

## Review Article

# Deep Learning in Cell Image Analysis

Junde Xu <sup>1,2</sup>, Donghao Zhou,<sup>1,2</sup> Danruo Deng,<sup>3</sup> Jingpeng Li,<sup>3</sup> Cheng Chen,<sup>3</sup>  
Xiangyun Liao,<sup>2</sup> Guangyong Chen <sup>3,4,5</sup> and Pheng Ann Heng<sup>3</sup>

<sup>1</sup>University of Chinese Academy of Sciences, China

<sup>2</sup>Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, China

<sup>3</sup>The Chinese University of Hong Kong, China

<sup>4</sup>Zhejiang Lab, China

<sup>5</sup>Zhejiang University, China

Correspondence should be addressed to Junde Xu; [jd.xu@siat.ac.cn](mailto:jd.xu@siat.ac.cn), Donghao Zhou; [dh.zhou@siat.ac.cn](mailto:dh.zhou@siat.ac.cn), and Guangyong Chen; [gychen@zhejianglab.com](mailto:gychen@zhejianglab.com)

Junde Xu and Donghao Zhou contributed equally to this work.

Received 24 June 2022; Accepted 8 August 2022; Published 7 September 2022

Copyright © 2022 Junde Xu et al. Exclusive Licensee Zhejiang Lab, China. Distributed under a Creative Commons Attribution License (CC BY 4.0).

Cell images, which have been widely used in biomedical research and drug discovery, contain a great deal of valuable information that encodes how cells respond to external stimuli and intentional perturbations. Meanwhile, to discover rarer phenotypes, cell imaging is frequently performed in a high-content manner. Consequently, the manual interpretation of cell images becomes extremely inefficient. Fortunately, with the advancement of deep-learning technologies, an increasing number of deep learning-based algorithms have been developed to automate and streamline this process. In this study, we present an in-depth survey of the three most critical tasks in cell image analysis: segmentation, tracking, and classification. Despite the impressive score, the challenge still remains: most of the algorithms only verify the performance in their customized settings, causing a performance gap between academic research and practical application. Thus, we also review more advanced machine learning technologies, aiming to make deep learning-based methods more useful and eventually promote the application of deep-learning algorithms.

## 1. Introduction

Cell image analysis plays an important role in biomedical research; it has become the main strategy for projecting how a life system interacts with environmental changes. For example, drug discovery is a crucial process for synthesizing and screening potential candidate medications, which incurs a high cost to study the responses of intact cells or entire organisms to specific chemical substances. Phenotypic screening has been demonstrated to be a superior strategy for small-molecule and first-in-class medicines based on phenotypic analysis of responses [1, 2]. Generally, to analyze the biological changes in cells, a phenotypic screening is performed in a high-content screening (HCS) method, where cells are stained with multichannel fluorescent probes and cultured in plates with multiple isolated wells subjected to different treatments [3–5]. Before entering clinical trials,

molecules must be validated in vitro [6], which unfortunately has a high attrition rate. Moreover, previous research [7–9] indicates that using HCS to select small molecules only has an estimated hit rate of 0–0.01%, which is highly dependent on the professional knowledge of different biologists and quality of the screening compound pool. Given these laborious and erratic procedures, the increasing demand for cell image analysis is crucial for accelerating and improving phenotypic screening.

The goal of cell image analysis is to analyze the phenotypic effects of various treatments and to reveal the relationships between them. The most widely studied tasks of cell image analysis include segmentation, tracking, and classification [4–10]. These tasks have drawn extensive attention from both academia and industry. Recently, a bioimage challenge called the Cell Tracking Challenge [11] for live cell segmentation and tracking was held under the auspices of the

IEEE International Symposium on Biomedical Imaging (ISBI). This challenge maintained a benchmark of 20 different treatments, including various imaging methods and cell types. New-generation biotechnology companies, such as Insitro and Recursion, have been using machine learning techniques for large-scale cell image analysis to promote drug discovery. Notably, Recursion also released a series of open-source datasets called RxRx, aiming to extract phenotypic features by classifying the treatments imposed on cells based solely on raw image inputs. To make the quantitative and statistical analyses of cell images automated and high throughput, many software packages are available, such as ImageJ [12], CellProfiler [13], Icy [14], and CellCognition [15], which typically contain plentiful plugins that allow biologists to design a customized pipeline to perform different tasks.

Deep learning, the most extensively used emerging machine-learning technique, has achieved remarkable success in computer vision and natural language processing [16–20]. In deep learning, a deep neural network (DNN) is trained as an end-to-end model to directly infer the desired labels from input data [21]. In contrast to traditional computer vision techniques, a DNN can automatically produce more effective representations than handcrafted representations by learning from a large-scale dataset. In cell images, deep learning-based methods also show promising results in cell segmentation [22–24] and tracking [25–27]. Such successful applications demonstrate the ability of DNNs to extract high-level features and shed light on the potential capability of using deep learning to reveal more sophisticated life laws behind cellular phenotypes [28]. In addition, a vital breakthrough in computer vision called representation learning [29–34] also provides confidence that phenotypic features can be learned end-to-end by DNNs more efficiently.

Deep learning has shown a powerful ability to extract useful information from raw inputs; however, it is highly influenced by the quality of the dataset. As shown in Figure 1, a typical deep-learning method consists of two modules: an *inference module* and a *retraining module*. When the test environment changes, if the inference module does not achieve satisfactory generalization performance, the DNN must be retrained to adapt to the new data domain using extra annotations. However, the annotations of cell images are considerably expensive than those of natural-scene images because they require expert knowledge to assign labels, and cell images themselves are difficult to collect. Moreover, rare cases are more valuable in cell image analysis than in natural-scene images. Thus, these factors further amplify the shortage of data hunger in deep-learning methods, limiting their practical application. However, because annotation is the most onerous workload, some promising machine-learning technologies, such as active learning, transfer learning, and noisy learning, have been proposed to address this problem. These technologies aim to train a more robust and generalizable DNN with minimal supervision.

In this study, we provide a comprehensive survey of the current progress of three critical tasks. In addition, we will

discuss the challenges of applying machine-learning algorithms to cell image analysis. In contrast to previous survey studies [28, 35], we provide a more technical perspective on deep learning in cell image analysis.

## 2. The Current Progress of Computational Methods in Cellular Image Processing

In this section, we discuss the current progress in applying deep learning to three crucial tasks in cell image analysis: segmentation, tracking, and classification.

**2.1. Segmentation.** Owing to its high homology with traditional computer vision tasks, cell segmentation is a popular topic in the computer vision community. Traditional cell semantic segmentation methods are based on image processing techniques such as level set [36], watershed [37], graph cut [38], optimization using intensity features [39], and physics-based restoration [40–42]. These methods lack flexibility and automation; thus, users regularly need to adjust their parameters to handle different images.

In contrast to cell semantic segmentation, cell instance segmentation requires not only discriminating which pixel belongs to the class of interest (i.e., the cell) but also distinguishing each individual cell. The existing popular methods can be divided into anchor- and region-based methods. Anchor-based methods are originally built for general purpose, such as Faster RCNN [43], Mask RCNN [44], and RetinaNet [45], which are successfully applied to cellular images [46–48]. In contrast, region-based methods adopt a deep convolutional neural network (CNN) as the feature extractor, such as VGG [49] and ResNet [50], and classify each predefined anchor across the input images. Using the anchor mechanism, each pixel can be assigned to multiple anchors, allowing multiple and even overlapped cells to be detected simultaneously. After the network predicts the probability of a cell being present in each anchor, a post-processing step known as non-maximum suppression (NMS) [51] is used to select the top-scoring anchors as the final result. Recently, region-based methods have gained increasing attention because of their simple and end-to-end training schemes. Region-based methods transform each pixel into the desired representation, from which subsequent algorithms can recover individual cells. A basic representation is a two-channel feature map that includes a cell probability channel and a cell boundary channel [27], where the cell probability channel indicates whether the pixel belongs to the cell and the cell boundary channel is used to divide the cell instance. This feature map is then fed into a post-processing algorithm to produce the final result. Song et al. [52] proposed a shape-based method to calculate the shape prior to refining the masks of the touching cells. Alternatively, some methods changed the binary cell probability map into a distance map and transferred the cell segmentation problem to a regression task. Kainz et al. [53] defined the output with regard to the Euclidean distance between the pixels and its closest annotated cell center but neglected the information about the cell boundaries. Bai and Urtasun [54] proposed a method called deepwatershed that adopted

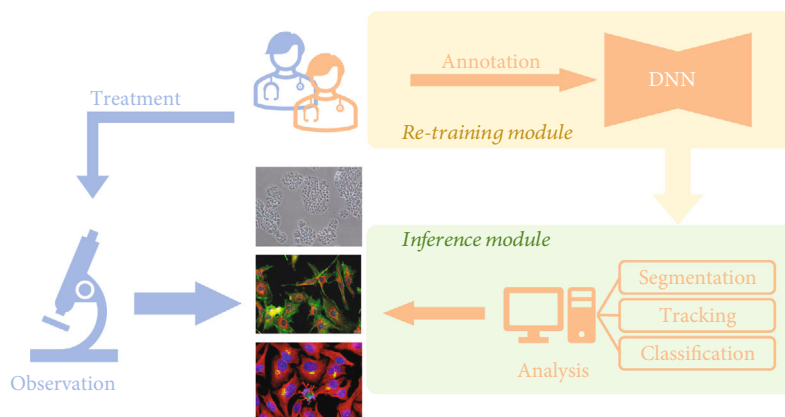


FIGURE 1: Overview of deep learning-based cell image analysis. A typical analysis pipeline consists of a retraining module and an inference module: the inference module directly produces estimated metrics. However, as the experimental setup changes, the parameters of the deep-learning model must be retrained.

a center-to-boundary distance map for instance segmentation, followed by a watershed algorithm [55] to produce final masks, and was successfully applied to cellular images [56–58]. Instead of predicting the distance of every pixel inside the cells, Schmidt et al. [59] used a star-convex polygon to represent a cell instance, which only required predicting the distance between the center and boundary at several particular angles. However, a star-convex polygon was not sufficient to represent nonconvex cells. Recently, a vector-field label was proposed for tracking objects [60] and adopted in instance segmentation [61]. In contrast to the distance map, Stringer et al. [62] proposed a vector-field label called Cellpose, which only focused on the local gradient of each pixel. A conspicuous property of Cellpose is that each vector points only to the neighboring pixel that is closer to the center and in the cell region. Thus, the resultant representations can easily model extremely nonconvex cells.

In summary, anchor- and region-based methods have their own advantages. First, anchor-based methods (Figure 2(a)) are generally more accurate than region-based methods (Figure 2(b)) if the Intersection over Union (IoU) threshold is not excessively high. Relative to region-based methods, anchor-based methods focus more on the wholeness of objects, whereas region-based methods focus on details such as accurate boundaries. Second, region-based methods require fewer computing resources than anchor-based methods do. This is because learning features with more “objectivity” often require a more complicated design and deeper neural networks. Third, region-based methods require significantly more burdensome post-processing than anchor-based methods. Post-processing not only reduces the overall efficiency but also introduces additional hyperparameters. In practice, most biology laboratories have limited computing resources. More accurate and faster algorithms with lower computational consumption remain to be explored.

Deep learning-based methods can achieve remarkable performance in cell segmentation after training with a large-scale and carefully annotated dataset. However, cell images can vary with different treatments, such as different cell types, stains, or even carbon dioxide concentrations.

Moreover, it is very expensive to collect a carefully annotated dataset because the annotation of cell images requires expert knowledge. These barriers cause many researchers to test their methods only on limited and imperfect data, which cannot yield promising results in practice. To this end, an increasing number of large-scale datasets covering multiple experimental environments have been proposed to reflect the practical performance of different algorithms. For example, in cell nucleus segmentation, the data science bowl challenge amassed a dataset of various images of nuclei with up to 30 different treatments or image types [63]. For whole-cell segmentation, Cellpose [62] proposed a generalized dataset containing 10 different treatments (the value is approximated because the Cellpose dataset contains images collected from Google) and up to 608 images, with 69 images held out for testing. In particular, this dataset also included 184 non-cell images shared with repeating convex patterns for better generalization. For more specific applications, Greenwald et al. built a large-scale tissue fluorescence dataset called TissueNet across six imaging platforms and nine organs, which also covered different disease states and species [24]. Edlund et al. manually annotated a large-scale dataset in live-cell imaging with label-free and phase-contrast images of 2D cell culture, named LIVECell, which consisted of more than 1.6 million annotated cells of eight morphologically distinct cell types, grown from early seeding to full confluence [64]. The reporting performance of such large-scale datasets is more convincing for researchers.

**2.2. Tracking.** Tracking is another fundamental task in cell image analysis. Monitoring cell behaviors throughout the lineage can provide useful information for drug discovery, including quantification of signaling dynamics [65], efforts to understand cell motility [66], and attempts to unravel the laws of bacterial cell growth [67]. Such an analysis must associate each cell entity over time. However, simultaneously tracking thousands of cells is challenging. First, in contrast to general object tracking, the phenotype of the cell is no longer a reliable feature to discriminate against cells because each cell shares a similar appearance in this task. Second, during cell culture, some cells, such as stem cells, can

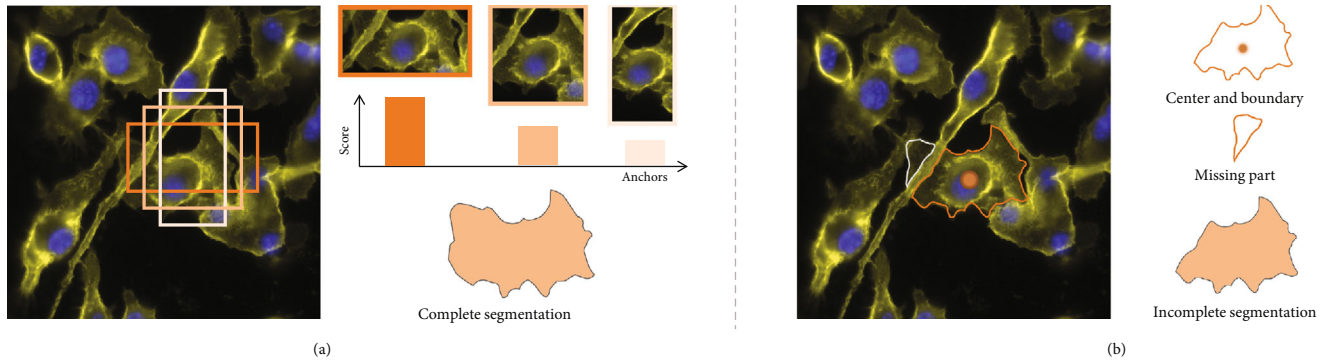


FIGURE 2: Comparison of anchor- and region-based methods. (a) Anchor-based methods segment cells from predefined anchors. (b) Region-based methods segment cells from auxiliary labels.

undergo serious deformation between frames. Third, owing to the phototoxicity and photobleaching during imaging, the frame rate of imaging is often limited.

Before the wide adoption of deep learning-based methods, cell tracking was associated with probabilistic models [68, 69] and active contour models [70–72]. These methods globally optimize a graph or probability map, where each connection between cells indicates a carefully designed energy function of a particular event, such as normal connection, mitosis, and move-in/out of the field of view. Most deep learning-based methods continue to be tracking-by-detection schemes [47, 73–75] and do not take advantage of the rich information of spatial-temporal cues. For example, the historical dynamics of cells can predict the location of each cell in the current frame, and the dynamics of cells can be accumulated by accurate cell detection. A few methods have been proposed for joint learning, cell detection/segmentation, and tracking. Payer et al. proposed a recurrent stacked hourglass network (ConvGRU) that jointly optimized the network with both segmentation labels and tracking information [76], where the network was forced to provide a similar embedding for linked cell pairs. However, this method could only handle cell images with high magnification (for more detailed features) and a high frame rate (for reducing phenotype change over time), which limited its application. Zhou et al. used two variants of U-Net to jointly perform segmentation and tracking [77]. However, they only leveraged multiframe input for better segmentation results and used heuristic functions (such as IoU) for the final tracking results. Thus, the tracking context was not implicitly involved in the network training. Hayashida et al. proposed a vector field map called MPM to simultaneously encode the location and motion of the cells [78]. Similar to another tracking algorithm for general objects [79], the MPM treats objects/cells as points. The MPM adopted two successive images as input and produced a number of shifted vectors, where the norm of vectors indicated cell centers and the direction of vectors indicated the formal location of cells. However, MPM has only been tested on a small fraction of tracks over a large-scale dataset owing to the lack of annotations. Thus, the practical performance of the MPM still must be proven.

**2.3. Classification.** Classification often serves as a downstream analysis task for phenotypic screening and cell profiling. After individual cells are located, each cell is conducted into a high-dimensional feature vector that contains various types of phenotypic information. Typical applications include classifying different gene mutations [80–82] and mechanism of action (MoA) [83, 84]. Before the popularity of deep learning, a classic workflow of processing the feature vector included quality control (to remove outlier samples), preprocessing (normalization, batch-effect correction, etc.), dimensionality reduction (using data-analysis strategies to select useful features while eliminating unnecessary or redundant features), and finally, a classification algorithm. The selection of a classifier depends on interpreting (clustering) or validating (classification) phenotypic features. Classic methods include hierarchical clustering [85, 86], nearest neighbors [87], Bayesian matrix factorization, neural networks, and random forests [88]. These methods rely heavily on the quality of the presteps.

With the impressive capability of CNN to extract more abstract features from images, researchers have started using CNN to analyze cell images end-to-end to substitute the onerous pipeline. However, cell images are often collected in a high-content scheme; thus, each image can have up to hundreds of cells with different phenotypes (specifically, outliers). To infer the correct type from both positive and negative cells, Kraus et al. used multiple instance learning (MIL) to train an integral CNN to jointly segment and predict the label of cell images [89]. This algorithm could generate class-specific segmentation masks, which proved its capability of filtering out outliers. Godinez et al. [90] adopted a multiscale CNN architecture to further eliminate the segmentation step and classify phenotypes into cohesive ones, which further reduced the annotation effort. They tested their algorithms on a real-world dataset, which showed a greater capability to distinguish phenotypes than the conventional pipeline that used handcrafted features. However, CNN classifiers often must be trained under the supervision of meaningful labels to learn useful features, but such labels in cell images are difficult to obtain. In realistic scenarios, we only have partial or no prior knowledge of the target compounds, which narrows the application of supervised classification.

Recently, a vital breakthrough in computer vision known as representation learning [29–34] aims to learn a good general representation without task-specific supervision. Extensive methods have been proposed for natural images and have yielded remarkable results. In contrast to natural images, metadata such as batch numbers, compound concentrations, and genetic perturbations are immediately available together with cell image data. Thus, many methods use the metadata of cell images as a pretext or surrogate classification task to train CNNs, expecting to obtain more discriminative feature representations of cell phenotypes [91–93]. Specifically, Caicedo et al. [92] used individual cells as inputs rather than the entire image. Individual cells as input can filter out the interference of the background and non-cell impurities but also cause the network to neglect global information such as cell densities. Spiegel et al. [93] further investigated the impact of the number of classes of pretext tasks and the difference between implicit and explicit learning [94]. Instead of using metadata for surrogate supervision, Janssens et al. [95] used deep clustering [96] to assign pseudolabels to each feature vector. Both obtained promising results on the BBBC021 dataset [97, 98], which is a public dataset for validating the MoAs of 103 different compound concentrations. More recently, Wang et al. proposed a framework called TEAMs [99] and achieved state-of-the-art performance on three cell-painting datasets [100]. Similar to [93], they also used metadata as supervision and built a framework based on conventional metric learning [94]. They further upgraded it with three modules to handle the negative sampling of metric learning and distribution shift between training/testing.

Another line of research involves learning cell embeddings by reconstructing cell images. The basic concept is that, provided a good representation of the cell phenotype, an accurate cell image can be reconstructed from it. Goldsborough et al. first used a generative adversarial network (GAN) [101] to generate cell images [102]. They used the output of the penultimate layer of the discriminator as a feature representation of cell phenotypes. Because the effectiveness of the discriminator of the GAN is highly dependent on the generator, this method cannot obtain a satisfactory result. Lu et al. [103] used an encoder-decoder network to reconstruct cells. They used a fully observed cell as an information source for phenotypes to paint an incomplete cell image with the target channel manually concealed. After the network converges, the encoder output is the final representation. Kobayashi et al. further proposed a new pipeline called Cytoself, based on a VQ-VAE-2 [104] model, to reconstruct endogenous tagged fluorescent images and classify tagged proteins simultaneously [105]. They tested their method on a dataset that tagged 1311 different proteins and showed that such self-supervised algorithms could learn useful feature representations that encoded the localization information of proteins, which is highly associated with the functionality of proteins [106]. Contrastive learning has also been applied to images of cells. For instance, Perakis et al. deployed the simCLR [107] framework to learn the cell phenotypic features [108]. However, they made only a marginal improvement.

Overall, both types of approaches have been proven to successfully learn useful representations of cell phenotypes for downstream analyses. However, more effective frameworks are yet to be developed. For example, the information contained in different treatments is valuable for the implicit representation of cells, as different molecules can cause different phenotypes. Yang et al. used molecular embeddings to indicate cell image synthesis and made significant improvements [109]. Using graph neural networks to predict molecular properties also achieves impressive results [109–113]. Thus, completely using multimodal information appears promising for cell image analysis.

### 3. The Challenges and Opportunities of Deep-Learning Methods in Cellular Image Processing

As discussed in the previous sections, deep learning has demonstrated an incredible ability to perform cell image analysis. However, there remains a significant performance gap between deep-learning algorithms in academic research and practical applications. Generally, sufficient and accurate training data are necessary for deep-learning algorithms to guarantee the generalization of the trained models. However, in practice, it is laborious and demanding to collect exhaustive annotations of cell images, which requires numerous biological experts and their efforts. Consequently, practical cell image datasets, which may contain defective training data, can dramatically degrade the performance of deep-learning algorithms. Thus, to further improve deep learning-based cell image analysis, defects in cell image datasets should also be considered and properly solved. In this section, we will discuss the major challenges of cell image analysis and the existing methods from a *data* perspective. We focus on three aspects of cell image datasets, *data quantity*, *data quality*, and *data confidence*, which are discussed in detail in the following sections.

*3.1. Deep Learning with Small But Expensive Dataset.* Currently, although cell images can be collected using microscopy and cameras in a high-throughput fashion, constructing a large-scale cell image dataset remains a strenuous task. This is because, compared with common images, cell images require knowledgeable biological experts to assign labels image by image, which is time-consuming and demanding. Thus, the scale of cell image datasets is often limited by the difficulty of annotation. Fortunately, alternative strategies are available for mitigating this problem. The first strategy is *dataset expansion* (Section 3.1.1), which aims to increase the quantity of training data from labeled or unlabeled images. Data augmentation, which is a widely adopted technique in deep learning, can aid acquire extra training images from labeled images by performing image transformation. In Section 3.1.1, we focus on a key technique called *active learning*, which can automatically select valuable unlabeled images to expand the scale of the training data by interacting with human experts. The second strategy is *knowledge transference* (Section 3.1.2), which aims to improve the performance of deep-learning models by

transferring knowledge contained in other datasets. In Section 3.1.2, we discuss a corresponding technique called *transfer learning*, which can be divided into model- and feature-based approaches.

*3.1.1. Collecting Datasets Efficiently.* As discussed previously, the manual annotations of cell images are laborious and expensive and can only be performed by professionals with rich knowledge. However, although the performance of deep-learning models increased as the amount of data, not all labeled data have the same value in learning effective feature representation. Active learning is proposed to solve these difficulties by selecting training samples with high learning values from the unlabeled data pool to annotate. Under the same cost, active learning does not increase the number of labeled data but constructs a more valuable training dataset that aims to increase the performance of trained learning algorithms.

To investigate the effectiveness of active learning in reducing the cost of phenotypic classification, Smith and Horvath [114] compared various combinations of active learning methods, such as least confident, vote entropy sampling, and margin sampling, with supervised learning algorithms, such as support vector machine (SVM), naïve Bayes, and random forest. Their experimental results on three phenotyping datasets demonstrated that active learning could achieve a performance similar to that of previous methods while largely reducing the labeling cost. Cell tracking aims to capture the dynamic movement of cells and is a fundamental tool in high-content screening for modern drug discovery. Compared with general object tracking, cell tracking faces some unique challenges, such as high similarity in the appearance of cells, low temporal resolution, and various cellular activities, thus requiring more labeled data. Lou et al. [115] proposed a structured learning model with an active learning strategy for cell tracking that has the advantages of automatic parameter learning, higher feature dimensions, and lower annotation cost. Their active learning strategy included four components: dividing images into representative patches, measuring the uncertainty of unlabeled structured data, updating the parameters of the model, and checking the terminal criteria. They evaluated the proposed cell tracking algorithm on five datasets, and the experimental results showed that their active learning method could only use 17% labeled data to achieve a performance similar to that of the baseline model with all training data. Screening methods have been widely used to identify drug candidates that perturb specific targets. Ideally, conducting experiments to test all combinations would be an effective method to discover the desired drug candidates. However, this approach is usually infeasible. Naik et al. [116] showed that an active learning method without any prior knowledge could effectively iteratively select a subset from a pool of biology experiments to learn various compound effects and perform better than the strategies that might be employed by humans. Automatic segmentation of nuclei aims to extract nuclei pixels from entire tissue slides, which is a core step in computer-aided pathology analysis. However, the generalization ability of nucleus segmentation methods with

fixed parameters is usually low because the morphology and texture of different types of nuclei vary significantly. Wen et al. [117] proposed the application of active learning with different classification methods to measure the quality of the nucleus segmentation results. This active learning procedure iteratively improved the generalization ability of the learning model under limited labeled samples. Directly combining a patched-based classifier with active learning makes it difficult to manually annotate selected patches with small sizes and a lack of context information. An active learning method with a core-set sampling strategy [118] tackled this challenge by merging uncertain patches into regions for annotation; this strategy did not affect the training procedure of the classifier that still operated on patches. To effectively leverage the information in labeled and unlabeled data, Lai et al. [119] proposed a label-efficient framework with active learning and semi-supervised learning for brain tissue segmentation in gigapixel pathology images, which surpassed fully supervised learning methods by using only 0.1% annotations.

*3.1.2. Transferring Knowledge from Other Large-Scale Datasets.* Transfer learning, which focuses on transferring knowledge across domains, is a promising machine learning methodology for solving data-scarcity problems. Inspired by how quickly humans learn new knowledge from similar experiences, transfer learning aims to leverage knowledge from related domains (also known as the source domain) to improve learning performance in the target domain while minimizing the number of labeled examples required [120]. Deep transfer learning (DTL), which is a combination of deep-learning architectures and transfer learning, is the most commonly used type of transfer learning in drug discovery [121]. DTL has yielded impressive results in applications such as automatic cell segmentation [122], prediction of protein subcellular localization [123], and prediction of MoA [124]. Compared with traditional machine learning approaches, deep learning uses DNNs with multiple hidden layers that can represent and learn more complex knowledge. Although the transferability of features decreases as the distance between the source and target domains increases, transferring features from distant tasks can be better than using random features [125]. For example, Khan et al. [126] used an ensemble of three CNNs (GoogleNet [127], VGGNet [49], and ResNet [50]) pretrained on the ImageNet [128] dataset to extract general features from breast cytology images so that the accuracy in the detection and classification of malignant cells is greater than 97%. Studies [123, 128–130] have shown that pretrained models deliver better predictive performance with less training time and fewer training samples. Based on current applications in phenotype feature representation for drug discovery, we classify transfer learning methods into parameter- and feature-based approaches.

*(1) Parameter-Based Approaches.* The parameters of the pretrained model reflect what the model learns in the source domain; therefore, knowledge can be transferred directly at the parameter level. An intuitive method is to use the

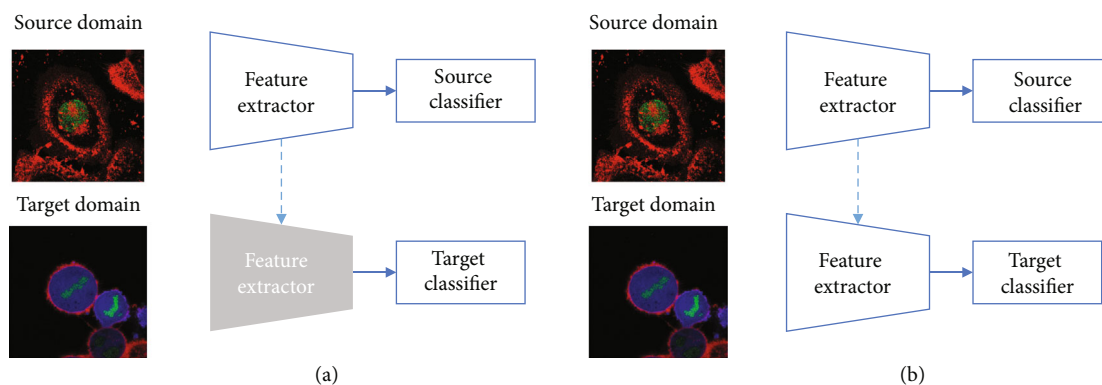


FIGURE 3: Schematic of parameter-based deep transfer learning methods. (a) Extract features of the target domain using a pretrained feature extractor without training. (b) The target learner considers the parameters of the pretrained feature extractor as the initial value and fine-tunes on the target domain. (Dashed line indicates parameter sharing; gray indicates freeze of training.)

parameters of the pretrained network directly as a feature extractor without additional training (Figure 1(a)). An image of the target domain is passed through the pretrained model to obtain its features, which serve as inputs to the downstream task. For example, Pawlowski et al. [131] extracted features from ImageNet pretrained neural networks and evaluated the task of classifying each treatment condition into its MoA using a 1-nearest neighbor classifier. Phan et al. [132] applied transfer learning from a pretrained network to extract generic features and then used the minimum redundancy maximum relevance (mRMR), a feature selection method, to obtain the most relevant features for classification (Figure 3).

The pretrained model can also be used to initialize the target model and be fine-tuned to the target task, as shown in Figure 1(b). Kraus et al. [123] trained a deep CNN (DeepLoc) for subcellular protein localization prediction. They showed that, in contrast to traditional approaches, the model could be successfully transferred to datasets with different genetic backgrounds acquired from other laboratories, even those with abnormal cellular morphology, by fine-tuning.

(2) *Feature-Based Approaches.* Feature-based approaches transform each original feature into a new representation for knowledge transfer. The goal is to find a common latent feature space in which the source and target data can have the same probability distribution. Therefore, the source data can be used as a training set for target tasks in the latent feature space, helping improve the performance of the model for the target data. There are two common methods to obtain domain-invariant features. One is to reduce the distribution difference between the source and target domain instances (Figure 2(a)). For example, Bermúdez-Chacón et al. [133] proposed a two-stream U-Net for electron microscopy image segmentation. One stream used source-domain samples, whereas the other used target data. They utilized the maximum mean discrepancy (MMD) and correlation alignment as domain regularization to use training data from the source domain to adjust the network weights in the target domain. Measurement MMD is widely used in transfer learning, which quantifies the distribution difference by cal-

culating the distance of the mean values of the instances in a reproducing kernel Hilbert space (RKHS). In addition to MMD, several measurement criteria have been adopted in transfer learning, including the Kullback-Leibler divergence [134], Jensen-Shannon divergence [135], and Wasserstein distance [136] (Figure 4).

The other is an adversarial-based method, which is promising for generating complex samples across different domains (e.g., GANs [101]). The original GAN is composed of a generator G and discriminator D. The goal of the generator G is to produce counterfeits of the actual data to confuse the discriminator. Discriminator D is fed a mixture of the actual data and counterfeits, and it aims to detect whether the data are actual or fake. Motivated by GAN, many transfer learning approaches have been established based on the assumption that a good feature representation contains almost no discriminative information regarding the original domains of the instances. Figure 2(b) shows an adversarial-based method that typically includes a shared-feature transformer, domain, source, and target classifier. The feature transformer, similar to the generator, aims to produce a domain-independent feature representation to confuse the domain classifier. The domain classifier plays the role of a discriminator, which attempts to detect whether the extracted features come from the source or target domains. The source and target classifiers produced label predictions for the source and target tasks, respectively. Adversarial-based learning methods have been widely used in recent years for not only transfer learning but also for data augmentation and addressing batch effects. For example, Boyd et al. [137] proposed domain-adversarial autoencoders to promote domain-invariant representations between cell lines, which not only improved the accuracy of MOA prediction but also enabled the comparison of the effects of drugs on different cell lines. Qian et al. [138] proposed a GAN-based batch equalization method that could transfer images from one batch to another while preserving the biological phenotype to address the batch effect.

3.2. *Deep Learning with Noisy and Imbalanced Labels.* As mentioned previously, annotating cell images requires

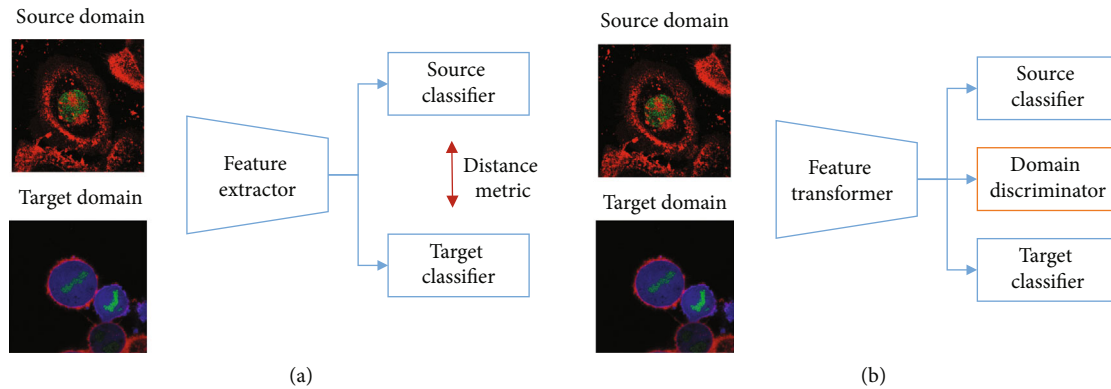


FIGURE 4: Schematic of feature-based deep transfer learning methods. (a) Transfer knowledge by finding a feature representation that reduces the difference between the source and target domains. (b) Find the domain-invariant features by training the transformed feature representation that could fool the domain classifier.

human annotators with profound biological knowledge. Therefore, the quality of the annotations of cell image datasets is highly dependent on the professional skills of human annotators, which may cause intractable issues influencing the training of deep-learning models. Specifically, assigning incorrect or incomplete labels to training images introduces numerous *label noise*, damaging the generalization of deep-learning models. Notably, even if the assigned labels are completely accurate, the preference for annotation may result in another issue referred to as *label imbalance*, where the numbers of labeled images for different classes are quite unbalanced.

To improve the robustness of supervised learning models against noisy labels, existing studies have proposed several types of strategies, such as regularization methods to reduce overfitting on noisy labels, robust architectures to model noise, sample selectors to filter out noise, and loss functions to underweight noisy samples. Caicedo et al. [92] presented an RNN-based regularization to remove unrelated features resulting from noisy labels for weakly supervised single-cell profiling. An unsupervised learning method for nuclear segmentation in brain images was proposed in [139] to iteratively train a mask R-CNN model with automatically generated noisy instance segmentation masks and refine the labels using an expectation and maximization (EM) procedure. Park et al. [140] proposed a robust neuron segmentation method that leveraged ADMSE loss to adaptively reduce the weights of noisy labels. Annotating data by multiple experts improves the quality of labels; however, inconsistency among experts could be a type of noisy label in training models. Xiao et al. [141] resolved this issue for pathological image segmentation by utilizing the surrounding context of pixels to compute the weights of the labels annotated by multiple experts.

Several efforts have been made to resolve this problem. Resampling and reweighting are two fundamental strategies to rebalance the distributions in model training from the perspective of input data and loss functions, respectively. Data resampling methods aim to construct a balanced dataset by oversampling the minority classes or undersampling the majority classes. For example, an undersampling strategy and *K*-means clustering were used in drug discovery to

remove the less important samples in the nonpotential drug class, which was the majority class, and the degree of importance was measured by the distance between samples and their cluster centroids [142]. In contrast to data resampling methods, loss reweight methods retain all data samples while assigning different weights to them to alleviate the effects of imbalance. Because the number of normal cells was considerably larger than the number of abnormal cells, a focal loss [143] was used in [144] to enlarge the weights of hard samples and reduce the weights of easy data for the morphological classification of red blood cells. Dice loss was improved with reweighting strategies for cell segmentation [145] and detection [146]. They achieved a better performance than vanilla loss. Because reweighted methods only change the loss functions and do not alter the network architectures, they can be easily combined with other machine-learning techniques to improve the performance of imbalanced tasks. For example, CBCM [147] integrates focal loss with transfer learning to classify images of bone marrow cells that follow a long-tailed data distribution. Recently, two-stage training frameworks [148, 149] have shown high performance on imbalanced natural image datasets, first on the original data distribution and then fine-tuned with rebalanced techniques. However, this paradigm has not been validated on imbalanced cell image tasks and may be a direction worth exploring for further studies.

**3.3. Uncertainty-Aware Cell Image Analysis.** In biological scenarios, deep learning has frequently been used as an efficient tool for processing biological images and outputting predictions for subsequent steps. Traditionally, annotations of cell images are always deterministic (for instance, deterministic cell boundaries or regions in a cellular image). It is rare and even nonexistent that cell image datasets would contain the confidence of annotations. Therefore, using these datasets, the predictions of traditional plain neural networks are deterministic. However, it is important to acquire the uncertainty of the prediction output using deep-learning models, which can help experts evaluate the confidence of these results and measure the robustness of deep-learning models with a probabilistic interpretation. Owing to the lack of confidence information in datasets, traditional plain



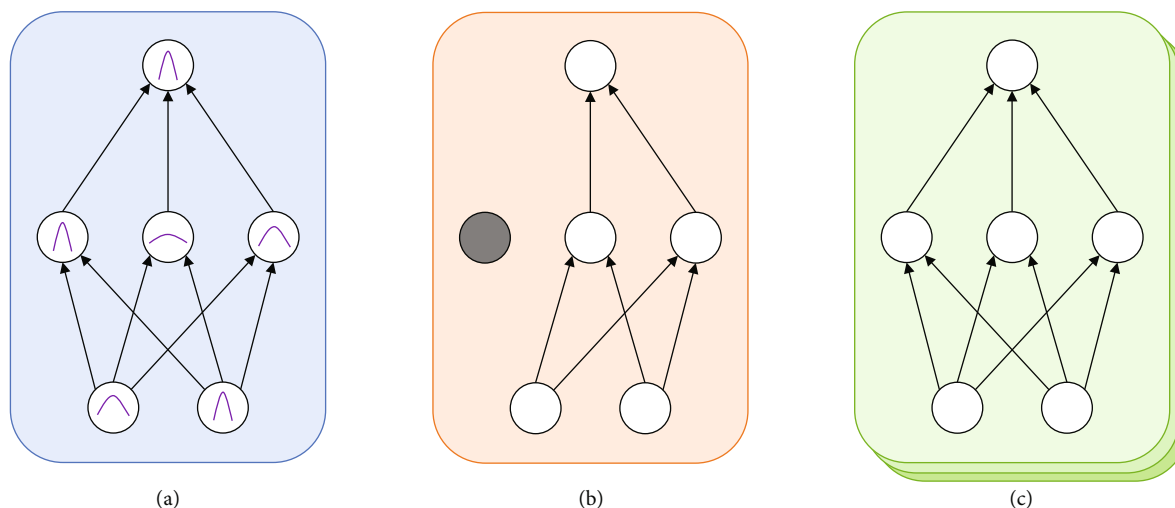


FIGURE 5: Schematic of general uncertainty estimation methods. (a) Bayes by Backprop: consider the weights of the neural network as Gaussian distributions (denoted by the purple lines). (b) MC dropout: adopt dropout in the training and inference phase, where the gray circle indicates the neuron dropped by dropout. (c) Deep ensembles: assemble multiple neural networks (denoted by the stacked rectangles).

neural networks are incapable of capturing uncertainty, which causes some unexpected problems, resulting in difficulties. For instance, in a drug discovery procedure, when applying a deep-learning classifier of cell phenotypes, a cell image with an unseen phenotype may be assigned to one of the classes presented in the training set because there is no mechanism to reflect the confidence of classification results and a plain neural network fails to indicate that it is a new phenotype. Therefore, uncertainty-aware learning is crucial for deep-learning applications in biological scenarios.

In response to this issue, many general methods have been proposed to estimate uncertainty in deep learning. Using a Bayesian method to mathematically model uncertainty, Blundell et al. proposed an algorithm called Bayes by Backprop, which uses variational Bayesian learning to introduce uncertainty in the weights of neural networks [150]. Instead of considering the weights of networks as fixed values, Bayes by Backprop assumed them to be independent Gaussian distributions and learned them using variational Bayesian learning. In addition to acting as an uncertainty estimation method, Bayes by Backprop can also be used to perform regularization based on the compression cost of the weights. However, the training and inference of Bayes by Backprop are time-consuming and memory inefficient, limiting its use in practical applications. To address this problem, Gal and Ghahramani used dropout, a common network regularization technique, to develop a new theoretical framework for uncertainty [151]. Gal and Ghahramani showed that using dropout correctly was mathematically equivalent to approximating the probabilistic deep Gaussian process and thus proposed Monte Carlo (MC) dropout for uncertainty estimation. MC dropout is performed using dropout in both training and inference, which does not sacrifice either the computational complexity or inference accuracy. By replacing dropout with other similar techniques (e.g., DropBlock [152], DropConnect [153], and SpatialDropout [154]), the MC dropout performance can be further improved. Moreover, it is worth noting that

Lakshminarayanan et al. proposed a different uncertainty estimation method called deep ensembles, which also uses an alternative approach rather than a Bayesian-based one to capture uncertainty in deep learning [155]. Deep ensembles are simple to implement and provide high-quality uncertainty estimates. For instance, it can yield a higher uncertainty in out-of-distribution data. For a clear comparison, schematic of Bayes by Backprop, MC dropout, and deep ensembles are shown in Figure 5.

Notably, some previous studies have considered the uncertainty in practical bioscience applications. Carrieri et al. predicted the host phenotype by maintaining compact representations of genetic material [156]. As one of the evaluation metrics, the uncertainty of predictions from four different classifiers was estimated to measure the performance of the proposed workflow, acquired by cross-validation, and a relevance vector machine (RVM). To reduce the cost of high-throughput screening using categorical matrix completion and active learning, Chen et al. designed an algorithm to guide experiments based on chemical compound effects on subcellular locations of various proteins [157]. In this algorithm, uncertainty estimation is performed for sparse matrix completion and implemented by margin sampling. Some in-depth studies have explored the uncertainty estimation of deep learning in biological scenarios. Using deep Bayesian learning, Gomariz et al. proposed a deep learning-based cell detection framework that could output the desired probabilistic predictions [158], where Bayesian regression techniques were used in uncertainty-aware density maps. In this study, MC dropout is used to capture aleatoric and epistemic uncertainty in the training data, which are used to generate spatial epistemic and aleatoric uncertainty maps as additional inputs for the classifier. A neural network that can capture uncertainty can distinguish between seen and unseen examples because uncertainty will decrease as more examples are observed, allowing the results output by networks to become more deterministic. Using this characteristic, Dürr et al. proposed the exploitation of

MC dropout to define different uncertainty measures for each phenotype prediction in a real-world biological dataset, which showed that these uncertainty measures can be used to recognize new or unclear phenotypes [159]. Thus, the uncertainty estimation of neural networks also shows potential for discovering new phenotypes. To track single cells in colonies without manual intervention, Theorell et al. developed a novel probabilistic tracking paradigm called uncertainty-aware tracking, which is based on a Bayesian approach to perform lineage hypothesis formation [160]. The introduction of uncertainty in this study improved the accuracy and tracking-induced errors.

#### 4. Conclusion

In this study, we provide a comprehensive survey of three critical tasks in cell image analysis—segmentation, tracking, and classification—which shows that deep learning has been widely applied to these tasks and achieves promising results. As a data-driven method, deep learning often suffers from a lack of high-quality datasets for biological scenarios. Consequently, a performance gap often exists between academic research and practical application. From a data perspective, we also discuss the challenges of applying machine-learning algorithms to cell image analysis. We hope that the discussed techniques and concepts can provide insights for both the biology and computer vision communities to propose more efficient solutions and promote the applications of deep learning in biomedical and life sciences.

#### Conflicts of Interest

The authors declare no conflicts of interest.

#### Authors' Contributions

Junde Xu and Donghao Zhou contributed equally to this work.

#### Acknowledgments

This work was supported by the National Key Research and Development Program of China (2022YFE020\0700), National Natural Science Foundation of China (Project No. 62006219), Natural Science Foundation of Guangdong Province (2022A1515011579), and Hong Kong Innovation and Technology Fund Project No. GHP/110/19SZ and ITS/170/20.

#### References

- [1] D. C. Swinney, "Phenotypic vs. target-based drug discovery for first-in-class medicines," *Clinical Pharmacology and Therapeutics*, vol. 93, no. 4, pp. 299–301, 2013.
- [2] D. C. Swinney and J. Anthony, "How were new medicines discovered?," *Nature Reviews. Drug Discovery*, vol. 10, no. 7, pp. 507–519, 2011.
- [3] M. Götte, C. Mohr, C. Y. Koo et al., "MiR-145-dependent targeting of junctional adhesion molecule A and modulation of fascin expression are associated with reduced breast cancer cell motility and invasiveness," *Oncogene*, vol. 29, no. 50, pp. 6569–6580, 2010.
- [4] J. C. Caicedo, S. Cooper, F. Heigwer et al., "Data-analysis strategies for image-based cell profiling," *Nature Methods*, vol. 14, no. 9, pp. 849–863, 2017.
- [5] S. J. Hassenbusch, R. K. Portenoy, M. Cousins et al., "Polyanalgesic consensus conference 2003: an update on the management of pain by intraspinal drug delivery— report of an expert panel," *Journal of Pain and Symptom Management*, vol. 27, no. 6, pp. 540–563, 2004.
- [6] P. Schneider, W. P. Walters, A. T. Plowright et al., "Rethinking drug design in the artificial intelligence era," *Nature Reviews. Drug Discovery*, vol. 19, no. 5, pp. 353–364, 2020.
- [7] A. Bender, D. Bojanic, J. W. Davies et al., "Which aspects of HTS are empirically correlated with downstream success?," *Current Opinion in Drug Discovery and Development*, vol. 11, no. 3, p. 327, 2008.
- [8] Y. Gilad, K. Nadassy, and H. Senderowitz, "A reliable computational workflow for the selection of optimal screening libraries," *Journal of Cheminformatics*, vol. 7, no. 1, pp. 1–17, 2015.
- [9] J. Bajorath, "Extending accessible chemical space for the identification of novel leads," *Expert Opinion on Drug Discovery*, vol. 11, no. 9, pp. 825–829, 2016.
- [10] E. Moen, D. Bannon, T. Kudo, W. Graf, M. Covert, and D. Van Valen, "Deep learning for cellular image analysis," *Nature Methods*, vol. 16, no. 12, pp. 1233–1246, 2019.
- [11] V. Ulman, M. Maška, K. E. G. Magnusson et al., "An objective comparison of cell-tracking algorithms," *Nature Methods*, vol. 14, no. 12, pp. 1141–1152, 2017.
- [12] M. D. Abràmoff, P. J. Magalhães, and S. J. Ram, "Image processing with ImageJ," *Biophotonics International*, vol. 11, no. 7, pp. 36–42, 2004.
- [13] A. E. Carpenter, T. R. Jones, M. R. Lamprecht et al., "Cellprofiler: image analysis software for identifying and quantifying cell phenotypes," *Genome Biology*, vol. 7, no. 10, pp. 1–11, 2006.
- [14] F. De Chaumont, S. Dallongeville, N. Chenouard et al., "Icy: an open bioimage informatics platform for extended reproducible research," *Nature Methods*, vol. 9, no. 7, pp. 690–696, 2012.
- [15] M. Held, M. H. A. Schmitz, B. Fischer et al., "CellCognition: time-resolved phenotype annotation in high-throughput live cell imaging," *Nature Methods*, vol. 7, no. 9, pp. 747–754, 2010.
- [16] A. Vouloudimos, N. Doulamis, A. Doulamis, and E. Protopapadakis, "Deep learning for computer vision: a brief review," *Computational Intelligence and Neuroscience*, vol. 2018, Article ID 7068349, 13 pages, 2018.
- [17] Z.-Q. Zhao, P. Zheng, S. Xu, and X. Wu, "Object detection with deep learning: a review," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 11, pp. 3212–3232, 2019.
- [18] E. Cambria and B. White, "Jumping NLP curves: a review of natural language processing research [review article]," *IEEE Computational Intelligence Magazine*, vol. 9, no. 2, pp. 48–57, 2014.
- [19] M. Reichstein, G. Camps-Valls, B. Stevens et al., "Deep learning and process understanding for data-driven Earth system science," *Nature*, vol. 566, no. 7743, pp. 195–204, 2019.
- [20] A. Vaswani, N. Shazeer, N. Parmar et al., "Attention is all you need," *Advances in Neural Information Processing Systems*, vol. 30, 2017.

- [21] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [22] Y. Al-Kofahi, A. Zaltsman, R. Graves, W. Marshall, and M. Rusu, "A deep learning-based algorithm for 2-D cell segmentation in microscopy images," *BMC Bioinformatics*, vol. 19, no. 1, pp. 1–11, 2018.
- [23] T. Falk, D. Mai, R. Bensch et al., "U-Net: deep learning for cell counting, detection, and morphometry," *Nature Methods*, vol. 16, no. 1, pp. 67–70, 2019.
- [24] N. F. Greenwald, G. Miller, E. Moen et al., "Whole-cell segmentation of tissue images with human-level performance using large-scale data annotation and deep learning," *Nature Biotechnology*, vol. 40, no. 4, pp. 555–565, 2022.
- [25] J.-B. Lugagne, H. Lin, and M. J. Dunlop, "DeLTA: automated cell segmentation, tracking, and lineage reconstruction using deep learning," *PLoS Computational Biology*, vol. 16, no. 4, article e1007673, 2020.
- [26] E. Moen, E. Borba, G. Miller et al., "Accurate cell tracking and lineage construction in live-cell imaging experiments with deep learning," *Biorxiv*, no. article 803205, 2019.
- [27] T. He, H. Mao, J. Guo, and Z. Yi, "Cell tracking using deep neural networks with multi-task learning," *Image and Vision Computing*, vol. 60, pp. 142–153, 2017.
- [28] S. N. Chandrasekaran, H. Ceulemans, J. D. Boyd, and A. E. Carpenter, "Image-based profiling for drug discovery: due for a machine-learning upgrade?," *Nature Reviews. Drug Discovery*, vol. 20, no. 2, pp. 145–159, 2021.
- [29] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: a review and new perspectives," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 8, pp. 1798–1828, 2013.
- [30] W. L. Hamilton, R. Ying, and J. Leskovec, "Representation learning on graphs: methods and applications," 2017, <https://arxiv.org/abs/1709.05584>.
- [31] M. Noroozi, H. Pirsiavash, and P. Favaro, "Representation learning by learning to count," in *Proceedings of the IEEE international conference on computer vision*, pp. 5898–5906, Honolulu, Hawaii, USA, 2017.
- [32] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," 2015, <https://arxiv.org/abs/1511.06434>.
- [33] S. Gidaris, P. Singh, and N. Komodakis, "Unsupervised representation learning by predicting image rotations," 2018, <https://arxiv.org/abs/1803.07728>.
- [34] S. Arora, H. Khandeparkar, M. Khodak, O. Plevrakis, and N. Saunshi, "A theoretical analysis of contrastive unsupervised representation learning," 2019, <https://arxiv.org/abs/1902.09229>.
- [35] A. Pratapa, M. Doron, and J. C. Caicedo, "Image-based cell phenotyping with deep learning," *Current Opinion in Chemical Biology*, vol. 65, pp. 9–17, 2021.
- [36] B. Liu, H.-D. Cheng, J. Huang, J. Tian, X. Tang, and J. Liu, "Fully automatic and segmentation-robust classification of breast tumors based on local texture analysis of ultrasound images," *Pattern Recognition*, vol. 43, no. 1, pp. 280–298, 2010.
- [37] K. Mkrtchyan, D. Singh, M. Liu, V. Reddy, A. Roy-Chowdhury, and M. Gopi, "Efficient cell segmentation and tracking of developing plant meristem," in *2011 18th IEEE International Conference on Image Processing*, pp. 2165–2168, Brussels, Belgium, 2011.
- [38] R. Bensch and O. Ronneberger, "Cell segmentation and tracking in phase contrast images using graph cut with asymmetric boundary costs," in *2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI)*, pp. 1220–1223, Brooklyn, NY, USA, 2015.
- [39] R. Bise and Y. Sato, "Cell detection from redundant candidate regions under nonoverlapping constraints," *IEEE Transactions on Medical Imaging*, vol. 34, no. 7, pp. 1417–1427, 2015.
- [40] H. Su, Z. Yin, S. Huh, and T. Kanade, "Cell segmentation in phase contrast microscopy images via semi-supervised classification over optics-related features," *Medical Image Analysis*, vol. 17, no. 7, pp. 746–765, 2013.
- [41] K. Li and T. Kanade, "Nonnegative mixed-norm preconditioning for microscopy image segmentation," in *International Conference on Information Processing in Medical Imaging*, pp. 362–373, Berlin, Heidelberg, 2009.
- [42] Z. Yin, T. Kanade, and M. Chen, "Understanding the phase contrast optics to restore artifact-free microscopy images for segmentation," *Medical Image Analysis*, vol. 16, no. 5, pp. 1047–1062, 2012.
- [43] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: towards real-time object detection with region proposal networks," *Advances in Neural Information Processing Systems*, vol. 28, 2015.
- [44] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2961–2969, Honolulu, Hawaii, USA, 2017.
- [45] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2117–2125, Honolulu, Hawaii, USA, 2017.
- [46] J. W. Johnson, "Adapting Mask-RCNN for automatic nucleus segmentation," 2018, <https://arxiv.org/abs/1805.00500>.
- [47] H.-F. Tsai, J. Gajda, T. F. W. Sloan, A. Rares, and A. Q. Shen, "Usiigaci: instance-aware cell tracking in stain-free phase contrast microscopy enabled by machine learning," *SoftwareX*, vol. 9, pp. 230–237, 2019.
- [48] R. Hollandi, A. Szkalisity, T. Toth et al., "A deep learning framework for nucleus segmentation using image style transfer," *Biorxiv*, no. article 580605, 2019.
- [49] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, <https://arxiv.org/abs/1409.1556>.
- [50] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778, Las Vegas, Nevada, USA, 2016.
- [51] A. Neubeck and L. Van Gool, "Efficient non-maximum suppression," in *18th International Conference on Pattern Recognition (ICPR'06)*, pp. 850–855, Hong Kong, China, 2006.
- [52] Y. Song, E. L. Tan, X. Jiang et al., "Accurate cervical cell segmentation from overlapping clumps in pap smear images," *IEEE Transactions on Medical Imaging*, vol. 36, no. 1, pp. 288–300, 2017.
- [53] P. Kainz, M. Urschler, S. Schuler, P. Wohlhart, and V. Lepetit, "You should use regression to detect cells," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 276–283, Cham, 2015.

- [54] M. Bai and R. Urtasun, "Deep watershed transform for instance segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5221–5229, Honolulu, Hawaii, USA, 2017.
- [55] L. Shafarenko, M. Petrou, and J. Kittler, "Automatic watershed segmentation of randomly textured color images," *IEEE Transactions on Image Processing*, vol. 6, no. 11, pp. 1530–1544, 1997.
- [56] C. F. Koyuncu, G. N. Gunesli, R. Cetin-Atalay, and C. Gunduz-Demir, "\_DeepDistance\_ : a multi-task deep regression model for cell detection in inverted microscopy images," *Medical Image Analysis*, vol. 63, article 101720, 2020.
- [57] D. Eschweiler, T. V. Spina, R. C. Choudhury, E. Meyerowitz, A. Cunha, and J. Stegmaier, "CNN-based preprocessing to optimize watershed-based cell segmentation in 3D confocal microscopy images," in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, pp. 223–227, Venice, Italy, 2019.
- [58] J. Cao, G. Guan, V. W. S. Ho et al., "Establishment of a morphological atlas of the *Caenorhabditis elegans* embryo using deep-learning-based 4D segmentation," *Nature Communications*, vol. 11, no. 1, pp. 1–14, 2020.
- [59] U. Schmidt, M. Weigert, C. Broaddus, and G. Myers, "Cell detection with star-convex polygons," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 265–273, Cham, 2018.
- [60] Z. Cao, T. Simon, S.-E. Wei, and Y. Sheikh, "Realtime multi-person 2D pose estimation using part affinity fields," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7291–7299, Honolulu, Hawaii, USA, 2017.
- [61] D. Neven, B. De Brabandere, M. Proesmans, and L. Van Gool, "Instance segmentation by jointly optimizing spatial embeddings and clustering bandwidth," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 8837–8845, Long Beach, CA, USA, 2019.
- [62] C. Stringer, T. Wang, M. Michaelos, and M. Pachitariu, "Cellpose: a generalist algorithm for cellular segmentation," *Nature Methods*, vol. 18, no. 1, pp. 100–106, 2021.
- [63] J. C. Caicedo, A. Goodman, K. W. Karhohs et al., "Nucleus segmentation across imaging experiments: the 2018 Data Science Bowl," *Nature Methods*, vol. 16, no. 12, pp. 1247–1253, 2019.
- [64] C. Edlund, T. R. Jackson, N. Khalid et al., "LIVECell—a large-scale dataset for label-free live cell segmentation," *Nature Methods*, vol. 18, no. 9, pp. 1038–1045, 2021.
- [65] J. E. Purvis and G. Lahav, "Encoding and decoding cellular information through signaling dynamics," *Cell*, vol. 152, no. 5, pp. 945–956, 2013.
- [66] J. C. Kimmel, A. Y. Chang, A. S. Brack, and W. F. Marshall, "Inferring cell state by quantitative motility analysis reveals a dynamic state system and broken detailed balance," *PLoS Computational Biology*, vol. 14, no. 1, article e1005927, 2018.
- [67] P. Wang, L. Robert, J. Pelletier et al., "Robust growth of *Escherichia coli*," *Current Biology*, vol. 20, no. 12, pp. 1099–1103, 2010.
- [68] S. Cooper, A. R. Barr, R. Glen, and C. Bakal, "NucliTrack: an integrated nuclei tracking application," *Bioinformatics*, vol. 33, no. 20, pp. 3320–3322, 2017.
- [69] K. E. G. Magnusson, J. Jaldén, P. M. Gilbert, and H. M. Blau, "Global linking of cell tracks using the Viterbi algorithm," *IEEE Transactions on Medical Imaging*, vol. 34, no. 4, pp. 911–929, 2014.
- [70] X. Wang, W. He, D. Metaxas, R. Mathew, and E. White, "Cell segmentation and tracking using texture-adaptive snakes," in *2007 4th IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pp. 101–104, Arlington, VA, USA, 2007.
- [71] K. Li, E. D. Miller, M. Chen, T. Kanade, L. E. Weiss, and P. G. Campbell, "Cell population tracking and lineage construction with spatiotemporal context," *Medical Image Analysis*, vol. 12, no. 5, pp. 546–566, 2008.
- [72] F. Amat, W. Lemon, D. P. Mossing et al., "Fast, accurate reconstruction of cell lineages from large-scale fluorescence microscopy data," *Nature Methods*, vol. 11, no. 9, pp. 951–958, 2014.
- [73] D. A. Van Valen, T. Kudo, K. M. Lane et al., "Deep learning automates the quantitative analysis of individual cells in live-cell imaging experiments," *PLoS Computational Biology*, vol. 12, no. 11, article e1005177, 2016.
- [74] J. M. Newby, A. M. Schaefer, P. T. Lee, M. G. Forest, and S. K. Lai, "Convolutional neural networks automate detection for tracking of submicron-scale particles in 2D and 3D," *Proceedings of the National Academy of Sciences*, vol. 115, no. 36, pp. 9026–9031, 2018.
- [75] S. U. Akram, J. Kannala, L. Eklund, and J. Heikkilä, "Cell tracking via proposal generation and selection," 2017, <https://arxiv.org/abs/1705.03386>.
- [76] C. Payer, D. Štern, T. Neff, H. Bischof, and M. Urschler, "Instance segmentation and tracking with cosine embeddings and recurrent hourglass networks," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 3–11, Cham, 2018.
- [77] Z. Zhou, F. Wang, W. Xi, H. Chen, P. Gao, and C. He, "Joint multi-frame detection and segmentation for multi-cell tracking," in *International Conference on Image and Graphics*, pp. 435–446, Cham, 2019.
- [78] J. Hayashida, K. Nishimura, and R. Bise, "MPM: joint representation of motion and position map for cell tracking," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 3823–3832, Seattle, Online, USA, 2020.
- [79] X. Zhou, V. Koltun, and P. Krähenbühl, "Tracking objects as points," in *European Conference on Computer Vision*, pp. 474–490, Cham, 2020.
- [80] P. Chang, J. Grinband, B. D. Weinberg et al., "Deep-Learning convolutional neural networks accurately classify genetic mutations in gliomas," *American Journal of Neuroradiology*, vol. 39, no. 7, pp. 1201–1207, 2018.
- [81] N. Coudray, P. S. Ocampo, T. Sakellaropoulos et al., "Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning," *Nature Medicine*, vol. 24, no. 10, pp. 1559–1567, 2018.
- [82] M. Chen, B. Zhang, W. Topatana et al., "Classification and mutation prediction based on histopathology H&E images in liver cancer using deep learning," *NPJ Precision Oncology*, vol. 4, no. 14, pp. 1–7, 2020.
- [83] A. O'Rourke, S. Beyhan, Y. Choi et al., "Mechanism-of-action classification of antibiotics by global transcriptome profiling," *Antimicrobial Agents and Chemotherapy*, vol. 64, no. 3, pp. e01207–e01219, 2020.

- [84] E. L. Berg, J. Yang, and M. A. Polokoff, "Building predictive models for mechanism-of-action classification from phenotypic assay data sets," *SLAS Discovery*, vol. 18, no. 10, pp. 1260–1269, 2013.
- [85] C. Bakal, J. Aach, G. Church, and N. Perrimon, "Quantitative morphological signatures define local signaling networks regulating cell morphology," *Science*, vol. 316, no. 5832, pp. 1753–1756, 2007.
- [86] M. H. Rohban, S. Singh, X. Wu et al., "Systematic morphological profiling of human gene and allele function via Cell Painting," *Elife*, vol. 6, 2017.
- [87] V. Ljosa, P. D. Caie, R. ter Horst et al., "Comparison of methods for image-based profiling of cellular morphological responses to small-molecule treatment," *Journal of Biomolecular Screening*, vol. 18, no. 10, pp. 1321–1329, 2013.
- [88] J. Simm, G. Klambauer, A. Arany et al., "Repurposed high-throughput images enable biological activity prediction for drug discovery," *bioRxiv*, no. article 108399, 2017.
- [89] O. Z. Kraus, J. L. Ba, and B. J. Frey, "Classifying and segmenting microscopy images with deep multiple instance learning," *Bioinformatics*, vol. 32, no. 12, pp. i52–i59, 2016.
- [90] W. J. Godinez, I. Hossain, S. E. Lazic, J. W. Davies, and X. Zhang, "A multi-scale convolutional neural network for phenotyping high-content cellular images," *Bioinformatics*, vol. 33, no. 13, pp. 2010–2019, 2017.
- [91] W. J. Godinez, I. Hossain, and X. Zhang, "Unsupervised phenotypic analysis of cellular images with multi-scale convolutional neural networks," *BioRxiv*, no. article 361410, 2018.
- [92] J. C. Caicedo, C. McQuin, A. Goodman, S. Singh, and A. E. Carpenter, "Weakly supervised learning of single-cell feature embeddings," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 9309–9318, Salt Lake City, Utah, USA, 2018.
- [93] S. Spiegel, I. Hossain, C. Ball, and X. Zhang, "Metadata-guided visual representation learning for biomedical images," *BioRxiv*, no. article 725754, 2019.
- [94] E. Hoffer and N. Ailon, "Deep metric learning using triplet network," in *International Workshop on Similarity-Based Pattern Recognition*, pp. 84–92, Cham, 2015.
- [95] R. Janssens, X. Zhang, A. Kauffmann, A. de Weck, and E. Y. Durand, "Fully unsupervised deep mode of action learning for phenotyping high-content cellular images," *Bioinformatics*, vol. 37, no. 23, pp. 4548–4555, 2021.
- [96] M. Caron, P. Bojanowski, A. Joulin, and M. Douze, "Deep clustering for unsupervised learning of visual features," in *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 132–149, Munich, Germany, 2018.
- [97] P. D. Caie, R. E. Walls, A. Ingleston-Orme et al., "High-content phenotypic profiling of drug response signatures across distinct cancer cells," *Molecular Cancer Therapeutics*, vol. 9, no. 6, pp. 1913–1926, 2010.
- [98] V. Ljosa, K. L. Sokolnicki, and A. E. Carpenter, "Annotated high-throughput microscopy image sets for validation," *Nature Methods*, vol. 9, no. 7, p. 637, 2012.
- [99] S. Wang, M. Lu, N. Moshkov, J. C. Caicedo, and B. A. Plummer, "Anchoring to exemplars for training mixture-of-expert cell embeddings," 2021, <https://arxiv.org/abs/2112.03208>.
- [100] M.-A. Bray, S. Singh, H. Han et al., "Cell Painting, a high-content image-based assay for morphological profiling using multiplexed fluorescent dyes," *Nature Protocols*, vol. 11, no. 9, pp. 1757–1774, 2016.
- [101] I. Goodfellow, J. Pouget-Abadie, M. Mirza et al., "Generative adversarial nets," *Advances in Neural Information Processing Systems*, vol. 27, 2014.
- [102] P. Goldsborough, N. Pawlowski, J. C. Caicedo, S. Singh, and A. E. Carpenter, "CytoGAN: generative modeling of cell images," *BioRxiv*, no. article 227645, 2017.
- [103] A. X. Lu, O. Z. Kraus, S. Cooper, and A. M. Moses, "Learning unsupervised feature representations for single cell microscopy images with paired cell inpainting," *PLoS Computational Biology*, vol. 15, no. 9, article e1007348, 2019.
- [104] A. Razavi, A. den Oord, and O. Vinyals, "Generating diverse high-fidelity images with VQ-VAE-2," *Advances in Neural Information Processing Systems*, vol. 32, 2019.
- [105] H. Kobayashi, K. C. Cheveralls, M. D. Leonetti, and L. A. Royer, "Self-supervised deep learning encodes high-resolution features of protein subcellular localization," *Nature Methods*, vol. 19, no. 8, pp. 995–1003, 2022.
- [106] N. H. Cho, K. C. Cheveralls, A. D. Brunner et al., "Opencell: endogenous tagging for the cartography of human cellular organization," *Science*, vol. 375, no. 6585, article eabi6983, 2022.
- [107] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A simple framework for contrastive learning of visual representations," in *International Conference on Machine Learning*, PMLR, pp. 1597–1607, Vienna, Austria, 2020.
- [108] A. Perakis, A. Gorji, S. Jain, K. Chaitanya, S. Rizza, and E. Konukoglu, "Contrastive learning of single-cell phenotypic representations for treatment classification," in *International Workshop on Machine Learning in Medical Imaging*, pp. 565–575, Cham, 2021.
- [109] K. Yang, S. Goldman, W. Jin et al., "Mol2Image: improved conditional flow models for molecule to image synthesis," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 6688–6698, Nashville, Tennessee, USA, 2021.
- [110] F. Scarselli, M. Gori, A. C. Tsoi, M. Hagenbuchner, and G. Monfardini, "The graph neural network model," *IEEE Transactions on Neural Networks*, vol. 20, no. 1, pp. 61–80, 2008.
- [111] O. Wieder, S. Kohlbacher, M. Kuenemann et al., "A compact review of molecular property prediction with graph neural networks," *Drug Discovery Today: Technologies*, vol. 37, pp. 1–12, 2020.
- [112] Z. Hao, M. Gori, A. C. Tsoi, M. Hagenbuchner, and G. Monfardini, "ASGN: an active semi-supervised graph neural network for molecular property prediction," in *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 731–752, Virtual Event, CA, USA, 2020.
- [113] J. Gasteiger, F. Becker, and S. Günnemann, "Gemnet: universal directional graph neural networks for molecules," *Advances in Neural Information Processing Systems*, vol. 34, pp. 6790–6802, 2021.
- [114] K. Smith and P. Horvath, "Active learning strategies for phenotypic profiling of high-content screens," *Journal of Biomolecular Screening*, vol. 19, no. 5, pp. 685–695, 2014.
- [115] X. Lou, M. Schiegg, and F. A. Hamprecht, "Active structured learning for cell tracking: algorithm, framework, and usability," *IEEE Transactions on Medical Imaging*, vol. 33, no. 4, pp. 849–860, 2014.
- [116] A. W. Naik, J. D. Kangas, D. P. Sullivan, and R. F. Murphy, "Active machine learning-driven experimentation to

- determine compound effects on protein patterns,” *eLife*, vol. 5, article e10047, 2016.
- [117] S. Wen, T. M. Kurc, L. Hou et al., “Comparison of different classifiers with active learning to support quality control in nucleus segmentation in pathology images,” *AMIA Summits on Translational Science Proceedings*, vol. 2018, pp. 227–236, 2018.
- [118] J. Carse and S. McKenna, “Active learning for patch-based digital pathology using convolutional neural networks to reduce annotation costs,” in *European Congress on Digital Pathology*, Springer, Cham, 2019.
- [119] Z. Lai, C. Wang, L. C. Oliveira, B. N. Dugger, S.-C. Cheung, and C.-N. Chuah, “Joint semi-supervised and active learning for segmentation of gigapixel pathology images with cost-effective labeling,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 591–600, Nashville, Tennessee, USA, 2021.
- [120] F. Zhuang, Z. Qi, K. Duan et al., “A comprehensive survey on transfer learning,” *Proceedings of the IEEE*, vol. 109, no. 1, pp. 43–76, 2021.
- [121] C. Cai, S. Wang, Y. Xu et al., “Transfer learning for drug discovery,” *Journal of Medicinal Chemistry*, vol. 63, no. 16, pp. 8683–8694, 2020.
- [122] M. Majurski, P. Manescu, S. Padi et al., “Cell image segmentation using generative adversarial networks, transfer learning, and augmentations,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, Long Beach, CA, USA, 2019.
- [123] O. Z. Kraus, B. T. Gryb, J. Ba et al., “Automated analysis of high-content microscopy data with deep learning,” *Molecular Systems Biology*, vol. 13, no. 4, p. 924, 2017.
- [124] A. Kensert, P. J. Harrison, and O. Spjuth, “Transfer learning with deep convolutional neural networks for classifying cellular morphological changes,” *SLAS Discovery: Advancing Life Sciences R&D*, vol. 24, no. 4, pp. 466–475, 2019.
- [125] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, “How transferable are features in deep neural networks?,” *Advances in Neural Information PROCESSING Systems*, vol. 27, 2014.
- [126] S. Khan, N. Islam, Z. Jan, I. U. Din, and J. J. P. C. Rodrigues, “A novel deep learning based framework for the detection and classification of breast cancer using transfer learning,” *Pattern Recognition Letters*, vol. 125, pp. 1–6, 2019.
- [127] C. Szegedy, W. Liu, Y. Jia et al., “Going deeper with convolutions,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–9, Boston, MA, USA, 2015.
- [128] O. Russakovsky, J. Deng, H. Su et al., “ImageNet large scale visual recognition challenge,” *International Journal of Computer Vision*, vol. 115, no. 3, pp. 211–252, 2015.
- [129] N. Bayramoglu and J. Heikkilä, “Transfer learning for cell nuclei classification in histopathology images,” in *European Conference on Computer Vision*, pp. 532–539, Cham, 2016.
- [130] W. Zhang, R. Li, T. Zeng et al., “Deep model based transfer and multi-task learning for biological image analysis,” *IEEE Transactions on Big Data*, vol. 6, no. 2, pp. 322–333, 2020.
- [131] N. Pawlowski, J. C. Caicedo, S. Singh, A. E. Carpenter, and A. Storkey, “Automating morphological profiling with generic deep convolutional networks,” *BioRxiv*, p. 85118, 2016.
- [132] H. T. H. Phan, A. Kumar, J. Kim, and D. Feng, “Transfer learning of a convolutional neural network for HEP-2 cell image classification,” in *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*, pp. 1208–1211, Prague, Czech Republic, 2016.
- [133] R. Bermúdez-Chacón, P. Márquez-Neila, M. Salzmann, and P. Fua, “A domain-adaptive two-stream U-Net for electron microscopy image segmentation,” in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, pp. 400–404, Washington, DC, USA, 2018.
- [134] W. Dai, G.-R. Xue, Q. Yang, and Y. Yu, “Co-clustering based classification for out-of-domain documents,” in *Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 210–219, San Jose, California, USA, 2007.
- [135] B. Chen, W. Lam, I. Tsang, and T.-L. Wong, “Location and scatter matching for dataset shift in text mining,” in *2010 IEEE International Conference on Data Mining*, pp. 773–778, Sydney, NSW, Australia, 2010.
- [136] J. Shen, Y. Qu, W. Zhang, and Y. Yu, “Wasserstein distance guided representation learning for domain adaptation,” *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 32, 2008.
- [137] J. C. Boyd, A. Pinheiro, E. Del Nery, F. Reyat, and T. Walter, “Domain-invariant features for mechanism of action prediction in a multi-cell-line drug screen,” *Bioinformatics*, vol. 36, no. 5, pp. 1607–1613, 2020.
- [138] W. W. Qian, C. Xia, S. Venugopalan et al., “Batch equalization with a generative adversarial network,” *Bioinformatics*, vol. 36, Supplement\_2, pp. i875–i883, 2020.
- [139] R. Xiaoyang, X. Li, B. Roysam, and H. Nguyen, *Toward Zero Human Efforts: Iterative Training Framework for Noisy Segmentation Label*, ResearchGate, 2020.
- [140] C. Park, K. Lee, S. Y. Kim, F. S. C. Cecen, S.-K. Kwon, and W.-K. Jeong, “Neuron segmentation using incomplete and noisy labels via adaptive learning with structure priors,” in *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*, pp. 1466–1470, Nice, France, 2021.
- [141] L. Xiao, Y. Li, L. Qv, X. Tian, Y. Peng, and S. K. Zhou, “Pathological image segmentation with noisy labels,” 2021, <https://arxiv.org/abs/2104.02602>.
- [142] V. S. Akondi, V. Menon, J. Baudry, and J. Whittle, “Novel K-means clustering-based undersampling and feature selection for drug discovery applications,” in *2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pp. 2771–2778, San Diego, CA, USA, 2019.
- [143] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, “Focal loss for dense object detection,” in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2980–2988, Honolulu, Hawaii, USA, 2017.
- [144] K. Pasupa, S. Vatathanavaro, and S. Tungjitnob, “Convolutional neural networks based focal loss for class imbalance problem: a case study of canine red blood cells morphology classification,” *Journal of Ambient Intelligence and Humanized Computing*, vol. 11, pp. 1–17, 2020.
- [145] N. Yudistira, M. Kavitha, T. Itabashi, A. H. Iwane, and T. Kurita, “Prediction of sequential organelles localization under imbalance using a balanced deep U-Net,” *Scientific Reports*, vol. 10, no. 1, pp. 1–11, 2020.
- [146] Y. B. Hagos, C. S. Lecat, D. Patel et al., “Cell abundance aware deep learning for cell detection on highly imbalanced pathological data,” in *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*, pp. 1438–1442, Nice, France, 2021.

- [147] L. Guo, P. Huang, D. Huang et al., “A classification method to classify bone marrow cells with class imbalance problem,” *Biomedical Signal Processing and Control*, vol. 72, article 103296, 2022.
- [148] Y. Zhang, X.-S. Wei, B. Zhou, and J. Wu, “Bag of tricks for long-tailed visual recognition with deep convolutional neural networks,” *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, pp. 3447–3455, 2021.
- [149] B. Kang, S. Xie, M. Rohrbach et al., “Decoupling representation and classifier for long-tailed recognition,” 2019, <https://arxiv.org/abs/1910.09217>.
- [150] C. Blundell, J. Cornebise, K. Kavukcuoglu, and D. Wierstra, “Weight uncertainty in neural network,” in *International Conference on Machine Learning, PMLR*, pp. 1613–1622, Lille, France, 2015.
- [151] Y. Gal and Z. Ghahramani, “Dropout as a Bayesian approximation: representing model uncertainty in deep learning,” in *International Conference on Machine Learning, PMLR*, pp. 1050–1059, New York City, NY, USA, 2016.
- [152] G. Ghiasi, T.-Y. Lin, and Q. V. Le, “Dropblock: a regularization method for convolutional networks,” *Advances in Neural Information Processing Systems*, vol. 31, 2018.
- [153] L. Wan, M. Zeiler, S. Zhang, Y. Le Cun, and R. Fergus, “Regularization of neural networks using dropconnect,” in *International Conference on Machine Learning, PMLR*, pp. 1058–1066, Atlanta, GA, USA, 2013.
- [154] J. Tompson, R. Goroshin, A. Jain, Y. LeCun, and C. Bregler, “Efficient object localization using convolutional networks,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 648–656, Boston, MA, USA, 2015.
- [155] B. Lakshminarayanan, A. Pritzel, and C. Blundell, “Simple and scalable predictive uncertainty estimation using deep ensembles,” *Advances in neural information processing systems*, vol. 30, 2017.
- [156] A. P. Carrieri, W. P. Rowe, M. Winn, and E. O. Pyzer-Knapp, “A fast machine learning workflow for rapid phenotype prediction from whole shotgun metagenomes,” *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, pp. 9434–9439, 2019.
- [157] J. Chen, J. Hou, and K.-C. Wong, “Categorical matrix completion with active learning for high-throughput screening,” *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 18, no. 6, 2021.
- [158] A. Gomariz, T. Portenier, C. Nombela-Arrieta, and O. Goksel, “Probabilistic spatial analysis in quantitative microscopy with uncertainty-aware cell detection using deep Bayesian regression,” *Science Advances*, vol. 8, no. 5, article eabi8295, 2022.
- [159] O. Dürr, E. Murina, D. Siegmund, V. Tolkachev, S. Steigele, and B. Sick, “Know when you don't know: a robust deep learning approach in the presence of unknown phenotypes,” *Assay and Drug Development Technologies*, vol. 16, no. 6, pp. 343–349, 2018.
- [160] A. Theorell, J. Seiffarth, A. Grünberger, and K. Nöh, “When a single lineage is not enough: uncertainty-aware tracking for spatio-temporal live-cell image analysis,” *Bioinformatics*, vol. 35, no. 7, pp. 1221–1228, 2019.