MDMR: BALANCING DIVERSITY AND REDUNDANCY FOR ANNOTATION-EFFICIENT FINE-TUNING OF PRETRAINED CELL SEGMENTATION MODELS

Eiram Mahera Sheikh, Alaa Tharwat, Constanze Schwan, Wolfram Schenck

Department of Engineering and Mathematics, Bielefeld University of Applied Sciences and Arts, Bielefeld, Germany

This work has been submitted to the IEEE for possible publication. Copyright may be transferred without notice, after which this version may no longer be accessible.

ABSTRACT

Pretrained cell segmentation models have simplified and accelerated microscopy image analysis, but they often perform poorly on challenging datasets. Although these models can be adapted to new datasets with only a few annotated images, the effectiveness of fine-tuning depends critically on which images are selected for annotation. To address this, we propose MDMR (Maximum Diversity Minimum Redundancy), a novel algorithm that selects the most informative subset of images by explicitly balancing diversity and redundancy in feature space. We evaluate MDMR under an extremely low annotation budget of just two images per dataset for fine-tuning the pretrained Cellpose Cyto2 model on four different 2D datasets from the Cell Tracking Challenge. MDMR consistently outperforms six competitive active learning and subset selection methods and approaches the performance of fully-supervised fine-tuning. The results show that explicitly balancing diversity and redundancy enables stable and annotation-efficient adaptation of pretrained cell segmentation models. Code is publicly available at: https://github.com/eiram-mahera/mdmr.

***Index Terms*—** Cell instance segmentation, Cell Tracking Challenge, Subset Selection, Active Learning

1. INTRODUCTION

Accurate cell segmentation is fundamental to quantitative microscopy, as it underpins downstream analyses such as cell tracking, quantitative morphology assessment, and phenotype characterization. However, manual segmentation is slow, requires expert supervision, and does not scale. Classical image-processing pipelines based on thresholding, watershed transforms, or morphological operations often fail for images with complex cellular structures, overlapping regions, or modality-specific artifacts. Deep learning models have achieved state-of-the-art accuracy [1], but they typically require large, diverse, and carefully annotated datasets for training. Acquiring such datasets for every experimental condition remains costly and time-consuming.

To reduce the need for exhaustive labeling, recent research has focused on developing pretrained generalist segmentation models that can generalize across multiple cell types and microscopy modalities. Cellpose [2, 3] is one such model, trained on heterogeneous microscopy data to capture transferable representations of cellular morphology. These models deliver strong out-of-the-box performance, but their accuracy can still degrade on new and challenging datasets. In these cases, fine-tuning the pretrained network on a small number of annotated images from the target dataset can substantially improve performance. However, the effectiveness of fine-tuning depends critically on which images are chosen. Selecting a small subset of images that will have the greatest impact on downstream performance after fine-tuning is therefore a central challenge in achieving efficient and targeted model adaptation.

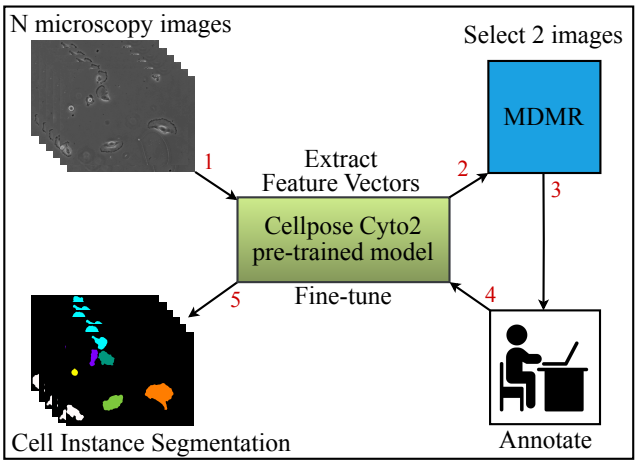

Fig. 1: Overview of the proposed MDMR pipeline. Features are extracted from microscopy images using the pretrained Cellpose Cyto2 model. MDMR selects two images that are maximally diverse and minimally redundant for annotation; these annotated images are used to fine-tune the model, which is finally used for segmentation.

This challenge has been addressed in part by active learning and subset-selection strategies, which aim to identify the most informative samples for annotation based on criteria such as uncertainty, diversity, or representativeness [4, 5, 6, 7, 8, 9, 10, 11, 12]. These approaches typically involve multiple rounds of sample selection, annotation, and model retraining. While such iterative schemes can be effective, they are impractical for biomedical image segmentation, where annotating a single image may require minutes to hours of expert work and repeated fine-tuning adds significant computational cost. Moreover, most prior studies evaluate their methods under moderate annotation budgets, often tens or hundreds of labeled images, which still represent substantial human effort for instance-level segmentation and do not reflect the extreme low-budget scenarios that many laboratories actually face.

To address these limitations, we propose MDMR (Maximum Diversity Minimum Redundancy), a novel algorithm for selecting a highly informative subset of images for fine-tuning pretrained segmentation models. MDMR explicitly balances diversity and redun-

dancy in the feature space of a pretrained network. Maximizing diversity ensures broad coverage of the target dataset, while minimizing redundancy avoids selecting visually or morphologically similar images. This explicit balance enables more stable and efficient fine-tuning, particularly under very small annotation budgets, and requires only a single selection step without iterative retraining or additional supervision. The approach follows the same high-level goals as diversity- and representativeness-based active learning and subset selection strategies, but it is designed specifically for the practical constraints of biomedical image analysis workflows.

We compare MDMR against representative methods from three major families of active learning and subset selection that also rely on diversity and representativeness criteria: clustering-based selection [13, 14], geometric coverage approaches [15, 16], and diversity-based sampling techniques [17, 18, 19]. Together, these baselines capture the main algorithmic perspectives used in prior work and provide a rigorous benchmark for evaluating our method. All methods are evaluated under an extremely low annotation budget of only two images per dataset, reflecting practical constraints in biomedical research where annotation time is a limiting factor. Experiments are conducted using the pretrained Cellpose Cyto2 model [3] on four 2D datasets from the Cell Tracking Challenge (CTC)[1]. Across all datasets, MDMR achieves higher segmentation accuracy than competing baselines and approaches the performance of fully-supervised fine-tuning, demonstrating that explicitly balancing diversity and redundancy enables stable and annotation-efficient fine-tuning of pretrained cell segmentation models under minimal supervision.

2. METHOD

We propose **MDMR (Maximum Diversity Minimum Redundancy)**, a novel algorithm that identifies a subset of informative samples by jointly maximizing diversity and minimizing redundancy in feature space. MDMR can operate on any type of feature representation: raw, handcrafted, or learned, and therefore integrates smoothly with pretrained segmentation models. In this work, we apply MDMR to the problem of annotation-efficient fine-tuning of a pretrained cell segmentation model under an extremely small labeling budget.

2.1. Feature Extraction and Overview

Figure 1 summarizes the overall workflow. First, we extract feature representations from all microscopy images in the dataset using the pretrained Cellpose Cyto2 model [3]. For each image, we obtain a 256-dimensional feature vector (referred to as a *style vector* in [2, 3]) that captures global image characteristics, including texture, cellular morphology, and imaging modality. Next, the proposed MDMR algorithm operates in this feature space to select two images that are maximally diverse and minimally redundant. These selected images are then manually annotated to obtain ground-truth instance segmentation masks. The annotated images are used to fine-tune the pretrained model, adapting it to the target dataset. Finally, the fine-tuned model is applied to segment the remaining unlabeled images from the same dataset.

2.2. MDMR Algorithm

Algorithm 1 outlines the full procedure. Let $\mathcal{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$ denote the set of N feature vectors in $\mathbb{R}^d$, and let B be the available annotation budget. Below we describe each component of MDMR.

[1] celltrackingchallenge.net

Algorithm 1: MDMR

Input: Feature vectors $\{\mathbf{x}_i\}_{i=1}^N \in \mathbb{R}^{N\times d}$; budget B.
Output: Set $\mathcal{S} \subset \{1, \ldots, N\}$ with $|\mathcal{S}| = B$.

1. Compute kernel matrix:
   $K[i,j] = \exp(-\gamma\|\mathbf{x}_i - \mathbf{x}_j\|^2) \quad \forall\, i,j \in \{1,\ldots,N\}$
   with $\gamma = 1/d$
2. Identify central anchor:
   $c = \arg\max_{i\in\{1,\ldots,N\}} \sum_{j=1}^N K[i,j]$.
3. Initialize $\mathcal{S} \leftarrow \{c\}$.
4. **while** $|\mathcal{S}| - 1 < B$ **do**:
   (a) For each $i \notin \mathcal{S}$, compute diversity score:
   $D(i) = 1 - \frac{1}{|\mathcal{S}|}\sum_{j\in\mathcal{S}} K[i,j]$
   (b) Let $D_{\max} = \max_i D(i)$ and $\mathcal{T} = \{i : D(i) = D_{\max}\}$
   (c) Let $r = B - (|\mathcal{S}| - 1)$.
   (d) If $|\mathcal{T}| \le r$ then:
   add all indices in $\mathcal{T}$ to $\mathcal{S}$.
   Else:
   randomly sample r indices from $\mathcal{T}$ (without replacement) and add them to $\mathcal{S}$.
5. $\mathcal{S} \leftarrow \mathcal{S} \setminus \{c\}$.
6. Return $\mathcal{S}$.

2.2.1. *Similarity computation*

We compute a Gaussian radial basis function (RBF) kernel [20] K, to quantify pairwise similarity between feature vectors:

$$K[i,j] = \exp\big(-\gamma\|\mathbf{x}_i - \mathbf{x}_j\|^2\big), \tag{1}$$

where γ controls the rate at which similarity decays with distance between the features. Following a common heuristic in high-dimensional feature spaces, we set

$$\gamma = \frac{1}{d}, \tag{2}$$

where d is the dimensionality of the feature vectors. This scaling maintains a consistent similarity range across datasets with different feature dimensions and provides a stable basis for measuring diversity and redundancy in the subsequent selection steps. The resulting matrix $K \in \mathbb{R}^{N\times N}$ forms the foundation for the MDMR subset selection algorithm.

2.2.2. *Central anchor initialization*

Using the similarity matrix K defined in Eq. (1), MDMR identifies the index c of the feature vector that is most similar to all other feature vectors in the dataset. Formally,

$$c = \underset{i\in\{1,\ldots,N\}}{\arg\max} \sum_{j=1}^{N} K[i,j]. \tag{3}$$

This index corresponds to the *central anchor sample*. The anchor serves as a fixed reference for measuring diversity and redundancy among the remaining samples during subset selection. Importantly, the sample corresponding to index c is *not* selected for annotation, it is used only as a reference to guide the diversity-based selection process.

2.2.3. Set initialization

Let $\mathcal{S}$ denote the set of indices corresponding to the samples selected for annotation. The selection process begins by including the central anchor index c obtained from Eq. (3):

$$\mathcal{S} \leftarrow \{c\}. \tag{4}$$

The anchor is included here for computing diversity scores, but it is not part of the final annotated subset. After the greedy selection procedure, the anchor c is removed from $\mathcal{S}$ to yield the final set of samples chosen for annotation.

2.2.4. Greedy subset selection

For each feature vector $\mathbf{x}_i$ with index $i \notin \mathcal{S}$, MDMR computes a *diversity score* that measures how dissimilar $\mathbf{x}_i$ is from the samples already in $\mathcal{S}$:

$$D(i) = 1 - \frac{1}{|\mathcal{S}|} \sum_{j \in \mathcal{S}} K[i, j]. \tag{5}$$

A higher $D(i)$ indicates that sample i is more diverse and less redundant with respect to the currently selected subset. At each iteration, the sample (or samples) with the highest diversity score is added to $\mathcal{S}$. If multiple samples have the same maximum score, MDMR includes all of them if the remaining annotation budget allows; otherwise, it randomly samples from the tied candidates to ensure unbiased tie resolution. This process continues until the desired annotation budget B is reached. Finally, the set $\mathcal{S}$ is returned excluding the central anchor c.

2.3. Complexity analysis

The computational cost of MDMR is dominated by the construction of the similarity matrix K. Computing the RBF kernel for N samples of dimensionality d requires $\mathcal{O}(N^2 d)$ time and $\mathcal{O}(N^2)$ memory. The subsequent greedy selection stage adds an additional $\mathcal{O}(BN)$ operations, where B is the annotation budget. Since $B \ll N$ in practical scenarios (e.g., B=2 in our experiments), the overall runtime is primarily determined by the kernel computation. Consequently, the total complexity of MDMR scales quadratically with the number of samples and linearly with the feature dimensionality. In practice, the algorithm remains efficient due to its vectorized implementation.

3. EXPERIMENTS & RESULTS

3.1. Datasets

We evaluate MDMR on four 2D datasets from the Cell Tracking Challenge. **DIC-C2DH-HeLa** has dense clusters of cells with very weak boundary contrast, DIC-enhanced internal structures, and the coexistence of flat interphase cells with bright and rounded mitotic cells. **Fluo-N2DH-GOWT1** has heterogeneous fluorescence signal inside nuclei due to unlabeled nucleoli and strong variability in intensity across cells. **PhC-C2DH-U373** has halo and shade-off artifacts from phase-contrast imaging combined with highly irregular cell shapes. **BF-C2DL-HSC** has strong visibility of the slide and additional debris that makes cell segmentation by a model difficult. To mitigate this, we crop the images to regions containing cells. In our experiments, we use four slightly different crops to ensure that results are not overly dependent on a specific crop size. Each dataset represents a different modality and cell type, with distinct challenges. We have mentioned the main challenges for each dataset here, for additional details, we refer the reader to the CTC website and papers [21, 22, 23].

3.2. Evaluation Metric

We use the **SEG** measure as our main evaluation metric, following the official CTC protocol. It is derived from the Jaccard index:

$$J(S, R) = \frac{|R \cap S|}{|R \cup S|},$$

where S is a segmented mask and R a reference mask. A prediction counts as matching if $|R \cap S| > 0.5 \cdot |R|$. Each reference object can be matched at most once. SEG is the mean Jaccard index across all matched objects in a video, yielding values between 0 (no match) and 1 (perfect match). CTC provides two reference annotation sets: sparse *gold truth* (GT) manually drawn by experts, and dense *silver truth* (ST) obtained by fusing outputs from top-performing algorithms. We evaluate against ST masks using the CTC evaluation software.

3.3. Baseline Methods

We benchmark MDMR against six competitive active learning and subset selection strategies spanning three methodological families:

Clustering-based methods: Clustering provides a natural strategy for subset selection by grouping similar samples and choosing representatives from each cluster. **K-means** clustering [13] partitions data into k clusters and selects points closest to centroids, ensuring representative coverage of the data distribution. **TypiClust** [14] improves on this idea by choosing typical (high-density) samples within clusters, which has been shown to be especially effective in low-budget regimes.

Coverage-based methods: Another family of approaches aims to cover the geometry of the data distribution in feature space. **CoreSet** [15] formulates subset selection as a k-center problem, where samples are chosen to minimize the maximum distance between any data point and the selected set. **Facility Location** [16] instead casts the problem as submodular optimization, selecting a subset that ensures each point in the dataset is well-represented by at least one chosen element. These approaches balance representativeness and diversity by emphasizing global coverage.

Diversity-based methods: These approaches directly optimize for diversity without relying on clustering or explicit geometric coverage. **FADS** [19] is a greedy diversity subsampling algorithm that iteratively selects samples with maximal dissimilarity, producing subsets that capture the variability of large datasets. Determinantal Point Processes (**DPP**) [17, 18, 24] provides a probabilistic formulation, assigning higher probability to subsets with low redundancy. By modeling negative correlations between items, DPPs offer a principled framework for diversity-driven selection.

3.4. Experimental Setup

All experiments use the pretrained Cellpose Cyto2 model [3] for feature extraction and fine-tuning. Following common practice, we use sequence 01 for training (subset selection and fine-tuning) and sequence 02 for evaluation across all datasets. Each method selects exactly two images ($B = 2$) for annotation. Models are fine-tuned on the selected images for up to 200 epochs. The optimal number of epochs is chosen by leave-one-out cross-validation (LOOCV)

Table 1: SEG scores (mean ± std over 100 runs) for all methods across four 2D CTC datasets. Pretrained: Cellpose Cyto2 without fine-tuning. Fully Supervised: fine-tuning using all the images. Boldface indicates the best method using 2 annotated images for fine-tuning.

Method	BF-C2DL-HSC	DIC-C2DH-HeLa	Fluo-N2DH-GOWT1	PhC-C2DH-U373
Pretrained	0.763	0.020	0.906	0.258
MDMR (ours)	**0.858 ± 0.008**	**0.842 ± 0.000**	**0.949 ± 0.000**	**0.857 ± 0.000**
Core-Set [15]	0.588 ± 0.275	0.807 ± 0.021	0.947 ± 0.002	0.839 ± 0.021
DPP [17, 18, 24]	0.615 ± 0.296	0.802 ± 0.024	0.946 ± 0.003	0.844 ± 0.010
FADS [19]	0.632 ± 0.320	0.807 ± 0.021	0.944 ± 0.004	0.848 ± 0.004
Facility Location [16]	0.703 ± 0.217	0.833 ± 0.000	0.946 ± 0.000	0.845 ± 0.000
K-means Clustering [13]	0.723 ± 0.089	0.816 ± 0.017	0.946 ± 0.001	0.844 ± 0.010
TypiClust [14]	0.699 ± 0.183	0.824 ± 0.000	0.947 ± 0.000	0.842 ± 0.000
Random Sampling	0.487 ± 0.279	0.797 ± 0.040	0.942 ± 0.005	0.845 ± 0.019
Fully Supervised	0.902 ± 0.021	0.842 ± 0.008	0.951 ± 0.001	0.850 ± 0.010

over the two annotated samples, performed independently for each method and dataset. In the full-data regime, we fine-tune on all training images with a validation split of 20% to select the best epoch (max 200). We use the same hyperparameters for all experiments: stochastic gradient descent with a learning rate 0.1 and weight decay 1×10^{-4}. To account for stochasticity in subset selection and fine-tuning, every experiment is repeated 100 times. We report mean ± standard deviation. We also report the performance of the pretrained Cellpose Cyto2 model (without fine-tuning) for context.

4. RESULTS AND DISCUSSION

Table 1 summarizes the performance of all methods across the four CTC datasets. Several consistent trends can be observed.

MDMR achieves the best overall performance. Across all datasets, MDMR attains the highest mean SEG score. On *BF-C2DL-HSC*, where strong brightfield artifacts and debris make segmentation particularly challenging, MDMR reaches 0.858, substantially higher than the next best method (0.723 for K-means) and far above random sampling (0.487). On *DIC-C2DH-HeLa*, which features densely packed cells with varying morphologies, MDMR achieves 0.842, matching the fully supervised result and outperforming all baselines. For the fluorescence dataset *Fluo-N2DH-GOWT1*, MDMR again yields the top score (0.949), though with smaller margins since most methods perform well for this dataset. Finally, on *PhC-C2DH-U373*, characterized by halo and shade-off artifacts, MDMR achieves 0.857, clearly ahead of competing approaches.

Performance approaches full supervision. In three of the four datasets, MDMR nearly matches the performance of fully supervised fine-tuning: 0.842 vs. 0.842 on HeLa, 0.949 vs. 0.951 on GOWT1, and 0.857 vs. 0.850 on U373. The only noticeable gap occurs on HSC (0.858 vs. 0.902), likely due to the fact that the video contains over 1,700 frames, and just two images cannot fully represent the dataset. Nevertheless, these results demonstrate that even minimal annotation can yield significant and stable performance gains, providing a strong foundation for incremental extension when more labels become available.

Baselines exhibit redundancy and instability. Clustering- and coverage-based methods such as Core-Set, Facility Location, and K-means perform reasonably on some datasets but lack consistency across modalities. Their results show high variance, particularly on HSC, where the selected samples often overlap in appearance or fail to capture key variations. Diversity-only methods such as DPP and FADS perform less reliably, sometimes selecting highly distinct but uninformative samples, which leads to unstable fine-tuning outcomes. Random sampling, as expected, remains the weakest and most inconsistent baseline.

Why MDMR succeeds. MDMR's advantage arises from its explicit balance between representativeness, diversity, and non-redundancy within the pretrained feature space. K-means partitions the data into clusters and selects samples near cluster centers, making them locally representative not globally. TypiClust favors samples from the most dense regions of the clusters, but this often neglects rare yet informative samples that differ from the dominant modes. Core-Set and Facility Location aim to span the feature space through geometric spread, yet they may overlook globally representative samples because they prioritize boundary points. DPP and FADS encourage dissimilarity between samples, which can overemphasize atypical or uninformative examples. MDMR unifies these complementary ideas: it first anchors selection on the sample most similar to all others, capturing the dataset's global structure, and then greedily adds samples that are maximally diverse yet minimally redundant relative to those already chosen. This combination yields an informative subset that captures both representativeness and diversity, supporting stable and efficient fine-tuning even when only two annotated images are available.

5. CONCLUSION AND FUTURE WORK

We introduced the Maximum Diversity Minimum Redundancy (MDMR), a novel algorithm that selects a small subset of samples by explicitly balancing diversity and redundancy in the feature space. We demonstrated its effectiveness for selecting the most informative images for fine-tuning the pretrained Cellpose Cyto2 model under an extremely low annotation budget of only two images. Despite its conceptual simplicity and deterministic nature, MDMR achieves competitive performance, highlighting the value of principled subset selection for annotation-efficient learning. The design of MDMR is inherently model- and task-agnostic as well as domain-independent, making it applicable across a broad spectrum of datasets. Future work will focus on a systematic evaluation of MDMR with different pretrained backbones and data modalities, and on extending the comparison to more advanced subset selection and active learning strategies.

6. COMPLIANCE WITH ETHICAL STANDARDS

This research study was conducted using the data publicly available at https://celltrackingchallenge.net/2d-datasets/. Ethical approval was not required as confirmed by the license attached with the open access data.

7. ACKNOWLEDGMENTS

This research was conducted within the framework of the project "SAIL: SustAInable Lifecycle of Intelligent SocioTechnical Systems" (grant no. NW21-059B). SAIL is receiving funding from the programme " Netzwerke 2021", an initiative of the Ministry of Culture and Science of the State of North Rhine-Westphalia, Germany. The sole responsibility for the content of this publication lies with the authors.

8. REFERENCES

[1] Ching-Wei Wang, Wei-Tang Lee, and Ting-Sheng Su, "A survey of deep learning methods on cell instance segmentation," *Neural Computing and Applications*, pp. 1–70, 2025.

[2] Carsen Stringer, Tim Wang, Michalis Michaelos, and Marius Pachitariu, "Cellpose: a generalist algorithm for cellular segmentation," *Nature methods*, vol. 18, no. 1, pp. 100–106, 2021.

[3] Marius Pachitariu and Carsen Stringer, "Cellpose 2.0: how to train your own model," *Nature methods*, vol. 19, no. 12, pp. 1634–1641, 2022.

[4] Benjamin Rombaut, Joris Roels, and Yvan Saeys, "Biosegment: Active learning segmentation for 3d electron microscopy imaging.," in *IAL@ PKDD/ECML*, 2022, pp. 7–26.

[5] Shuo Zhao, Yu Zhou, and Jianxu Chen, "Active learning pipeline for biomedical image instance segmentation with minimal human intervention," in *BVM Workshop*. Springer, 2025, pp. 217–222.

[6] Jiajun Cao, Jan Wenzel, Shanghang Zhang, Josephine Lampe, Hongxiao Wang, Jiachen Yao, Zhicheng Zhang, Shuo Zhao, Yu Zhou, Chao Chen, et al., "Rethinking deep learning in bioimaging through a data centric lens," *npj Imaging*, vol. 3, no. 1, pp. 29, 2025.

[7] Miguel Molina-Moreno, Marcel P Schilling, Markus Reischl, and Ralf Mikut, "Asap: Automated style-aware similarity measurement for selection of annotated pre-training datasets in 2d biomedical imaging," *IEEE Access*, 2025.

[8] Daniele Falcetta, Hava Chaptoukaev, Francesco Galati, and Maria A. Zuluaga, "One-shot active learning for vessel segmentation," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2025*, James C. Gee, Daniel C. Alexander, Jaesung Hong, Juan Eugenio Iglesias, Carole H. Sudre, Archana Venkataraman, Polina Golland, Jong Hyo Kim, and Jinah Park, Eds., Cham, 2026, pp. 470–480, Springer Nature Switzerland.

[9] Zhifan Jiang, Vishwesh Nath, Holger R Roth, Abhijeet Parida, Nicholas Foreman, Michael J Fisher, Roger J Packer, Syed Muhammad Anwar, Robert A Avery, and Marius George Linguraru, "Kalm: Knowledge-driven active learning for medical image segmentation using localized similarity," in *2025 IEEE 22nd International Symposium on Biomedical Imaging (ISBI)*. IEEE, 2025, pp. 1–5.

[10] Anyang Ji, Qingbo Kang, Wei Xu, Changfan Wang, Kang Li, and Qicheng Lao, "Confounder-aware medical data selection for fine-tuning pretrained vision models," in *2025 IEEE 22nd International Symposium on Biomedical Imaging (ISBI)*. IEEE, 2025, pp. 1–5.

[11] Hyeongu Kang, Mujin Kim, Young Sin Ko, Yesung Cho, and Mun Yong Yi, "Wise: Efficient wsi selection for active learning in histopathology," *Computerized Medical Imaging and Graphics*, vol. 118, pp. 102455, 2024.

[12] Carsten T. Lüth, Jeremias Traub, Kim-Celine Kahl, Till J. Bungert, Lukas Klein, Lars Krämer, Paul F Jaeger, Fabian Isensee, and Klaus Maier-Hein, "nnactive: A framework for evaluation of active learning in 3d biomedical segmentation," *Transactions on Machine Learning Research*, 2025.

[13] John Burkardt, "K-means clustering," *Virginia Tech, Advanced Research Computing, Interdisciplinary Center for Applied Mathematics*, vol. 5, 2009.

[14] Guy Hacohen, Avihu Dekel, and Daphna Weinshall, "Active learning on a budget: Opposite strategies suit high and low budgets," *arXiv preprint arXiv:2202.02794*, 2022.

[15] Ozan Sener and Silvio Savarese, "Active learning for convolutional neural networks: A core-set approach," *arXiv preprint arXiv:1708.00489*, 2017.

[16] Kai Wei, Rishabh Iyer, and Jeff Bilmes, "Submodularity in data subset selection and active learning," in *International conference on machine learning*. PMLR, 2015, pp. 1954–1963.

[17] Alexei Borodin, "Determinantal point processes," *arXiv preprint arXiv:0911.1153*, 2009.

[18] Alex Kulesza, Ben Taskar, et al., "Determinantal point processes for machine learning," *Foundations and Trends® in Machine Learning*, vol. 5, no. 2–3, pp. 123–286, 2012.

[19] Boyang Shang, Daniel W Apley, and Sanjay Mehrotra, "Diversity subsampling: Custom subsamples from large data sets," *INFORMS Journal on Data Science*, vol. 2, no. 2, pp. 161–182, 2023.

[20] Corinna Cortes and Vladimir Vapnik, "Support-vector networks," *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.

[21] Martin Maška, Vladimír Ulman, David Svoboda, Pavel Matula, Petr Matula, Cristina Ederra, Ainhoa Urbiola, Tomás España, Subramanian Venkatesan, Deepak MW Balak, et al., "A benchmark for comparison of cell tracking algorithms," *Bioinformatics*, vol. 30, no. 11, pp. 1609–1617, 2014.

[22] Vladimír Ulman, Martin Maška, Klas EG Magnusson, Olaf Ronneberger, Carsten Haubold, Nathalie Harder, Pavel Matula, Petr Matula, David Svoboda, Miroslav Radojevic, et al., "An objective comparison of cell-tracking algorithms," *Nature methods*, vol. 14, no. 12, pp. 1141–1152, 2017.

[23] Martin Maška, Vladimír Ulman, Pablo Delgado-Rodriguez, Estibaliz Gómez-de Mariscal, Tereza Nečasová, Fidel A Guerrero Peña, Tsang Ing Ren, Elliot M Meyerowitz, Tim Scherr, Katharina Löffler, et al., "The cell tracking challenge: 10 years of objective benchmarking," *Nature Methods*, vol. 20, no. 7, pp. 1010–1020, 2023.

[24] Guillaume Gautier, Guillermo Polito, Rémi Bardenet, and Michal Valko, "DPPy: DPP Sampling with Python," *Journal of Machine Learning Research - Machine Learning Open Source Software (JMLR-MLOSS)*, 2019, Code at http://github.com/guilgautier/DPPy/ Documentation at http://dppy.readthedocs.io/.