Cost-effective Active Learning for Nucleus Detection Using Crowdsourced Annotations with Dynamic Weighting Adjustment

Jiao Tang[1], Yuankun Zu[2], Qi Zhu[1], Peng Wan[1], Daoqiang Zhang (✉)[1], and Wei Shao (✉)[1]

[1] College of Artificial Intelligence, Nanjing University of Aeronautics and Astronautics, Key Laboratory of Brain-Machine Intelligence Technology, Ministry of Education, Nanjing 211106, China
dqzhang@nuaa.edu.cn, shaowei20022005@nuaa.edu.cn
[2] School of Computer Science and Engineering, Southeast University, Nanjing 210096, China

Abstract. Accurate nucleus detection in pathology images is crucial for disease diagnosis. Deep learning based methods require extensive annotations of nuclei, which are time-consuming for pathologists. Active learning (AL) provides an attractive paradigm for reducing annotation efforts by iteratively selecting the most valuable samples for annotation. However, most AL methods do not consider utilizing crowdsourced annotations from multiple workers with varying expertise levels and labeling costs, limiting their practical applicability. Recent approaches design AL strategies that adaptively select the most cost-effective worker for each sample, but these methods solely focus on the classification task, overlooking the development of an AL framework with crowdsourced annotations for the detection task. Additionally, they struggle to adapt to the changes in model performance during AL iterations, resulting in inefficiencies in sample selection and cost management. Based on the above considerations, we propose C2AL, a novel cost-effective AL framework using crowdsourced annotations for nucleus detection in pathology images. Specifically, we design a new criterion in the form of score function and a dynamic weighting adjustment strategy to iteratively select the most cost-effective sample-worker pairs from the crowdsourced data. Then, based on the selected sample-worker pairs, the labeled pool is updated and the detection model is trained for performance evaluation. To the best of our knowledge, this is the first AL framework for detecting nuclei in the crowdsourced environment, and the experimental results on one real-world and two simulated crowdsourced datasets demonstrate that C2AL achieves higher detection accuracy at lower annotation costs compared to existing methods.

Keywords: Active Learning · Crowdsourcing · Nucleus Detection

1 Introduction

Nucleus detection in pathology images is crucial for cancer progression analysis [2]. Designing an effective deep learning based algorithm requires a large amount of labeled data, which increases the annotation burden for experts [10][1]. Active learning (AL) provides an attractive paradigm for reducing annotation efforts by iteratively selecting the most valuable samples for annotation [18], which makes the process of labeling new data more efficient. However, most AL methods do not consider utilizing crowdsourced annotations from multiple workers with varying expertise levels and labeling costs, limiting their practical applicability. As a matter of fact, specialists (SPs) are more likely to make correct decisions than non-pathologists (NPs). However, it is infeasible to assign all unlabeled images to SPs for annotation due to their high annotation costs [14]. On the other hand, we cannot rely solely on the annotation from NPs since they often struggle with labeling nuclei with complex morphology [15]. Hence, it is important to optimize the selection process to balance accurate detection with lower annotation costs.

Recent methods have focused on designing AL strategies that select the most cost-effective workers and samples, which can be broadly divided into two types: sequential and joint selection. Sequential selection is a two-step process involving sample selection followed by labeler assignment [29][8][7][28]. However, sequential strategies do not consider the workers' performance during sample selection, leading to the potential low-quality annotations. Joint selection, which considers both samples and workers simultaneously, has been proposed to address this issue [17][12]. Herde *et al.* [11] proposed the multi-worker probabilistic AL method for estimating annotation performance to improve the joint selection of sample-worker pairs. Similarly, Chakraborty *et al.* [4] framed the optimal sample and worker selection as a constrained optimization problem and used linear programming relaxation to select sample-worker batches.

Although much progress has been achieved, existing AL frameworks in crowdsourced environment solely focus on the classification task, the development of an AL framework with crowdsourced annotations for the detection task (i.e., nucleus detection) is overlooked. Intuitively, the detection task is more challenging since it includes both the classification and localization branches [16]. Additionally, existing methods struggle to adapt to changes in model performance during AL iterations, resulting in inefficiencies in sample selection and cost management. For instance, in the early stages, when the model lacks confidence and requires more informative samples, existing methods fail to prioritize them, leading to slower progress in model performance. As the model improves, these methods fail to adjust the worker selection accordingly, leading to unnecessary reliance on highly skilled annotators for simple samples, which reduces cost efficiency.

Based on the above considerations, we propose C2AL, a novel cost-effective AL framework for detecting various types of nuclei in pathology images under the crowdsourced environment. Specifically, we introduce a multi-criteria sample-worker selection process that considers sample uncertainty, worker cost, and worker credibility in each AL cycle. Furthermore, we introduce dynamic weighting adjustment, a strategy that adjusts the importance of each sample-

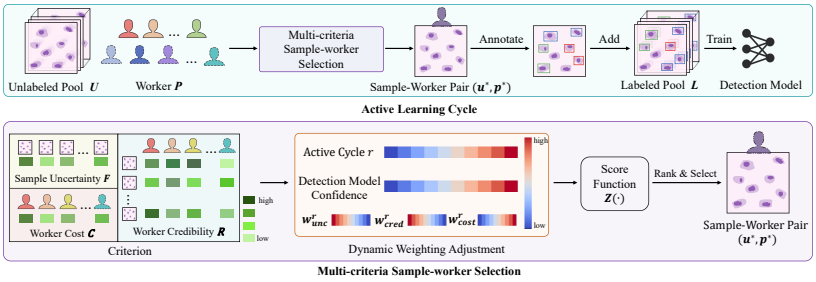

Fig. 1. Flowchart of C2AL. In each AL cycle, We combine sample uncertainty, worker credibility, and worker cost into a unified score function to select the most cost-effective sample-worker pairs for annotation. The method dynamically adjusts the weights based on model performance, ensuring efficient learning while minimizing annotation costs. The visual design is partly inspired by Figure 1 in [22].

worker pair based on the model's progress during the learning process. We evaluate the effectiveness of C2AL through experiments on one real-world and two simulated crowdsourced datasets. The results demonstrate that C2AL not only achieves higher detection accuracy but also reduces annotation costs compared to the existing methods.

2 Method

Flowchart of C2AL is shown in Fig.1. We start with a small labeled pool $\boldsymbol{L} = \{\boldsymbol{l}_k | k \in (1, \cdots, n_l)\}$ with n_l labeled images, and a large pool of unlabeled data $\boldsymbol{U} = \{\boldsymbol{u}_j | j \in (1, \cdots, n_u)\}$ with n_u unlabeled images. The labeled pool $\boldsymbol{L}$ contains ground-truth annotations $\boldsymbol{Y}$, and we have prior knowledge from annotations $\hat{\boldsymbol{Y}}$ provided by n_p workers $\boldsymbol{P} = \{\boldsymbol{p}_i | i \in (1, \cdots, n_p)\}$ on the labeled pool. The goal is to optimize the selection of sample-worker pairs from the unlabeled pool in each AL cycle. For unlabeled image $\boldsymbol{u}_j$ and worker $\boldsymbol{p}_i$, C2AL defines a score function $Z(\boldsymbol{u}_j, \boldsymbol{p}_i; r)$ based on the AL cycle r and three key criteria: Sample Uncertainty, Worker Credibility, and Worker Cost. The criteria are weighted and combined to identify the most cost-effective sample-worker pairs. Notably, the weights for each criterion are dynamically adjusted based on the model's progress during AL iterations. After ranking the scores, the most cost-effective sample-worker pair $(\boldsymbol{u}^*, \boldsymbol{p}^*)$ is selected for annotation. The annotated sample is then added to the labeled pool, and the model is updated accordingly. The process repeats until the cost budget is exhausted or the desired accuracy is achieved.

2.1 Sample Uncertainty: Identifying Informative Samples

Considering sample uncertainty is essential in AL, as informative samples can enhance model's generalization ability [27]. We use entropy [21] to quantify each

sample's uncertainty. A sample with higher entropy indicates greater uncertainty in the model's predictions, making it more valuable for annotation [6]. In the context of nucleus detection, we define the entropy of sample $\boldsymbol{u}_j$ as follows:

$$F(\boldsymbol{u}_j) = \max_{b \in (1,\cdots,B_j)} \sum_{\boldsymbol{g}_t} -H_b^t \cdot log(H_b^t), \tag{1}$$

where B_j denotes the number of predicted bounding boxes in $\boldsymbol{u}_j$, and H_b^t is the probability of the b-th bounding box being classified as category $\boldsymbol{g}_t$. The uncertainty of $\boldsymbol{u}_j$ is defined as the maximum entropy across all bounding boxes.

2.2 Worker Credibility: Assessing Worker Reliability

In object detection, which combines the classification and localization task [30], we consider both aspects to ensure reliable annotations. For the classification task, we assume that a worker's reliability is positively correlated with their ability to correctly classify different types of nuclei, which is derived from the prior knowledge $\hat{\boldsymbol{Y}}$ and the ground-truth labels $\boldsymbol{Y}$, calculated from the worker's accuracy in annotating each nucleus type $\boldsymbol{g}_t$, denoted as $h_i(\boldsymbol{g}_t)$. We then define the worker's ability to label image $\boldsymbol{l}_k$ as $q_i(\boldsymbol{l}_k)$, averaged across different nucleus types. For each unlabeled image $\boldsymbol{u}_j$, we assume the worker's labeling ability on $\boldsymbol{u}_j$ is consistent with that on $\boldsymbol{l}_k$ if their nucleus compositions are similar. The model predicts the nucleus composition of $\boldsymbol{u}_j$ as $\boldsymbol{v}_j = [\boldsymbol{v}_{jt}]$, where $\boldsymbol{v}_{jt}$ is the proportion of nucleus type $\boldsymbol{g}_t$ in $\boldsymbol{u}_j$. To estimate $q_i(\boldsymbol{u}_j)$, we select the K most similar images $\boldsymbol{I}_j = \{\boldsymbol{I}_{j1}, \ldots, \boldsymbol{I}_{jK}\}$ in the labeled pool $\boldsymbol{L}$ based on nucleus composition using cosine similarity, with corresponding compositions denoted as $\boldsymbol{E}_j = \{\boldsymbol{E}_{j1}, \cdots, \boldsymbol{E}_{jK}\}$. The worker's ability on $\boldsymbol{u}_j$ is computed as a weighted average:

$$q_i(\boldsymbol{u}_j) = \frac{\sum_{s=1}^{K} \langle \boldsymbol{v}_j, \boldsymbol{E}_{js} \rangle q_i(\boldsymbol{I}_{js})}{\sum_{s=1}^{K} \langle \boldsymbol{v}_j, \boldsymbol{E}_{js} \rangle}, \tag{2}$$

where $\langle \cdot \rangle$ denotes the cosine similarity [25]. For the localization task, we assess the worker's ability based on the average intersection-over-union (IoU) [20] o_i and the missing detection rate m_i, computed by comparing the worker's annotations $\hat{\boldsymbol{Y}}$ with the ground truth $\boldsymbol{Y}$. The overall credibility of worker $\boldsymbol{p}_i$ for annotating image $\boldsymbol{u}_j$ is then given by the combination of classification ability $q_i(\boldsymbol{u}_j)$, localization IoU o_i, and missing detection rate m_i:

$$R_i(\boldsymbol{u}_j) = q_i(\boldsymbol{u}_j) \cdot (1 - m_i) \cdot o_i. \tag{3}$$

Thus, workers with higher classification accuracy, IoU, and lower missing detection rate are considered more reliable for annotating image $\boldsymbol{u}_j$. Notably, a worker's credibility is not fixed but varies depending on the specific image. For instance, even a less specialized worker may excel at annotating certain types of nuclei [3], and we should make full use of their expertise.

2.3 Worker Cost: Optimizing Annotation Costs

Intuitively, workers with higher expertise and accuracy naturally demand higher compensation, as accurate labeling requires specialized knowledge and time from experts like pathologists [3]. Thus, the cost for worker $\boldsymbol{p}_i$, denoted as C_i, can be modeled based on their labeling ability. We use $h_i(\boldsymbol{g}_t)$ (defined in Sec. 2.2) to represent the cost for annotating nuclei of type $\boldsymbol{g}_t$. The total cost for worker $\boldsymbol{p}_i$ is the average of $h_i(\boldsymbol{g}_t)$ across all nucleus types:

$$C_i = \frac{1}{n_g} \sum_{t=1}^{n_g} h_i(\boldsymbol{g}_t), \tag{4}$$

where n_g is the number of nucleus types. Therefore, workers with more expertise in nucleus annotation come with higher costs and vice versa.

2.4 Score Function with Dynamic Weighting Adjustment

We propose an evaluation function $Z(\cdot)$ for active selection, combining sample uncertainty $F(\boldsymbol{u}_j)$, worker credibility $R_i(\boldsymbol{u}_j)$, and worker cost C_i. To select the most cost-effective sample-worker pair $(\boldsymbol{u}^*, \boldsymbol{p}^*)$, we aim to maximize uncertainty $F(\boldsymbol{u}_j)$ and worker credibility $R_i(\boldsymbol{u}_j)$, while minimizing cost C_i. The score function is defined as:

$$Z(\boldsymbol{u}_j, \boldsymbol{p}_i; r) = \frac{w_{unc}^r F(\boldsymbol{u}_j) \cdot w_{cred}^r R_i(\boldsymbol{u}_j)}{w_{cost}^r C_i}, \tag{5}$$

where r denotes the r-th AL cycle, and w_{unc}^r, w_{cred}^r, w_{cost}^r are dynamic weights that adjust based on the current performance and progress of the model. Initially, when the model is less confident, more weights are given to sample uncertainty and worker credibility to explore uncertain regions and ensure reliable annotations. As the model improves and gains confidence, the focus shifts toward minimizing worker cost, optimizing annotation expenses. The weights are defined as:

$$w_{unc}^r = \frac{1}{1 + e^{-F_{avg}^r}}, w_{cred}^r = \frac{1}{1 + e^{-F_{avg}^r}}, w_{cost}^r = \frac{1}{1 + e^{F_{avg}^r}}, \tag{6}$$

where F_{avg}^r represents the average uncertainty for unlabeled images in the r-th AL cycle. After ranking the scores in each cycle, the most cost-effective sample-worker pair is selected as:

$$(\boldsymbol{u}^*, \boldsymbol{p}^*) = \max_{\boldsymbol{u}_j, \boldsymbol{p}_i} Z(\boldsymbol{u}_j, \boldsymbol{p}_i; r). \tag{7}$$

The annotated sample is then added to the labeled pool, followed by model updating. The process repeats until the cost budget is exhausted or the desired accuracy is achieved.

3 Experiment

Datasets. We evaluate C2AL on the NuCLS [3] and Lizard [9] datasets. NuCLS is a crowdsourced dataset containing over 220,000 labeled nuclei from the BRCA cohort of TCGA [23]. It includes 1,797 images, of which 53 have multi-rater annotations and ground truth, while the remaining 1,744 contain only ground truth. Lizard is a comprehensive dataset designed for colonic nucleus detection, containing 208 images and 495,179 labeled nuclei. It is worth noting that Lizard dose not include crowdsourced annotations. To address the issue of limited multi-rater images in NuCLS, we conduct experiments on the real crowdsourced NuCLS dataset and extend the 1,744 images from NuCLS and Lizard into simulated crowdsourced datasets, referred to as SimNuCLS and SimLizard, respectively. For SimNuCLS, we calculate the averaged Intersection over Union (IoU, o_i), missing detection rate (m_i), and classification accuracy ($h_i(\boldsymbol{g}_t)$) for each worker based on the multi-rater annotations, and extend the 1,744 images using the following rules: **Bounding Box Offset:** let N_i denote the number of labeled bounding boxes by $\boldsymbol{p}_i$ in multi-rater annotations, with f_i of these having IoU greater than S, where S is the maximum averaged IoU between ground-truth bounding boxes and annotations across all workers. We simulate $\boldsymbol{p}_i$'s annotations by adjusting the positions of Bo_i percentage of bounding boxes compared with ground truth, where $Bo_i = (N_i - f_i)/N_i$. A worker with better bounding box annotation ability will have fewer adjusted bounding boxes, and vice versa. Based on Bo_i, we calculate the shift amount Sa_i for each adjustment along a random direction (i.e., top, bottom, left, right) as $Sa_i = (S - o_i)/Bo_i$. The shift amount is larger for workers with lower IoU (o_i). **Bounding Box Missing:** To simulate missing bounding boxes, we randomly drop some boxes based on the worker's missing detection rate m_i. **Nucleus Type Mis-classification:** For each bounding box, the nucleus type with ground truth of $\boldsymbol{g}_t$ is misclassified with probability $1 - h_i(\boldsymbol{g}_t)$. For SimLizard, we apply the same rules but generate random values for o_i, m_i, and $h_i(\boldsymbol{g}_t)$ for 4 SPs and 6 NPs. Specifically, for SPs, o_i is in the range of (0.60-0.80), m_i in (0.10-0.20), and $h_i(\boldsymbol{g}_t)$ in (0.80-0.90). For NPs, o_i is in (0.40-0.60), m_i in (0.20-0.30), and $h_i(\boldsymbol{g}_t)$ in (0.30-0.40).

Experimental Settings. For each dataset, we randomly split it into five folds, with four for training and remaining for evaluation. We randomly select 10% samples from training set as initial labeled set. In each AL cycle, we set the number of samples required annotation as 5% of the training set. We use Faster R-CNN [19] as the detection model, and K in Eq. 2 is fixed as 5. We use mean Average Precision (mAP) at IoU threshold of 0.5 as the evaluation metric.

Comparison with the State-of-the-Arts. We compare C2AL with the following 8 baseline methods to evaluate its effectiveness: *i*) **Random**: Randomly select sample-worker pairs. *ii*) **MaPAL** [11]: A multi-worker probabilistic AL method to improve the joint selection of sample-worker pairs. *iii*) **LNCL** [5]:

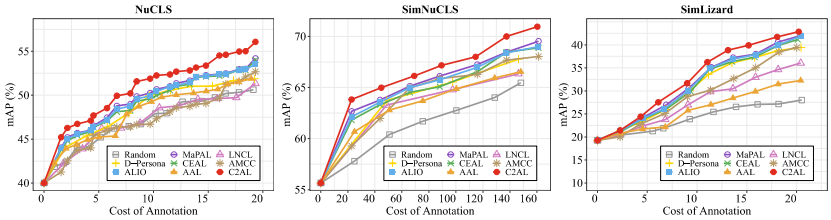

Fig. 2. The mAP curves with total costs increasing on three datasets.

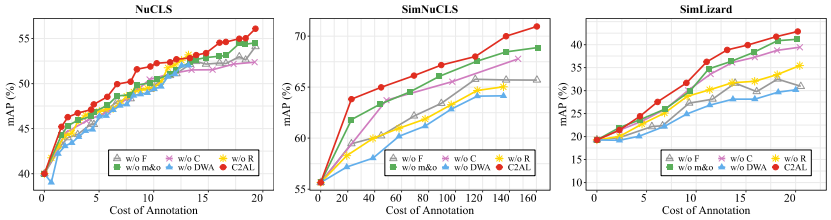

Fig. 3. Ablation study on the NuCLS, SimNuCLS, and SimLizard datasets.

Learn from noisy crowd labels using logics. *iv*) **D-Persona** [24]: A unified framework for multi-rater medical image analysis. *v*) **CEAL** [12]: An AL framework under the crowdsourced environment using score function. *vi*) **AMCC** [26]: A cost-effective AL framework under crowdsourced environment using the alternative minimization strategy. *vii*) **ALIO** [4]: Optimize the selection of sample-worker pairs by solving a constrained optimization problem. *viii*) **AAL** [13]: A multi-server multi-worker framework for AL. For methods without AL process (LNCL and D-Persona), we use entropy to assess sample uncertainty (introduced in Sec. 2.1) and select the most informative sample in each AL cycle. Fig. 2 plots the mAP curves as total costs increase across all three datasets. It is obviously that C2AL consistently outperforms other methods. On one hand, for the same detection performance, C2AL generally requires lower annotation costs, as it selects the most cost-effective sample-worker pairs that can reduce the total annotation costs. On the other hand, C2AL achieves higher mAP at a given cost. For example, on the real crowdsourced NuCLS dataset, C2AL achieves an mAP of 51.89% at a cost of 10, while its top competitors, MaPAL [11] and ALIO [4], reach 50.25% and 49.94%, respectively. Similar trends are observed in the simulated datasets (SimNuCLS and SimLizard). Moreover, the mAP on SimNuCLS is higher than on NuCLS, as SimNuCLS includes more images for training. These results demonstrate C2AL's potential to achieve strong nucleus detection performance with sufficient crowdsourced data.

Ablation Study. To further evaluate the effectiveness of C2AL, we compare it with several variants: **w/o F**: Select sample-worker pairs without considering sample uncertainty (Eq. 1). **w/o C** : Select sample-worker pairs without

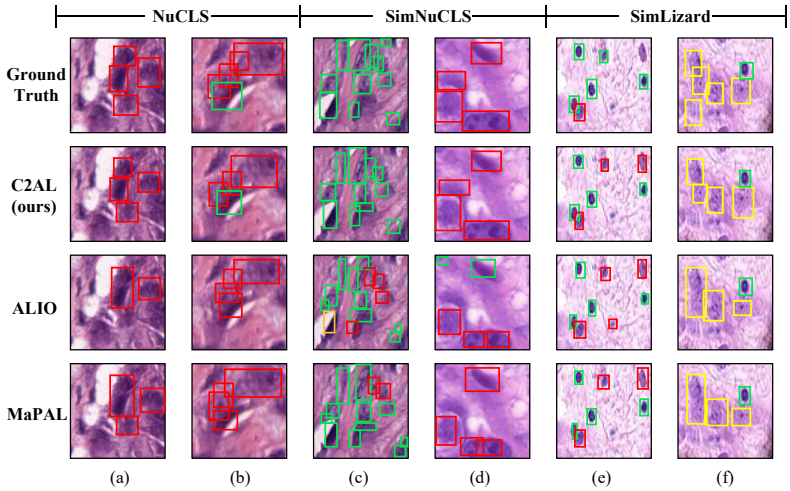

Fig. 4. Comparison of the visualization results among C2AL, ALIO and MaPAL at the same annotation cost on the NuCLS, SimNuCLS, and SimLizard datasets.

considering worker cost (Eq. 4). **w/o R**: Select sample-worker pairs without considering worker credibility (Eq. 3). **w/o m&o**: Only considering the classification branch to calculate the reliability of labeling (only keep the first term in Eq. 3). **w/o DWA**: Select sample-worker pairs without Dynamic Weighting Adjustment. As shown in Fig. 3, C2AL outperforms **w/o F**, **w/o C**, and **w/o R**, highlighting the importance of considering sample uncertainty, worker cost, and worker credibility for a cost-effective AL algorithm. Additionally, C2AL consistently achieves higher mAP than **w/o m&o**, indicating that both classification and localization branches are critical for object detection. Notably, in the NuCLS ablation study, **w/o R** exceeds C2AL at a certain cost because it neglects the reliability of workers and selects a large number of low-cost NPs for annotation. In such situations, more labeled images are involved in **w/o R** that may temporally lead to higher mAP values. However, once all images are labeled, C2AL outperforms **w/o R**. Finally, C2AL surpasses **w/o DWA**, as the dynamic weighting adjustment allows for more effective selection of sample-worker pairs by adaptively adjusting the importance of factors such as sample uncertainty and worker cost, resulting in higher accuracy at a lower cost.

Visualization Results. Fig. 4 presents sample visualizations of C2AL, ALIO [4], and MaPAL [11] on the NuCLS, SimNuCLS, and SimLizard datasets at the same annotation cost, with ALIO and MaPAL as the best competitors. It can be observed that C2AL outperforms both methods across all datasets. Specifically, in nucleus classification (b,c,d), C2AL shows better accuracy than ALIO and MaPAL. Additionally, in comparison with ALIO and MaPAL, C2AL can successfully detect touching nuclei (a,f). On the contrary, ALIO and MaPAL

tend to over-detect large single nucleus as multiple components (d), whereas C2AL detects them accurately. Moreover, C2AL has lower missing detection rates (a,d) and over-detection rates (e) compared to ALIO and MaPAL.

4 Conclusion

In this paper, we propose a novel AL framework C2AL for detecting different types of nuclei in a crowdsourced environment. C2AL aims to achieve accurate detection while minimizing annotation costs by using a dynamic weighting-adjusted score function to select cost-effective sample-worker pairs. We evaluate C2AL on both simulated and real crowdsourced datasets, and the results show that C2AL outperforms the existing methods, highlighting its potential for practical applications in nucleus detection under the crowdsourced environment.

Acknowledgments. This work was supported by the National Natural Science Foundation of China (Nos. 62136004, 62272226, 62102188), Key Research and Development Plan of Jiangsu Province, China under Grant BE2022842, Postdoctoral Fellowship Program of CPSF (No. GZC20242233).

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

References

1. Abels, E., Pantanowitz, L., Aeffner, F., Zarella, M.D., Van der Laak, J., Bui, M.M., Vemuri, V.N., Parwani, A.V., Gibbs, J., Agosto-Arroyo, E., et al.: Computational pathology definitions, best practices, and recommendations for regulatory guidance: a white paper from the digital pathology association. The Journal of pathology **249**(3), 286–294 (2019)
2. Ahn, J.C., Teng, P.C., Chen, P.J., Posadas, E., Tseng, H.R., Lu, S.C., Yang, J.D.: Detection of circulating tumor cells and their implications as a biomarker for diagnosis, prognostication, and therapeutic monitoring in hepatocellular carcinoma. Hepatology **73**(1), 422–436 (2021)
3. Amgad, M., Atteya, L.A., Hussein, H., Mohammed, K.H., Hafiz, E., Elsebaie, M.A., Alhusseiny, A.M., AlMoslemany, M.A., Elmatboly, A.M., Pappalardo, P.A., et al.: Nucls: A scalable crowdsourcing approach and dataset for nucleus classification and segmentation in breast cancer. GigaScience **11** (2022)
4. Chakraborty, S.: Asking the right questions to the right users: Active learning with imperfect oracles. In: Proceedings of the AAAI conference on artificial intelligence. vol. 34, pp. 3365–3372 (2020)
5. Chen, Z., Sun, H., He, H., Chen, P.: Learning from noisy crowd labels with logics. In: 2023 IEEE 39th International Conference on Data Engineering (ICDE). pp. 41–52. IEEE (2023)
6. Deng, Y.: Uncertainty measure in evidence theory. Science China Information Sciences **63**, 1–19 (2020)
7. Fang, M., Yin, J., Tao, D.: Active learning for crowdsourcing using knowledge transfer. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 28 (2014)

8. Fang, M., Zhu, X., Li, B., Ding, W., Wu, X.: Self-taught active learning from crowds. In: 2012 IEEE 12th international conference on data mining. pp. 858–863. IEEE (2012)
9. Graham, S., Jahanifar, M., Azam, A., Nimir, M., Tsang, Y.W., Dodd, K., Hero, E., Sahota, H., Tank, A., Benes, K., et al.: Lizard: a large-scale dataset for colonic nuclear instance segmentation and classification. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 684–693 (2021)
10. Hartman, D.J., Van Der Laak, J.A., Gurcan, M.N., Pantanowitz, L.: Value of public challenges for the development of pathology deep learning algorithms (2020)
11. Herde, M., Kottke, D., Huseljic, D., Sick, B.: Multi-annotator probabilistic active learning. In: 2020 25th International Conference on Pattern Recognition (ICPR). pp. 10281–10288. IEEE (2021)
12. Huang, S.J., Chen, J.L., Mu, X., Zhou, Z.H.: Cost-effective active learning from diverse labelers. In: IJCAI. pp. 1879–1885 (2017)
13. Huang, S.J., Zong, C.C., Ning, K.P., Ye, H.: Asynchronous active learning with distributed label querying. In: IJCAI. pp. 2570–2576 (2021)
14. Irshad, H., Montaser-Kouhsari, L., Waltz, G., Bucur, O., Nowak, J., Dong, F., Knoblauch, N.W., Beck, A.H.: Crowdsourcing image annotation for nucleus detection and segmentation in computational pathology: evaluating experts, automated methods, and the crowd. In: Pacific Symposium on Biocomputing Co-chairs. pp. 294–305 (2014)
15. Karachiwalla, R., Pinkow, F.: Understanding crowdsourcing projects: A review on the key design elements of a crowdsourcing initiative. Creativity and innovation management **30**(3), 563–584 (2021)
16. Kaur, R., Singh, S.: A comprehensive review of object detection with deep learning. Digital Signal Processing **132**, 103812 (2023)
17. Moon, S., Carbonell, J.G.: Proactive learning with multiple class-sensitive labelers. In: 2014 International Conference on Data Science and Advanced Analytics (DSAA). pp. 32–38. IEEE (2014)
18. 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 computing surveys (CSUR) **54**(9), 1–40 (2021)
19. Ren, S., He, K., Girshick, R., Sun, J.: Faster r-cnn: Towards real-time object detection with region proposal networks. Advances in neural information processing systems **28** (2015)
20. Rezatofighi, H., Tsoi, N., Gwak, J., Sadeghian, A., Reid, I., Savarese, S.: Generalized intersection over union: A metric and a loss for bounding box regression. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 658–666 (2019)
21. Shewry, M.C., Wynn, H.P.: Maximum entropy sampling. Journal of Applied Statistics **14**(2), 165–170 (1987)
22. Tang, J., Yue, Y., Wan, P., Wang, M., Zhang, D., Shao, W.: Osal-nd: Open-set active learning for nucleus detection. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 351–361. Springer (2024)
23. Tomczak, K., Czerwińska, P., Wiznerowicz, M.: Review the cancer genome atlas (tcga): an immeasurable source of knowledge. Contemporary Oncology/Współczesna Onkologia **2015**(1), 68–77 (2015)
24. Wu, Y., Luo, X., Xu, Z., Guo, X., Ju, L., Ge, Z., Liao, W., Cai, J.: Diversified and personalized multi-rater medical image segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11470–11479 (2024)

25. Xia, P., Zhang, L., Li, F.: Learning similarity with cosine similarity ensemble. Information Sciences **307**, 39–52 (2015)
26. Yu, G., Tu, J., Wang, J., Domeniconi, C., Zhang, X.: Active multilabel crowd consensus. IEEE transactions on neural networks and learning systems **32**(4), 1448–1459 (2020)
27. Yuan, T., Wan, F., Fu, M., Liu, J., Xu, S., Ji, X., Ye, Q.: Multiple instance active learning for object detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5330–5339 (2021)
28. Zhao, L., Zhang, Y., Sukthankar, G.: An active learning approach for jointly estimating worker performance and annotation reliability with crowdsourced data. arXiv preprint arXiv:1401.3836 (2014)
29. Zhong, J., Tang, K., Zhou, Z.H.: Active learning from crowds with unsure option. In: IJCAI. pp. 1061–1068 (2015)
30. Zou, Z., Chen, K., Shi, Z., Guo, Y., Ye, J.: Object detection in 20 years: A survey. Proceedings of the IEEE (2023)