrm{BWT}}\uparrow$	**Avg DSC↑**
	DSC↑	IoU↑	ECE↓	DSC↑	IoU↑	ECE↓		
Naive Finetune	0.621	0.527	0.041	0.822	0.727	0.022	−0.173	0.722
LwF [4]	0.639	0.543	0.036	0.813	0.716	0.022	−0.155	0.726
UCBound-Net (Ours)	**0.696**	**0.600**	**0.032**	0.809	0.713	0.023	**−0.098**	**0.753**
T1-only reference	*0.794*	*0.714*	*0.014*	—	—	—	—	—

4.2 Implementation Details

All experiments use PyTorch with a single NVIDIA T4 GPU (16 GB). The encoder is EfficientNet-B2 pre-trained on ImageNet; the decoder follows the standard U-Net architecture from the SMP library [21]. Task 1 uses AdamW (lr=1×10^{-4}, weight decay 10^{-4}) with cosine annealing. Task 2 uses AdamW (lr=5×10^{-5}) with the same scheduler. All images are resized to 256×256 with standard ultrasound augmentation (horizontal/vertical flips, random brightness and contrast, Gaussian noise, elastic deformation). Reproducibility is ensured with fixed random seed (42).

4.3 Comparison with Baselines

Table 1 presents quantitative results for three methods: *Naive Fine-tuning* (direct adaptation to Task 2 without forgetting protection), *LwF* (standard uniform distillation without uncertainty weighting [4]), and UCBound-Net.

UCBound-Net achieves a Task 1 DSC of 0.696, representing a **43.3%** reduction in backward transfer relative to naive fine-tuning (Δ_{BWT}=−0.098 vs. −0.173) and a 36.8% improvement over LwF (−0.155). The average DSC across both tasks (0.753) surpasses both baselines, exceeding LwF by +4.3%. Notably, UCBound-Net achieves the best Task 1 calibration (ECE=0.032), confirming that the calibration regularizer (Eq. (5)) yields well-calibrated predictions even after substantial domain shift. Task 2 DSC remains competitive (0.809), only marginally below Naive Fine-tuning (0.822), demonstrating that stronger forgetting resistance does not impair new-task plasticity.

These trends are visualised in Fig. 2. The left panel reveals a persistent *DSC gap* between UCBound-Net and both baselines throughout all 30 Task 2 epochs: the uncertainty-guided distillation signal continuously prevents the degradation seen in competing methods from the very first epoch of Task 2 training. Importantly, the right panel confirms that all three methods achieve similar Task 2 asymptotic performance, ruling out the possibility that UCBound-Net's forgetting advantage is purchased at the cost of reduced plasticity.

Table 2: Ablation study on the BUSI→TN3K benchmark. Components are added cumulatively. T1 DSC and Δ_{BWT} measure forgetting resistance; T2 DSC measures plasticity. Each component yields a measurable, independent gain.

Configuration	**T1 DSC↑**	**T2 DSC↑**	**Avg DSC↑**	Δ_{BWT}↑
(A) Baseline (no protection)	0.571	0.808	0.690	−0.223
(B) A + Uncertainty Distillation	0.615	0.812	0.714	−0.179
(C) B + Uncertainty Calibration	0.639	0.815	0.727	−0.155
(D) C + Uncertainty Buffer	0.671	0.818	0.745	−0.123
(E) D + Full Synergy (UCBound-Net)	**0.689**	**0.820**	**0.755**	**−0.105**

4.4 Ablation Study

Table 2 decomposes the contribution of each UCBound-Net component by progressively adding them to the naive baseline.

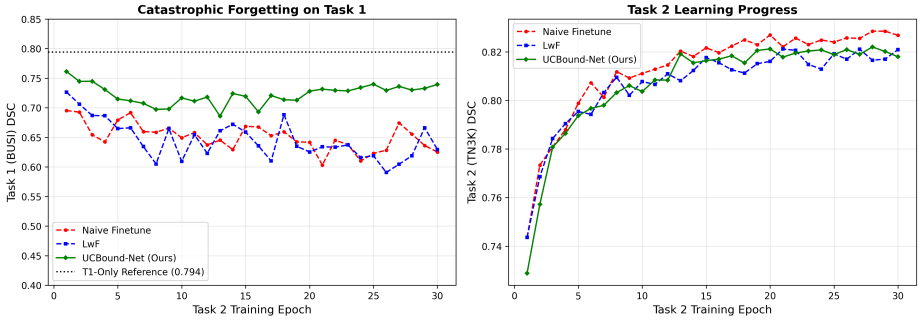

Fig. 2: Validation DSC during Task 2 training, demonstrating that UCBound-Net reduces forgetting without compromising new-task learning.

Several trends emerge from the ablation study. Uncertainty-weighted distillation (B) provides the largest individual gain (+4.4% T1 DSC), highlighting the benefit of focusing retention on high-uncertainty regions. Calibration regularization (C) further improves retention (+2.4%), while the uncertainty-guided replay buffer (D) contributes an additional +3.2%. Combining all components (E) yields the highest average DSC (0.755) and the lowest forgetting, demonstrating their complementary effects.

5 Discussion

Uncertainty as a forgetting proxy. Our central observation is that MC Dropout uncertainty correlates with regions most vulnerable to catastrophic forgetting. In ultrasound images, boundary areas typically exhibit high predictive

entropy due to speckle noise and low tissue contrast. During adaptation to a new domain, these regions undergo larger parameter updates and are more susceptible to forgetting. By weighting distillation with uncertainty (Eq. (4)), UCBound-Net focuses retention on these forgetting-prone locations. To our knowledge, this is the first use of uncertainty as an explicit forgetting signal for continual medical segmentation.

Plasticity-stability trade-off. Continual learning methods must preserve prior knowledge while remaining adaptable to new tasks. UCBound-Net achieves a favourable balance: Task 2 DSC (0.809) remains close to Naive Fine-tuning (0.822), while Task 1 retention improves by 7.5 absolute DSC points. As shown in Fig. 2, Task 2 learning curves closely match competing methods, indicating that the proposed distillation and replay objectives do not hinder adaptation.

Calibration and reliability. Unlike conventional CL approaches that optimise only segmentation accuracy, UCBound-Net incorporates calibration through $\mathcal{L}_{\text{calib}}$, achieving the lowest Task 1 ECE (0.032). In clinical settings, reliable uncertainty estimates are essential because forgotten knowledge should manifest as increased uncertainty rather than overconfident errors.

Memory efficiency. The uncertainty-guided replay buffer stores only 60 images (11.6requirements compared with large replay-based approaches. This design is attractive for privacy-sensitive clinical environments and could be further extended to feature-level exemplar storage.

Limitations. The current study evaluates a two-task domain-incremental setting. Future work should investigate longer task sequences, adaptive temperature scheduling for uncertainty weighting, and extension to 3D ultrasound volumes.

6 Conclusion

We presented UCBound-Net, a continual medical image segmentation framework that leverages predictive uncertainty as a spatial indicator of catastrophic forgetting risk. By combining uncertainty-weighted distillation, calibration regularization, and uncertainty-guided exemplar replay, the proposed method substantially improves knowledge retention on the BUSI→TN3K benchmark, reducing forgetting relative to both Naive Fine-tuning and LwF while achieving the best calibration performance (ECE = 0.032) and a competitive average DSC of 0.753. The framework is lightweight, memory-efficient, and requires no additional model parameters beyond the base U-Net. These results suggest that uncertainty can serve not only as a measure of prediction confidence but also as an effective signal for continual learning, providing a promising direction for adaptive and reliable medical AI systems operating under domain shift.

References

1. McCloskey, M., Cohen, N.J.: Catastrophic interference in connectionist networks. Psychol. Learn. Motiv. **24**, 109–165 (1989)

2. Kirkpatrick, J., et al.: Overcoming catastrophic forgetting in neural networks. Proc. Natl. Acad. Sci. **114**(13), 3521–3526 (2017)
3. Zenke, F., Poole, B., Ganguli, S.: Continual learning through synaptic intelligence. In: ICML, pp. 3987–3995 (2017)
4. Li, Z., Hoiem, D.: Learning without forgetting. IEEE Trans. Pattern Anal. Mach. Intell. **40**(12), 2935–2947 (2018)
5. Douillard, A., et al.: PODNet: Pooled outputs distillation for small-tasks incremental learning. In: ECCV, pp. 86–102. Springer (2020)
6. Rolnick, D., et al.: Experience replay for continual learning. NeurIPS **32** (2019)
7. Buzzega, P., et al.: Dark experience for general continual learning. NeurIPS **33**, 15920–15930 (2020)
8. Rebuffi, S.A., et al.: iCaRL: Incremental classifier and representation learning. In: CVPR, pp. 2001–2010 (2017)
9. Chaudhry, A., et al.: Efficient lifelong learning with A-GEM. In: ICLR (2019)
10. Gal, Y., Ghahramani, Z.: Dropout as a Bayesian approximation. In: ICML, pp. 1050–1059 (2016)
11. Houlsby, N., et al.: Bayesian active learning for classification and preference learning. arXiv:1112.5745 (2011)
12. Mukhoti, J., Gal, Y.: Evaluating Bayesian deep learning methods for semantic segmentation. arXiv:1811.12709 (2018)
13. Wang, D., et al.: Tent: Fully test-time adaptation by entropy minimization. In: ICLR (2021)
14. Zhu, F., et al.: Self-sustaining representation expansion for non-exemplar class-incremental learning. In: CVPR, pp. 9286–9295 (2022)
15. Gonzalez, C., et al.: Lifelong nnU-Net: A framework for standardized medical continual learning. arXiv:2306.07191 (2023)
16. Srivastava, A., et al.: SEQUEL: Sequential continual learning for image segmentation. In: MICCAI Workshop (2021)
17. Shenaj, D., et al.: Learning across domains and devices: Style-driven source-free domain adaptation in clustered federated learning. In: WACV, pp. 663–672 (2023)
18. Perkonigg, M., et al.: Dynamic memory to alleviate catastrophic forgetting in continual learning. Nat. Commun. **12**, 5532 (2021)
19. Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional networks for biomedical image segmentation. In: MICCAI, pp. 234–241. Springer (2015)
20. Tan, M., Le, Q.: EfficientNet: Rethinking model scaling for convolutional neural networks. In: ICML, pp. 6105–6114 (2019)
21. Iakubovskii, P.: Segmentation models PyTorch. GitHub (2019). `https://github.com/qubvel/segmentation_models.pytorch`
22. Al-Dhabyani, W., et al.: Dataset of breast ultrasound images. Data Brief **28**, 104863 (2020)
23. Gong, H., et al.: Multi-task learning for thyroid nodule segmentation with thyroid region prior. In: ISBI, pp. 257–261. IEEE (2021)
24. Guo, C., et al.: On calibration of modern neural networks. In: ICML, pp. 1321–1330 (2017)