CellFMCount: A Fluorescence Microscopy Dataset, Benchmark, and Methods for Cell Counting

Abdurahman Ali Mohammed[1,6], **Catherine Fonder**[2,3,5,6], **Ying Wei**[1,6],
Wallapak Tavanapong[1,6], **Donald S. Sakaguchi**[2,3,4,5,6], **Qi Li**[1,6], **Surya K. Mallapragada**[2,3,4,5,6]
[1]*Department of Computer Science*
[2]*Department of Genetics, Development, and Cell Biology*
[3]*Molecular, Cellular, and Developmental Biology Program*
[4]*Neuroscience Program* [5]*Nanovaccine Institute*
[6]*Iowa State University, Ames, IA 50011*

***Abstract*—Accurate cell counting is essential in various biomedical research and clinical applications, including cancer diagnosis, stem cell research, and immunology. Manual counting is labor-intensive and error-prone, motivating automation through deep learning techniques. However, training reliable deep learning models requires large amounts of high-quality annotated data, which is difficult and time-consuming to produce manually. Consequently, existing cell-counting datasets are often limited, frequently containing fewer than 500 images. In this work, we introduce a large-scale annotated dataset comprising $3{,}023$ images from immunocytochemistry experiments related to cellular differentiation, containing over $430{,}000$ manually annotated cell locations. The dataset presents significant challenges: high cell density, overlapping and morphologically diverse cells, a long-tailed distribution of cell count per image, and variation in staining protocols. We benchmark three categories of existing methods: regression-based, crowd-counting, and cell-counting techniques on a test set with cell counts ranging from 10 to $2{,}126$ cells per image. We also evaluate how the Segment Anything Model (SAM) can be adapted for microscopy cell counting using only dot-annotated datasets. As a case study, we implement a density-map-based adaptation of SAM (SAM-Counter) and report a mean absolute error (MAE) of 22.12, which outperforms existing approaches (second-best MAE of 27.46). Our results underscore the value of the dataset and the benchmarking framework for driving progress in automated cell counting and provide a robust foundation for future research and development.**

***Index Terms*—Cell counting, Deep Learning, Density map estimation, Segment Anything Model**

I. Introduction

Cell counting is a fundamental task in biomedical research, essential for quantifying cell populations, studying cellular dynamics, and investigating complex biological processes. Accurate and scalable cell counting is critical for disease diagnosis, monitoring disease progression [1], [2], helping identify biomarkers [3], [4], and evaluating treatment responses [5]. It supports drug discovery by screening therapeutic candidates and analyzing drug efficacy, and is indispensable in regenerative medicine to assess treatment outcomes.

This work is partially supported by the National Science Foundation under Grant No. 2152117. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

In neural regeneration research, precise quantification of neural stem cells, neurons, and glial cells across developmental stages and injury models helps to uncover mechanisms underlying neurogenesis and repair. These insights have direct implications for the treatment of neurological disorders such as stroke, spinal cord injury, and neurodegenerative diseases. Cell counting is critical for stem cell therapy research, which attempts to use adult stem cells from patients [6]–[10] to treat their peripheral nerve injuries.

Traditionally, cell counting has relied heavily on manual annotation, which is labor-intensive, time-consuming, and challenging to scale for large datasets or images with dense cells (i.e., high-density cell images). Three paradigms of automated methods have emerged: the *detection-based*, *regression-based*, and *density map estimation* methods.

Detection-based methods localize individual cells using bounding boxes or segmentation masks, typically via convolutional neural networks or transformer-based detectors. While offering high interpretability and enabling instance-level analysis, these methods often struggle with dense regions due to overlapping cells and non-maximal suppression errors. Regression-based methods directly predict a single scalar count, reducing annotation effort but sacrificing the ability to localize cells and making annotation difficult to trace. Density map estimation (DME) approaches predict a 2D density map where the sum of all pixel values corresponds to the total count. DME was widely used for crowd counting [11], [12], of which few were adopted for cell counting. This approach provides spatial context, robustness to occlusion, and the ability to train from limited annotations (dot annotations) compared to detection-based methods.

Despite significant advancements, automated cell counting remains a challenge. Existing automated methods are hindered by the scarcity of large, diverse, and consistently annotated datasets. Public benchmarks with limited annotations, such as [13]–[17], have driven the field forward but are limited in scale, density variation, and biological diversity. **Challenges for fluorescence microscopy cell counting** include extreme variability in cell densities (number of cells per image), ranging from fewer than ten to thousands per image, cell type, cell

TABLE I
MICROSCOPY CELL COUNTING DATASETS WITH DOT ANNOTATIONS

Datasets	Cell Types	# Images	#Cells	Mean CPI ± std	Image Size
VGG [13]	Bacterial	200	35,192	176 ± 61	256x256
MBM [14]	Bone marrow	44	5,553	126 ± 33	600x600
ADI [16]	Adipose tissue	200	31,017	165 ± 44	150x150
DCC [15]	Various types	176	5,906	34 ± 22	306×322 to 798×788
IDCIA [17]	AHPC	262	22,155	84 ± 104	800x600
CellFMCounter (ours)	AHPC & rpc	3,023	431,321	143 ± 316	600×447 to 1600×1200

***AHPC**: Adult Hippocampal Progenitor Cells. **rpc**: Murine Retinal Progenitor Cells. **CPI**: #Cells per image*
Synthetic Brightfield Fluorescence

development stage, staining strategies, image magnification, and lack of clear object contrast as seen in crowd counting.

Contributions. To overcome these limitations, we introduce **CellFMCount**, a large-scale, annotated fluorescence microscopy dataset explicitly designed to facilitate robust and generalizable cell counting. CellFMCount contains 3,023 images from diverse biological contexts, imaging protocols, and staining conditions, featuring over 430,000 annotated cells. Each cell is labeled with a dot near its visual center, providing scalable supervision without requiring extensive segmentation annotations. Beyond serving as a cell-counting benchmark, CellFMCount supports numerous downstream tasks. Its variability enables exploration of effective data augmentation and domain adaptation. Dot annotations facilitate weak supervision approaches, such as pseudo-labeling for detection-based cell counting. Additionally, the extensive scale and diversity of the dataset are suitable for investigating sample-efficient learning paradigms such as active learning.

We also present an adaptation of the Segment Anything Model (SAM) [18], SAM-Counter, for density-map estimation (DME)–based cell counting. Our implementation pairs SAM's pretrained encoder with a lightweight density estimation head to generate density maps, showing competitive accuracy without requiring full segmentation masks.

In summary, we make the following contributions.

- Introduce **CellFMCount**, a large-scale, high-diversity, annotated fluorescence microscopy dataset for training and evaluating cell counting models under highly varying number of cells per image and appearance variation.
- Demonstrate how SAM can be repurposed for DME-based counting through our SAM-Counter, highlighting the feasibility of leveraging foundation-model features for microscopy analysis.
- Benchmark **thirteen cell-counting methods** with dot annotations as training ground truth. The methods include regression-based and density-based techniques, leading models adapted from the crowd-counting literature, and our SAM-Counter. Our evaluations provide unified and extensive baselines reflective of real-world complexity, setting a foundation for future research in automated microscopy-based cell counting.

All data, annotations, code, and pretrained models have been released to promote transparency, reproducibility, and progress in automated microscopy analysis. The CellFMCount dataset is available at https://doi.org/10.5281/zenodo.17088532, and the code repository (including training scripts and pretrained SAM-Counter models) is available at https://github.com/NRT-D4/CellFMCount.

II. RELATED WORK

This section describes existing datasets and methods for cell counting.

A. Cell Datasets

Many datasets were developed to support research in automatic cell counting and analysis, each highlighting different imaging techniques, biological samples, and annotation strategies.

Cell datasets with segmentation masks or bounding boxes: The ACCT dataset includes fluorescent images of neurons and microglia and corresponding segmentation masks for detection-based counting [19]. Similarly, the CIDACC dataset captures the particular difficulties in counting dense and irregularly shaped Chlorella vulgaris cells [20]. The Nasal Cytology Dataset (NCD) addresses clinical diagnosis by including nasal cell images with annotated bounding boxes, accounting for rare cell types and uneven distributions [21].

Cell datasets with dot annotations: Table I lists existing datasets in this category. ***Synthetic Datasets:*** One of the earlier and widely used datasets is the VGG Synthetic Fluorescence Microscopy Dataset [13], which contains computer-generated images that mimic the look of real fluorescence microscopy images. ***Real datasets:*** They better reflect the challenges found in actual biological samples. The Modified Bone Marrow (MBM) [14] and Adipose Tissue (ADI) [16] datasets include images of blood and fat tissue, respectively. These datasets highlight the variability in cell shape and distribution that occurs naturally across different tissue types. The Dublin Cell Counting (DCC) dataset features a variety of cell images captured under different settings for evaluation of how well models generalize to new data [15]. The IDCIA dataset has annotated fluorescence microscopy images of cells labeled with different antibody markers, which can significantly alter their appearance [17].

Our CellFMCount dataset addresses the limitations of existing cell datasets in terms of dataset sizes, significant variability in cell densities, cell type, antibody and cell staining strategies, and image magnification.

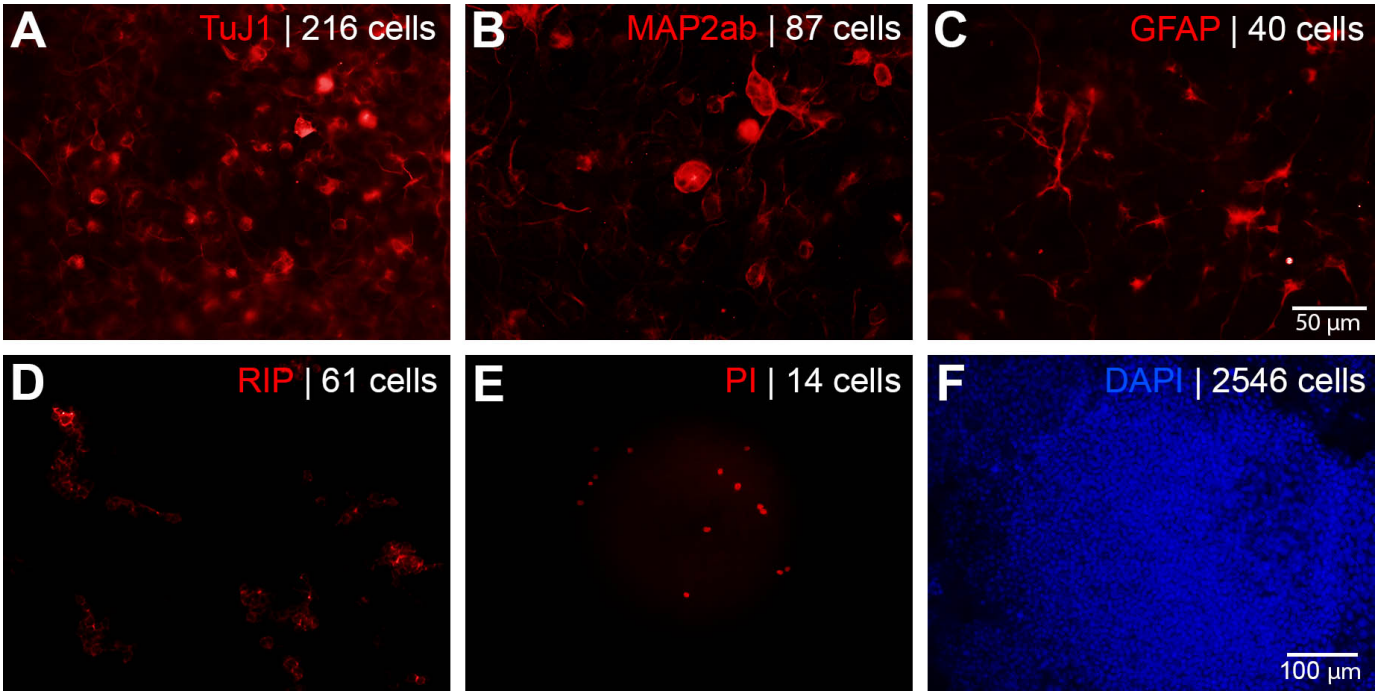

Fig. 1. **Cell-type diversity and morphological variation across immunolabeled samples.** Cells were immunolabeled with antibodies for markers of specific cell types, including immature neurons (TuJ1; A), maturing neurons (MAP2ab; B), astrocytes (GFAP; C), and oligodendrocytes (RIP; D). Cells were also stained with a cell viability dye to mark for dead cells (PI; E) as well as a nuclear stain (DAPI; F). Images show drastic differences in both cell morphology and density, ranging from sparse cells (e.g., PI, 14 cells) to dense fields (e.g., DAPI, 2546 cells). Scale bar = 50 μm for 40x image fields (Row 1) or 100 μm for 20x image fields (Row 2). Images were pseudo-colored for visualization.

B. Methods for Cell Counting

Early techniques combined intensity thresholding, edge detection, and watershed-like algorithms to separate touching objects, then counted connected components [22]. Although these methods are fast and annotation-free, they struggle when cell shape and contrast vary widely. Tools such as CellProfiler [23] have extended traditional pipelines by integrating segmentation models with user-guided parameter tuning. However, they require users to specify object properties or provide annotated masks for model training or refinement, limiting scalability and automation, especially in heterogeneous or dense datasets.

Further developments involved methods focused on learning mappings from local image features to a density map. Lempitsky and Zisserman [13] pioneered this line of research by employing linear regression with dense SIFT features to predict density maps. Subsequent work replaced linear regression with regression forests to enhance density estimation accuracy. Arteta et al. [24] extended this pipeline by introducing a local feature vocabulary and ridge regression in an interactive framework.

Recent methods shift toward deep neural networks. Xie et al. [25] applied fully convolutional regression networks (FCRN) with Gaussian-filtered outputs, while Count-ception [16] used sliding-window aggregation over receptive fields to construct density maps, improving accuracy but risking overfitting in background regions. SAUNet [26] enhanced U-Net with self-attention and modified batch normalization for small datasets, and Xue et al. [27] proposed partitioning images into subregions processed by separate networks. Wang et al. [28] introduced a concatenated fully convolutional regression network with auxiliary (CFCRN+Aux) supervision to improve intermediate feature learning and generalization across datasets.

A zero-shot SAM-based approach for cell counting was introduced [29], but is less accurate compared to the convolution neural network models trained on the IDCIA dataset. These efforts demonstrate SAM's versatility in addressing microscopy challenges, from complex morphologies to scalable analysis.

III. Proposed Dataset: Data Curation

We curated the images from stem cell differentiation experiments using mouse-derived Retinal Progenitor Cells (rpcs) and rat-derived Adult Hippocampal Progenitor Cells (AHPCs) to investigate the effects of different stimuli on neural progenitor cell differentiation. Differentiating stem cells into mature and functional cells requires carefully designed experiments. A typical cell differentiation experiment might involve comparing the effect of different growth factor concentrations on cell differentiation of the cell population during a seven-day period.

Fig. 2 shows the CellFMCount dataset creation pipeline from stem cell differentiation experiments, manual annotation of cell images, to data cleaning, producing high-quality annotated cell locations.

Step 1: Stem Cell Differentiation Experiments

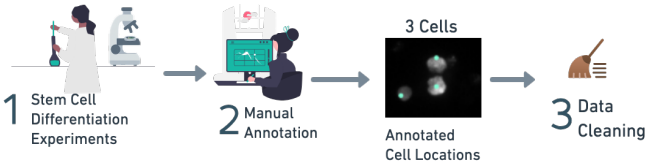

Fig. 2. CellFMCount dataset creation pipeline.

The biological experiments followed a typical process for stem cell differentiation. Once a starting growth factor concentration has been identified, additional concentrations above and below the literature-based determination were used. A control would be incorporated where no growth factor was added. When planning an experiment, duplicate samples are included just in case a mishap occurs and a sample is destroyed (4 concentrations of the growth factor $\times$ 2 samples/condition = 8 samples $\times$ number of antibodies to be screened (5 antibodies) = 40 total samples).

At the end of the seven days, these cell cultures were fixed and immunolabeled with a panel of cell-type-specific antibody markers (see Table II), followed by a fluorescent secondary antibody, such as Cy3 (Cyanine 3) or AF488 (Alexa Fluor 488) conjugated secondary antibody, to visualize the immunolabeled cells. Cells can also be incubated in propidium iodide (PI), a fluorescent cell viability dye, to determine the cell viability of the cells at the conclusion of the experiment. Some experiments used co-labeling with two primary antibodies; however, during imaging, each fluorescent channel displayed only one marker at a time. Since images were not overlaid prior to analysis, markers were quantified separately. All samples were also counterstained with a nuclear stain such as DAPI to facilitate counting the total cell population. All samples were imaged under fluorescence microscopy using a 20x or 40x objective. The above process is also known as immunocytochemistry (ICC).

Ten image regions (fields) were selected systematically, and 2-3 images were captured (antibody immunolabeling and DAPI) for each image field per biological condition. The following counts would be made in each field: *the total number of cells (DAPI-stained nuclei) and the number of cells expressing the primary antibody of interest*, Ki67, TuJ1, MAP2ab, RIP, or GFAP. This data would be used to calculate the percentage of cells immunolabeled with each respective antibody marker. For each antibody marker and each condition, we typically gathered 20 images (10 for DAPI-stained images and 10 for antibody-labeled image fields) for each condition. Students supervised by the biologist authors performed experiments and collected images. They used either an inverted Leica fluorescent microscope (Leica DMI4000B) equipped with standard epiflorescence illumination and a Leica DFC310 FX digital camera or an upright Leica fluorescent microscope (Leica DM5000B) equipped with standard epifluorescence illumination and a Q Imaging Retiga 2000R (Q Imaging) digital camera.

Step 2: Data Annotation

TABLE II
SUMMARY OF NUCLEAR STAINS AND IMMUNOLABELING MARKERS WITH THEIR BIOLOGICAL RELEVANCE.

Marker	Biological Target / Description
DAPI	Universal nuclear stain (labels DNA in all cells)
PI	Propidium Iodide; labels dead or permeable nuclei
RIP	Oligodendrocyte lineage marker
GFAP	Astrocyte marker
TuJ1	Neuronal marker for immature neurons
MAP2ab	Neuronal marker for maturing neurons
Ki67	Proliferating Cells

To perform manual annotation, all images were imported into ImageJ (FIJI distribution) [30]. Graduate and undergraduate students trained to evaluate antibody labeled images performed the initial annotation on their assigned sets of images. The students used ImageJ CellCounter plugin to place a single dot on each nucleus, marking each cell of interest by clicking directly on the image. To ensure high annotation fidelity, every undergraduate annotation set was reviewed and corrected as needed by a graduate student with experience in cell analysis. The final dot coordinates and associated metadata were exported from CellCounter in XML format, preserving annotation structure.

Step 3: Data Cleaning

This process ensured that all annotations were complete, accurate, and machine learning-ready. First, we verified that each XML file exported from the ImageJ CellCounter plugin matched its corresponding microscopy image. Any XML files without a corresponding image or vice versa were removed to maintain data integrity. The duplicate image annotation pairs were then removed to prevent data leakage and biased performance estimates during model development and evaluation. All image and annotation filenames were standardized to numerical identifiers to simplify file referencing and ensure compatibility with automated processing pipelines. A separate metadata CSV file was created to maintain traceability, containing the original filenames, associated antibody or cell staining markers, imaging magnification, and other relevant experimental details for each sample.

The original dot annotations in XML format were converted into machine learning-friendly CSV files. Each CSV file contains the X and Y coordinates of all annotated cells in the corresponding image. This format facilitates straightforward integration into different machine learning pipelines used in counting tasks.

Following this cleaning and standardization process, the final CellFMCount dataset comprises 3,023 fluorescence microscopy images containing over 430,000 manually annotated cell locations.

IV. Dataset Description

A. Staining Modalities and Cell Count Statistics

The CellFMCount dataset includes images stained with three channels: DAPI, Cy3, and AF488. Of the 3,023 images, approximately 45% include DAPI, 47% Cy3, and the rest

AF488 (Table III). DAPI is routinely used together with other markers as a universal nuclear counterstain, which explains its high representation in the dataset and makes it essential for determining the total cell population. DAPI-stained images are therefore denser, while Cy3 and AF488 along with other markers highlight specific structures or subpopulations and produce sparser images. Across all channels, cell counts follow a long-tailed distribution with most images containing relatively few cells.

TABLE III
SUMMARY STATISTICS OF CELL COUNTS ACROSS IMAGING CHANNELS. CPI: #CELLS PER IMAGE.

Stain	#Images	#Cells	Mean CPI± std	Min / Max CPI
DAPI	1,373	386,702	281.6 ± 426.9	4 / 2,546
Cy3	1,428	42,797	30.0 ± 51.2	0 / 383
AF488	222	1,822	8.2 ± 16.6	0 / 128

B. Marker Diversity and Labeling Strategies

CellFMCount includes six immunocytochemistry markers that identify different biological structures or cellular states involved in neural differentiation. Depending on the experimental design, these markers are applied individually or in various combinations. They are used to label nuclei, neuronal structures, glial cells, or indicators of proliferation. This supports a detailed characterization of cellular heterogeneity and phenotypic diversity during the differentiation process.

Table IV shows descriptive statistics of the dataset by antibody marker type. **Ki67 + TuJ1** denotes that both Ki67 and TuJ1 were used (colabeling). Ki67 marks actively proliferating cells, while TuJ1 highlights immature neurons, enabling identification of proliferating neuronal precursors. Similarly, combinations like **MAP2ab + RIP** or **GFAP + Ki67** allow assessment of phenotypically distinct or transitioning cell states.

TABLE IV
CELL COUNT STATISTICS FOR EACH MARKER TYPE. MARKERS WITH LOW AVERAGE COUNTS AND HIGH STANDARD DEVIATION MAY PRESENT DETECTION CHALLENGES. CPI: #CELLS PER IMAGE

Marker	#Images	Mean CPI ± std	Median CPI	Min / Max CPI
DAPI	1,373	281.6 ± 426.9	110	4 / 2,546
PI	552	13.4 ± 36.1	0	0 / 12
Ki67 + TuJ1	295	37.0 ± 59.0	10	0 / 383
MAP2ab	242	20.9 ± 23.03	11	0 / 128
GFAP + Ki67	178	1.3 ± 2.39	0	0 / 12
RIP	120	18.1 ± 13.24	15	2 / 67
TuJ1	119	74.1 ± 49.33	71	1 / 216
Ki67 + RIP	67	129.9 ± 79.45	131	1 / 318
GFAP	57	10.4 ± 19.12	1	0 / 84
MAP2ab + RIP	20	39.2 ± 24.35	35.5	6 / 91

Dataset Challenges. The wide range in image counts and average cell densities across markers introduces significant class imbalance and sparsity. For instance, DAPI-stained images are dense and abundant, while markers like GFAP, MAP2ab, TuJ1, and RIP occur in fewer than 300 images and often exhibit very low cell counts. This imbalance introduces challenges for model generalization, particularly in the context of density map-based cell counting. Markers corresponding to rare cellular populations yield highly sparse density maps, making learning more difficult and prone to overfitting. Fig. 1 shows cell images with varying morphologies across markers. DAPI highlights densely packed nuclei, while MAP2ab reveals elongated structures. These differences introduce spatial complexity that challenges standard feature extraction methods.

We restrict our benchmarking to the DAPI-stained subset, which offers both high cell density and consistent availability across conditions, providing a stable and representative basis for model assessment.

C. DAPI-stained Subset Characterization

The DAPI-stained image subset includes 1,373 grayscale fluorescence microscopy images with 386,702 annotated cell locations. Table IV shows highly variable cell count per image (CPI), with a long-tailed distribution. While the average is 281.6 cells per image (std: 426.9), over 25% of images contain fewer than 38 cells, and some images contain over 2,500 cells. This reflects substantial biological heterogeneity, differences in experimental conditions (e.g., growth conditions), and variation in the field of view.

Table V shows statistics of the DAPI-stained subset by cell types. Two neural progenitor cell types: adult rat hippocampal progenitor cells (AHPC) and murine retinal progenitor cells (RPC), were imaged at 20× and 40× magnification, respectively. Notably, RPC images exhibit higher average cell counts and lower variance in density despite having smaller fields of view.

TABLE V
DESCRIPTIVE STATISTICS FOR DAPI-STAINED AHPC AND RPC SUBSETS. CPI: #CELLS PER IMAGE.

Cell type	#Images	Mean CPI± std	Median	Min / Max CPI
AHPC (20×)	1,214	279.3 ± 451.7	93	4 / 2,546
RPC (40×)	159	299.7 ± 125.6	258	75 / 686

V. MODEL BENCHMARKING

We present a comprehensive evaluation of twelve established methods, including regression-based, density-map, and crowd-counting approaches, as well as a case study adaptation of the Segment Anything Model (SAM) for density-map estimation. For this study, we focused on DAPI-stained images from CellFMCount, as they are the most abundant, display consistent nuclear contrast, and represent a widely used staining protocol in microscopy. These images feature high cell densities, overlapping nuclei, and staining variability, making them a challenging yet representative benchmark. Future work will extend our analysis to other markers (e.g., Cy3, AF488) to study cross-channel generalization and model robustness in multimodal settings.

A. Problem Formulation

Given an input image $\mathbf{x} \in \mathbb{R}^{c \times H \times W}$ with c channels and an image height and width of H by W, the goal is to estimate the number of cells present in $\mathbf{x}$.

a) Regression-based Counting: The objective is to learn a function $f_\theta : \mathbb{R}^{c\times H\times W} \to \mathbb{R}_{\geq 0}$ that maps an input image to a scalar prediction $\hat{y}$ denoting the predicted cell count.

$$\hat{y} = f_\theta(\mathbf{x}) \tag{1}$$

The model parameters θ are optimized by minimizing a task-specific objective $\mathcal{L}_{\text{reg}}$ over the training dataset:

$$\underset{\theta}{\operatorname{argmin}}\ \mathcal{L}_{\text{reg}}(f_\theta(\mathbf{x}), y), \tag{2}$$

where $y \in \mathbb{R}_{\geq 0}$ is the ground truth cell count for image $\mathbf{x}$.

b) Density-based Counting: Alternatively, counting can be formulated as estimating a density map $D_\theta(\mathbf{x}) \in \mathbb{R}^{H_f\times W_f}$ and then summing the density values over the entire estimated density map.

$$\hat{y} = \sum_{h=1}^{H_f}\sum_{w=1}^{W_f} D_\theta(\mathbf{x})_{h,w}, \tag{3}$$

where H_f and W_f are the height and width of the output. The density map ground truth $D(\mathbf{x})$ is typically constructed by placing normalized Gaussian kernels at annotated cell locations. The parameters θ are learned by minimizing a density-specific objective $\mathcal{L}_{\text{dens}}$.

$$\underset{\theta}{\operatorname{argmin}}\ \mathcal{L}_{\text{dens}}(D_\theta(\mathbf{x}), D(\mathbf{x})). \tag{4}$$

Both formulations aim to infer accurate cell counts from microscopy images, with the density-based approach additionally offering localization information.

B. Proposed SAM-Counter Model

SAM-Counter repurposes the pretrained Segment Anything Model (SAM) image encoder [18] as the backbone of a density-map estimation pipeline for cell counting. We remove SAM's prompt encoders and mask decoder, retaining only the Vision Transformer (ViT) image encoder, which is then paired with lightweight convolutional layers to produce density maps. See Fig. 3.

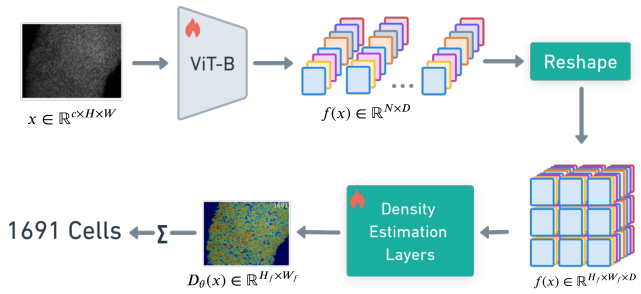

Fig. 3. The architecture of SAM-Counter. The input image is encoded using a fine-tuned ViT encoder from SAM [18]. The density estimation layers then estimate the density map, whose sum yields the cell count. See (5) for reshaping.

SAM's encoder, originally trained on over one billion segmentation masks sourced from diverse natural images, provides spatial representations with rich semantic structure and strong cross-domain generalization. To adapt these features to microscopy data, we reshape the ViT output and feed it to lightweight convolutional layers that produce density maps for cell counting. We then fine-tune the entire encoder–decoder stack on our annotated cell images, enabling the model to capture domain-specific characteristics such as variable cell morphology, staining variability, and imaging noise. The resulting model is highly accurate, leveraging SAM's large-scale pretraining to achieve robust performance even with limited training samples.

Given an input image $\mathbf{x} \in \mathbb{R}^{c\times H\times W}$, the SAM-encoder divides it into non-overlapping patches of size $P \times P$. Each patch is flattened and projected into a D-dimensional embedding, producing $N = \frac{H}{P}\cdot\frac{W}{P}$ tokens. Passing these through a Vision Transformer (ViT) yields contextualized embeddings that we reshape into a spatial feature map:

$$\mathbf{f}(\mathbf{x}) \in \mathbb{R}^{H_f\times W_f\times D}, \quad \text{where } H_f = \tfrac{H}{P},\ W_f = \tfrac{W}{P}. \tag{5}$$

The spatial feature map $\mathbf{f}(\mathbf{x})$ is processed by convolutional density estimation layers. These layers consist of sequential convolutional blocks, implemented as a series of 1×1 convolutional layers with ReLU activations. The final 1×1 convolution reduces the channel dimension to 1, yielding a single-channel density map.

We selected the ViT-Base (ViT-B) architecture as the encoder to balance performance and computational efficiency. Preliminary experiments show that ViT-Large and ViT-Huge offer only marginal improvements in performance [18]. In contrast, these larger models significantly increased memory usage, training time and inference latency.

We fine-tune the SAM encoder jointly with the density estimation head in an end-to-end manner using Mean Squared Error (MSE) loss. This allows the model to adapt to domain-specific features in microscopy data. Freezing the encoder led to poor performance in our experiments (see Section VI-B).

C. Evaluation Metrics

We utilize the following metrics: Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), and Acceptable Count Percentage (ACP) [17]. See Table VI.

Let $y_i \in \mathbb{R}_{\geq 0}$ denote the ground truth cell count for the i-th image, $\hat{y}_i \in \mathbb{R}_{\geq 0}$ the predicted count, and $n \in \mathbb{N}$ the total number of test samples. The indicator function $[\![\cdot]\!]$ returns 1 if the condition inside is true, and 0 otherwise.

TABLE VI
COMMON PERFORMANCE METRICS

$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}\lvert y_i - \hat{y}_i\rvert$	$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$
$\text{MAPE} = \frac{100\%}{n}\sum_{i=1}^{n}\left\lvert\frac{y_i - \hat{y}_i}{y_i}\right\rvert$	$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$
$\text{ACP} = \frac{100\%}{n}\sum_{i=1}^{n}[\![\,\lvert\hat{y}_i - y_i\rvert \leq 0.05\, y_i\,]\!]$	

All the metrics except ACP capture absolute and relative prediction errors. Lower values are desirable. ACP follows the domain experts' desire to accept the result only when the predicted count is within a practically acceptable error tolerance of 5%. Higher ACP values are desirable.

D. Training and Testing Sets

All DAPI images were split into two non-overlapping sets: training (80%) and testing (20%) sets. To ensure representative training and testing sets despite the long-tailed distribution of cell counts, we employed a stratified splitting strategy. First, we discretized the cell count into 5 distinct bins using Jenk's natural breaks optimization [31]. Stratification was then performed based on the combined categories of these cell count bins and the magnifications used, ensuring proportional representation in both the training and test sets. This approach aims to maintain similar cell count and magnification distributions across splits. As shown in Fig. 4, both the training and test sets exhibit similar cell count distributions.

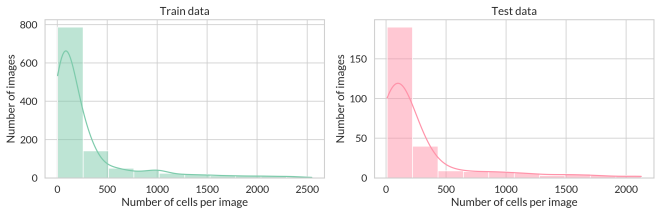

Fig. 4. Distribution of cell counts per DAPI-stained images in training and testing sets.

E. Model Selection

Thirteen distinct models were evaluated for cell counting, including CNN-based regression-based models and density estimation methods, and crowd-counting models. These models were chosen for their strong performance in cell counting and in other domains. The regression-based models employ a backbone feature extractor network followed by a fully connected layer to output a scalar value. Four different pretrained backbones, namely VGG-16 [32], ResNet-50, ResNet-18 [33], and EfficientNet B7 [34] were fine-tuned on our training dataset. The density map estimation models include four cell counting models [16], [25], [26], [28] and three crowd-counting models that were previously adopted for cell-counting tasks [11], [12], [35]. We used the original authors' implementations of the compared cell counting and crowd counting methods with a minor modification to use our dataset. We resized all images to 224x224 for faster training and because the pretrained models were trained on this image size. The ground truth density maps were generated by convolving each dot annotation with a normalized 5×5 Gaussian kernel. This kernel size offers a good balance between smoothing and preserving detail in areas with densely packed cells. No augmentations were used unless the original implementation uses augmentation during training. The training of all the models involved searching for optimal values for batch size and learning rate based on a validation split. The model checkpoints with the lowest validation MAE were saved. Table VII shows the optimal hyperparameter values. A single round of SAM-Counter training on one NVIDIA A100 GPU took 36 hours.

TABLE VII
OPTIMAL HYPERPARAMETER VALUES BASED ON VALIDATION MAE FOR EACH MODEL UNDER STUDY

Model	Batch size	Learning rate
ResNet-18 [33]	8	4.92×10^{-4}
ResNet-50 [33]	16	7.8×10^{-3}
VGG-16 [32]	8	2.57×10^{-4}
EfficientNet B7 [34]	16	8.667×10^{-4}
MCNN [12]	32	3.79×10^{-4}
CSRNet [11]	32	3.31×10^{-3}
MAN [35]	4	$1.0 \times 10 - 5$
SAUNet [26]	75	1.0×10^{-3}
C-FCRN+AUX [28]	100	1.0×10^{-3}
Count-ception [16]	32	5.166×10^{-3}
FCRN-A [25]	16	9.708×10^{-3}
SAM-Counter (Ours)	8	1.0×10^{-6}

VI. RESULTS

Table VIII reports the performance of the thirteen compared models on the test dataset. In all performance tables, the best performance for each method category is underlined. The best overall performances are bold.

TABLE VIII
PERFORMANCE COMPARISON OF DIFFERENT MODELS ON THE TEST DATA

Model	MAE ↓	MSE ↓	RMSE ↓	MAPE ↓	ACP ↑
Regression-based					
ResNet-18 [33]	63.49	16651.17	129.04	30%	13.82%
ResNet-50 [33]	61.73	14058.55	118.57	33%	11.64%
VGG-16 [32]	38.15	7131.87	84.45	17%	24.73%
EfficientNet B7 [34]	41.76	8189.01	90.49	17%	19.27%
Adopted Density Map Estimation from Crowd Counting					
MCNN [12]	42.31	9943.65	99.72	18%	25.09%
CSRNet [11]	27.46	4576.63	67.65	**9%**	38.55%
MAN [35]	27.68	5014.03	70.81	**9%**	42.55%
Density Map Estimation for Cell Counting					
SAUNet [26]	42.02	7866.28	88.69	24%	22.91%
C-FCRN+AUX [28]	37.45	15133.47	123.02	14%	33.82%
Count-ception [16]	34.79	7770.48	88.15	13%	34.91%
FCRN-A [25]	31.96	6616.88	81.34	12%	32.36%
SAM-based Cell Counting					
IDCC-SAM (zero shot) [29]	211.72	217437.74	466.3	54%	2.55%
SAM-Counter (Ours)	**22.12**	**2470.71**	**49.71**	11%	**48.73%**

Across all evaluated categories, the proposed **SAM-Counter** achieves the strongest overall performance. It attains the lowest MAE (22.12), MSE (2470.71), and RMSE (49.71), alongside the highest ACP (48.73%), indicating both accuracy and reliability in count prediction. CSRNet and MAN stand out among the baselines, suggesting robustness to input variability. They both tie for the minimal MAPE (9%) with MAN offering stronger ACP (42.55%).

Regression-based models such as ResNet-18 and ResNet-50 underperform across all metrics, with notably higher MAE and MSE values. In contrast, classical crowd-counting models (e.g., CSRNet, MAN) and domain-specific architectures (e.g.,

Count-ception, FCRN-A) show consistently better alignment with the spatial distribution of cellular features. The poor performance of IDCC-SAM (MAE of 211.72, ACP of 2.55%) illustrates the inadequacy of directly applying general-purpose segmentation models to cell counting without task-specific tuning.

TABLE IX
MAE ACROSS DIFFERENT CELL TYPES: AHPC AND RPC. LOWER VALUES INDICATE BETTER PERFORMANCE.

Model	AHPC ↓	RPC ↓
Regression-based		
VGG-16 [32]	39.23	29.95
ResNet-50 [33]	62.95	52.40
ResNet-18 [33]	65.51	48.15
EfficientNet B7 [34]	43.04	31.97
Adopted Density Map Estimation from Crowd Counting		
MCNN [12]	43.38	34.17
CSRNet [11]	28.16	22.15
MAN [35]	28.84	**18.88**
Density Map Estimation for Cell Counting		
SAU-Net [26]	43.10	33.83
C-FCRN+AUX [28]	38.76	27.50
Count-ception [16]	36.32	23.20
FCRN-A [25]	32.06	31.21
SAM-based Cell Counting		
IDCC-SAM (zero shot) [29]	228.96	80.78
SAM-Counter (Ours)	**22.32**	20.61

Table IX shows that SAM-Counter achieves the lowest error on AHPC images (22.32), demonstrating robust performance on data with high variability and sparsity. On RPC images, MAN slightly outperforms SAM-Counter with an MAE of 18.88 versus 20.61, suggesting that its architecture may be better suited for higher magnification images containing larger cells.

Among the baselines, CSRNet delivers competitive performance across both cell types, particularly on RPC (22.15), highlighting the general effectiveness of density map-based approaches. Count-ception and C-FCRN+AUX also exhibit notably lower error on RPC compared to AHPC, likely due to the reduced variability and tighter count distribution in the higher-magnification subset.

By contrast, regression-based models perform consistently worse on both conditions, with especially high errors on AHPC images, underscoring their limitations in handling spatial heterogeneity and large-scale count variation.

Overall, these results show that both model design choices and the nature of the dataset significantly impact performance in biological tasks.

Performance across varying numbers of cells per image: We evaluate model performance over three cell-count ranges: *low* (0–250 cells, 198 images), *medium* (251–500 cells, 34 images), and *high* (more than 500 cells, 43 images), based on ground-truth annotations.These ranges were defined in collaboration with domain biologists to reflect common variations in microscopy data. Table X reports macro MAE (↓) and ACP (↑) for each range.

SAM-Counter achieves the lowest MAE in the medium (21.41) and high (82.33) ranges and the highest ACP across all densities (47.47%, 58.82%, 46.51%). Improvements are particularly clear in the high-density setting, where it reduces MAE by over 30% compared to the next-best model (CSRNet, 119.99) and nearly doubles ACP (46.51% vs. 27.91%). For low-density images, SAM-Counter remains competitive (MAE 9.17) but is slightly outperformed by MAN (7.31) and CSRNet (7.70).

TABLE X
MACRO MAE (↓) AND ACP (↑) FOR DIFFERENT CELL DENSITY LEVELS. LOWER MAE AND HIGHER ACP INDICATE BETTER PERFORMANCE. UNDERLINED VALUES ARE BEST WITHIN THEIR RESPECTIVE CATEGORIES; BOLD VALUES ARE BEST OVERALL.

Model	Low		Medium		High	
	MAE ↓	ACP ↑	MAE ↓	ACP ↑	MAE ↓	ACP ↑
Regression-based						
VGG-16 [32]	13.70	22.22	45.48	32.35	144.92	30.23
ResNet-50 [33]	25.92	11.11	78.93	8.82	212.97	16.28
ResNet-18 [33]	22.77	13.13	56.11	23.53	256.80	9.30
EfficientNet B7 [34]	14.45	20.20	41.31	14.71	167.82	18.60
Adopted Density Map Estimation from Crowd Counting						
MCNN [12]	15.75	23.74	37.01	23.53	168.78	32.56
CSRNet [11]	7.70	38.38	25.54	52.94	119.99	27.91
MAN [35]	**7.31**	43.43	23.82	47.06	124.53	34.88
Density Map Estimation for Cell Counting						
SAU-Net [26]	17.62	22.22	41.56	29.41	154.75	20.93
C-FCRN+AUX [28]	10.29	30.81	28.90	44.12	169.25	39.53
Count-ception [16]	11.47	36.87	31.84	35.29	144.49	25.58
FCRN-A [25]	10.50	32.32	34.51	29.41	128.78	34.88
SAM for Cell Counting						
IDCC-SAM (zero shot) [29]	45.10	2.02	152.47	5.88	1025.79	2.33
SAM-Counter (Ours)	9.17	**47.47**	**21.41**	**58.82**	**82.33**	**46.51**

MAN achieves the best MAE in sparse images and a strong ACP (43.43%), while CSRNet provides the best non-SAM high-density MAE (119.99) and the highest ACP at medium density (52.94%). Among cell-specific DME models, C-FCRN+AUX achieves the lowest MAE for low and medium densities (10.29, 28.90) and the highest ACP within its group at high density (39.53%).

Within model families, VGG-16 attains the best MAE for low (13.70) and high (144.92) densities, whereas EfficientNet-B7 is strongest at medium counts (41.31). MAN leads at low and medium densities among crowd-counting models, while CSRNet performs better at high density. FCRN-A is strongest for dense images (128.78), whereas C-FCRN+AUX and Count-ception perform better for sparse settings.

A. Qualitative Analysis

Fig. 5 shows qualitative results for the best-performing models selected based on the lowest Mean Absolute Error (MAE) from Table VIII: CSRNet from crowd counting methods, FCRN-A from cell counting methods, and SAM-Counter. These models were evaluated on diverse test scenarios, ranging from sparse to highly congested scenes.

The top row presents the input images, while the second row displays the ground truth density maps with the total number of cells indicated. Subsequent rows illustrate predicted density maps by CSRNet, FCRN-A, and SAM-Counter, respectively. Heatmap colors indicate density intensity, with red representing high density and blue indicating lower density regions.

Fig. 5(1) illustrates a sparse scenario, where all three models show strong visual consistency with ground truth, accurately capturing individual cell locations. CSRNet and SAM-Counter precisely match the GT count, while FCRN-A slightly underestimates. In Fig. 5(2), a highly dense and challenging scenario with clustered cells is presented. All models produce smoother density maps, with CSRNet and FCRN-A significantly underestimating the count, whereas SAM-Counter demonstrates improved localization, though it slightly underestimates the total count.

Fig. 5(3) shows moderately dense and variably sized cells. CSRNet and FCRN-A yield irregular density predictions, while SAM-Counter provides a uniform density distribution closely aligning with GT. Fig. 5(4) represents a scenario with fewer cells and moderate density variations, where SAM-Counter closely matches the ground truth count and provides accurate cell localization, whereas CSRNet and FCRN-A slightly underestimate and produce more diffused density predictions. Fig. 5(5) depicts another sparse distribution scenario with varied cell sizes. SAM-Counter maintains consistent accuracy, closely mirroring GT distributions despite a slight overestimation. CSRNet and FCRN-A again exhibit challenges in precise localization and overrepresent overall density.

Overall, SAM-Counter tends to produce well-defined and visually consistent density maps across diverse scenarios, though all three models exhibit specific strengths and challenges depending on scenario complexity and cell distribution characteristics.

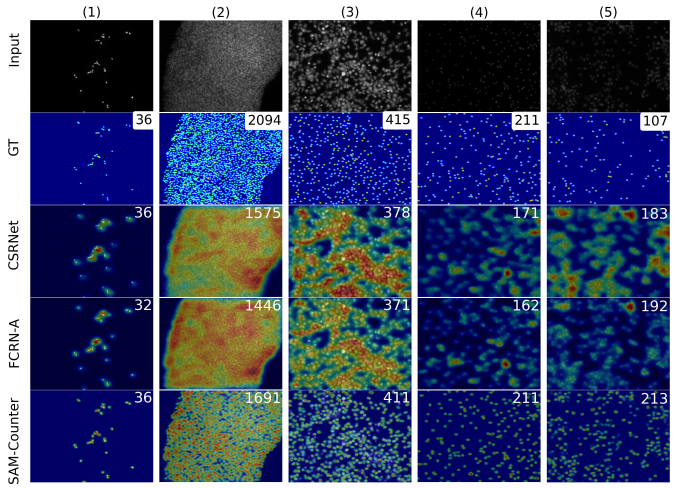

Fig. 5. Visualization of density map predictions of the three best-performing models on five representative test samples: one per column.

B. Ablation Results

To evaluate the impact of design choices when adapting SAM, we investigated two configurations. One variable is whether to freeze or fine-tune the SAM encoder. The other is the number of layers in the density estimation head. Table XI reports the test MAE for each configuration.

Table XI shows the impact of encoder fine-tuning and head depth. Fine-tuning the SAM encoder reduces test MAE from 44.24 (frozen) to 22.12. This confirms that adapting the encoder to microscopy images improves performance.

TABLE XI
ABLATION RESULTS ON SAM-COUNTER: ENCODER FREEZING AND DENSITY-HEAD DEPTH (TEST MAE).

Experiment	Setting	# Params	Test MAE
Encoder	Frozen	41.2 K	44.24
	Trainable	89.7 M + 41.2 K	22.12
Number of layers	1 layer	89.7 M + 257	22.99
	2 layers	89.7 M + 33K	24.58
	3 layers	89.7 M + 41.2 K	22.12

The number of layers in the density estimation head also affects accuracy. A three-layer head performs best, with an MAE of 22.12. One and two layers result in slightly higher errors. The head remains lightweight in all cases. Even the largest version adds only about 41K parameters, compared to 89.7M in the encoder.

We evaluated whether SAM-Counter trained on DAPI-stained images could generalize to other staining types. The model was trained only on DAPI and tested on images stained with Cy3 and AF488, which correspond to different antibody markers. On these test images, SAM-Counter achieved MAE of 606.26. This performance drop supports our earlier observation: images stained with different markers show substantial visual differences (see Fig. 1), which limit generalization across different fluorescence markers.

C. Limitations and Broader Impacts

While CellFMCount presents a diverse benchmark with multiple cell types, conditions, and magnifications, it does not cover all imaging modalities or biological settings. SAM-Counter is based on SAM [18], which requires 1024×1024 input images. Future work could explore parameter-efficient fine-tuning (e.g., LoRA) or model distillation to reduce resource requirements without sacrificing accuracy.

Our dataset and method aim to support research in neuroregeneration, cancer, and stem cell therapy by improving accuracy and consistency in cell counting. The data come from experiments on murine and rat cells. We demonstrate the potential of adapting foundation models for biomedical tasks and encourage responsible use and validation in downstream applications.

VII. CONCLUSION

We introduced CellFMCount, a large-scale, diverse fluorescence microscopy dataset for robust cell counting under real-world variability. We showed how foundation models like SAM can be repurposed for density map estimation via our SAM-Counter, and benchmarked thirteen state-of-the-art methods on DAPI-stained images. Our results establish strong baselines and highlight both the challenges and opportunities in automated cell counting.

REFERENCES

[1] Antony Orth, Diane Schaak, and Ethan Schonbrun. Microscopy, Meet Big Data. *Cell Systems*, 4(3):260–261, March 2017.

[2] Martin S Blumenreich. The white blood cell and differential count. *Clinical Methods: The History, Physical, and Laboratory Examinations. 3rd edition*, 1990.

[3] Jarno Drost and Hans Clevers. Organoids in cancer research. *Nature Reviews Cancer*, 18(7):407–418, 2018.

[4] Suprem R. Das, Metin Uz, Shaowei Ding, Matthew T. Lentner, John A. Hondred, Allison A. Cargill, Donald S. Sakaguchi, Surya Mallapragada, and Jonathan C. Claussen. Electrical Differentiation of Mesenchymal Stem Cells into Schwann-Cell-Like Phenotypes Using Inkjet-Printed Graphene Circuits. *Advanced Healthcare Materials*, 6(7):1601087, 2017.

[5] Mei-Yin C. Polley, Samuel C. Y. Leung, Lisa M. McShane, Dongxia Gao, Judith C. Hugh, Mauro G. Mastropasqua, Giuseppe Viale, Lila A. Zabaglo, Frédérique Penault-Llorca, John M. S. Bartlett, Allen M. Gown, W. Fraser Symmans, Tammy Piper, Erika Mehl, Rebecca A. Enos, Daniel F. Hayes, Mitch Dowsett, Torsten O. Nielsen, and International Ki67 in Breast Cancer Working Group of the Breast International Group and North American Breast Cancer Group. An international Ki67 reproducibility study. *Journal of the National Cancer Institute*, 105(24):1897–1906, December 2013.

[6] Gerburg Keilhoff, Alexander Goihl, Felix Stang, Gerald Wolf, and Hisham Fansa. Peripheral nerve tissue engineering: Autologous schwann cells vs. transdifferentiated mesenchymal stem cells. *Tissue Engineering*, 12(6):1451–1465, 2006. PMID: 16846343.

[7] Maria Brohlin, Daljeet Mahay, Lev N. Novikov, Giorgio Terenghi, Mikael Wiberg, Susan G. Shawcross, and Liudmila N. Novikova. Characterisation of human mesenchymal stem cells following differentiation into schwann cell-like cells. *Neuroscience Research*, 64(1):41–49, 2009.

[8] Mari Dezawa, Izumi Takahashi, Michiyo Esaki, Masahiko Takano, and Hajime Sawada. Sciatic nerve regeneration in rats induced by transplantation of in vitro differentiated bone-marrow stromal cells. *European Journal of Neuroscience*, 14(11):1771–1776, 2001.

[9] A Ladak, J Olson, EE Tredget, and T Gordon. Differentiation of mesenchymal stem cells to support peripheral nerve regeneration in a rat model. *Experimental neurology*, 228(2):242–252, 2011.

[10] Anup D Sharma, Svitlana Zbarska, Emma M Petersen, Mustafa E Marti, Surya K Mallapragada, and Donald S Sakaguchi. Oriented growth and transdifferentiation of mesenchymal stem cells towards a schwann cell fate on micropatterned substrates. *Journal of bioscience and bioengineering*, 121(3):325–335, 2016.

[11] Yuhong Li, Xiaofan Zhang, and Deming Chen. CSRNet: Dilated Convolutional Neural Networks for Understanding the Highly Congested Scenes. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1091–1100, June 2018. ISSN: 2575-7075.

[12] Yingying Zhang, Desen Zhou, Siqin Chen, Shenghua Gao, and Yi Ma. Single-Image Crowd Counting via Multi-Column Convolutional Neural Network. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 589–597, June 2016. ISSN: 1063-6919.

[13] Victor Lempitsky and Andrew Zisserman. Learning To Count Objects in Images. In *Advances in Neural Information Processing Systems*, volume 23. Curran Associates, Inc., 2010.

[14] Philipp Kainz, Martin Urschler, Samuel Schulter, Paul Wohlhart, and Vincent Lepetit. You Should Use Regression to Detect Cells. In Nassir Navab, Joachim Hornegger, William M. Wells, and Alejandro F. Frangi, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, Lecture Notes in Computer Science, pages 276–283, Cham, 2015. Springer International Publishing.

[15] Mark Marsden, Kevin McGuinness, Suzanne Little, Ciara E. Keogh, and Noel E. O'Connor. People, Penguins and Petri Dishes: Adapting Object Counting Models to New Visual Domains and Object Types Without Forgetting. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 8070–8079, 2018.

[16] Joseph Paul Cohen, Genevieve Boucher, Craig A. Glastonbury, Henry Z. Lo, and Yoshua Bengio. Count-ception: Counting by Fully Convolutional Redundant Counting. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pages 18–26, 2017.

[17] Abdurahman Ali Mohammed, Catherine Fonder, Donald S. Sakaguchi, Wallapak Tavanapong, Surya K. Mallapragada, and Azeez Idris. IDCIA: Immunocytochemistry Dataset for Cellular Image Analysis. In *Proceedings of the 14th Conference on ACM Multimedia Systems*, pages 451–457, Vancouver BC Canada, June 2023. ACM.

[18] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C. Berg, Wan-Yen Lo, Piotr Dollár, and Ross Girshick. Segment anything. *arXiv:2304.02643*, 2023.

[19] Theodore J Kataras, Tyler J Jang, Jeffrey Koury, Hina Singh, Dominic Fok, and Marcus Kaul. Acct is a fast and accessible automatic cell counting tool using machine learning for 2d image segmentation. *Scientific Reports*, 13(1):8213, 2023.

[20] Evangelos Pistolas, Eleni Kyratzopoulou, Lamprini Malletzidou, Evangelos Nerantzis, Chairi Kiourt, and Nikolaos Kazakis. Cidacc: Chlorella vulgaris image dataset for automated cell counting. *Data in Brief*, 57:110941, 2024.

[21] Mauro Giuseppe Camporeale, Giovanni Dimauro, Matteo Gelardi, Giorgia Iacobellis, Mattia Sebastiano Ladisa, Sergio Latrofa, and Nunzia Lomonte. Nasal mucosa cell dataset (nmcd), 2024.

[22] Felix Buggenthin, Carsten Marr, Michael Schwarzfischer, Philipp S Hoppe, Oliver Hilsenbeck, Timm Schroeder, and Fabian J Theis. An automatic method for robust and fast cell detection in bright field images from high-throughput microscopy. *BMC bioinformatics*, 14:1–12, 2013.

[23] David R Stirling, Madison J Swain-Bowden, Alice M Lucas, Anne E Carpenter, Beth A Cimini, and Allen Goodman. Cellprofiler 4: improvements in speed, utility and usability. *BMC bioinformatics*, 22:1–11, 2021.

[24] Carlos Arteta, Victor Lempitsky, J. Alison Noble, and Andrew Zisserman. Detecting overlapping instances in microscopy images using extremal region trees. *Medical Image Analysis*, 27:3–16, 2016. Discrete Graphical Models in Biomedical Image Analysis.

[25] Weidi Xie, J. Alison Noble, and Andrew Zisserman. Microscopy cell counting and detection with fully convolutional regression networks. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, 6(3):283–292, May 2018.

[26] Yue Guo, Jason Stein, Guorong Wu, and Ashok Krishnamurthy. SAU-Net: A Universal Deep Network for Cell Counting. In *Proceedings of the 10th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics*, BCB '19, pages 299–306, New York, NY, USA, September 2019. Association for Computing Machinery.

[27] Yao Xue, Nilanjan Ray, Judith Hugh, and Gilbert Bigras. Cell Counting by Regression Using Convolutional Neural Network. In Gang Hua and Hervé Jégou, editors, *Computer Vision – ECCV 2016 Workshops*, Lecture Notes in Computer Science, pages 274–290, Cham, 2016. Springer International Publishing.

[28] Shenghua He, Kyaw Thu Minn, Lilianna Solnica-Krezel, Mark A. Anastasio, and Hua Li. Deeply-supervised density regression for automatic cell counting in microscopy images. *Medical Image Analysis*, 68:101892, February 2021.

[29] Samuel Fanijo, Ali Jannesari, and Julie Dickerson. Idcc-sam: A zero-shot approach for cell counting in immunocytochemistry dataset using the segment anything model. *Bioengineering*, 12(2):184, 2025.

[30] Caroline A. Schneider, Wayne S. Rasband, and Kevin W. Eliceiri. NIH Image to ImageJ: 25 years of image analysis. *Nature Methods*, 9(7):671–675, July 2012.

[31] JENKS G. F. The data model concept in statistical mapping. *International Yearbook of Cartography*, 7:186–190, 1967.

[32] Shuying Liu and Weihong Deng. Very deep convolutional neural network based image classification using small training sample size. In *2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR)*, pages 730–734, 2015.

[33] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2016-December:770–778, 12 2015.

[34] Mingxing Tan and Quoc Le. EfficientNet: Rethinking model scaling for convolutional neural networks. In Kamalika Chaudhuri and Ruslan Salakhutdinov, editors, *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 6105–6114. PMLR, 09–15 Jun 2019.

[35] Hui Lin, Zhiheng Ma, Rongrong Ji, Yaowei Wang, and Xiaopeng Hong. Boosting crowd counting via multifaceted attention. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 19628–19637, 2022.