Proceedings of ICBioMed 2025 Symposium: AI for Healthcare: Advanced Medical Data Analytics and Smart Rehabilitation
DOI: 10.54254/2753-8818/2025.AU28768

A Review of Deep Learning-Based Methods for Cell Image Recognition and Tracking

Qiancheng Zhou

Portland College, Nanjing University of Posts and Telecommunications, Nanjing, China
2084392260@qq.com

Abstract: By enabling precise quantification, cell image analysis advances understanding of cellular processes and structures. However, traditional methods have limitations in terms of accuracy and efficiency for dense target segmentation, continuous trajectory recognition, and large-scale data processing. Thus, this paper examines deep learning approaches for cell recognition and tracking, highlighting advancements in tracking across frames, automatic feature extraction, model adaptability, and their effectiveness to handle diverse and complex cellular environments. Through the review of relevant literature, including convolutional neural networks (CNNs), Mask R-CNN, HOG-SVM, as well as transfer learning methods, the potential applications in real-time processing, multimodal fusion, and high-throughput analysis are discussed. The results demonstrate that deep learning techniques enable precise segmentation, stable cross-frame tracking, and strong feature extraction in complex, dense cellular environments. Unlike traditional algorithms, deep learning methods notably reduce segmentation errors and tracking interruptions, all while maintaining solid generalization with minimal labeled data.

Keywords: Imaging system, Target detection, Cell segmentation, Convolutional neural network, Three-dimensional tracking

1. Introduction

Early cell image analysis used thresholding and spatiotemporal features. For example, PhagoSight applies improved Otsu thresholding and spatiotemporal features for 3D visualization of phagocytic cells [1]. However, this method remains limited in segmenting dense or overlapping regions and in accurately recognizing trajectories. Similarly, active contour models rely on clear boundaries, while traditional tracking methods suffer from low resolution and subjectivity, thereby making it difficult to capture complex dynamics. In contrast, deep learning methods such as HOG-SVM, R-CNN/Fast R-CNN, U-Net/SegNet, and Mask R-CNN with Kalman filtering can extract cell features, improve high-throughput accuracy, and adapt to different cell types and imaging conditions through transfer learning [2]. Thus, this study investigates the development of cell image recognition and tracking methods and critically evaluates their limitations, paying particular attention to the practical use of traditional segmentation and tracking techniques and to the advantages and shortcomings of deep learning approaches in complex settings. Through a literature review, it analyzes the features of various methods and assesses their adaptability across different data, computational, and application

© 2025 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2753-8818/2025.AU28768

contexts, highlighting the constraints of current techniques and outlining directions for subsequent methodological refinement and experimental planning.

2. Evolution and limitations of traditional cell image processing paradigms

2.1. Applicability and limitations of threshold segmentation methods

Image segmentation is a key task in traditional computer vision, once based on simple features such as gray, edge, and region. Threshold-based segmentation methods, valued for their simplicity and speed, were extensively applied in the early extraction and preprocessing of cell structures and are generally classified into three categories [3]. Firstly, gray-level thresholding methods estimate an optimal threshold from the image histogram to separate target and background. Typical methods include the Moment Preserving, Maximum Entropy, and Minimum Error Thresholding, with Otsu's Method maximizing inter-class variance to boost foreground-background separability.

$$\arg\max_T \sigma_b^2(T)=\omega_1(T)\cdot\omega_2(T)\cdot[\mu_1(T)-\mu_2(T)]^2 \tag{1}$$

where $\omega_1(T)$ and $\omega_2(T)$ represent the pixel proportions of the foreground and background under the current threshold, while $\mu_1(T)$ and $\mu_2(T)$ denote the average gray levels of the corresponding areas. This method achieves image binarization through statistical distribution optimization and has strong adaptability and stability.

Secondly, the local gray gradient-based edge detection method focuses on boundary positioning and, by combining fuzzy C-means clustering, extracts the target area, making it ideal for images with clear edges or high contrast. Thirdly, the region-based methods, such as region growing and watershed algorithms, segment through spatial adjacency and local similarities, preserving shape and handling connections. However, threshold-based methods still have obvious bottlenecks. Due to their strong dependence on parameters, they require presetting thresholds, structuring elements or seed points, and perform poorly in handling complex backgrounds, adhered cells and weakly edged targets. Moreover, these methods are sensitive to noise and brightness variations, often requiring manual correction, which limits their use in high-throughput automated cell image analysis.

2.2. Optimization and limitations of the active contour model

The Active Contour Model (ACM) combines three traditional segmentation strategies of gray-level statistics, local gradients and region detection, enhancing and unifying these methods. Despite its solid theoretical foundation, ACM faces several practical limitations that impact its segmentation performance and application range. Specifically, the energy functional is highly sensitive to model parameters. The standard energy expression is as follows.

$$E=\oint\{[E_{int}(X(s))+E_{ext}(X(s))]\}\,ds \tag{2}$$

The internal energy term constrains the curve's shape and is expressed as follows.

$$E_{int}=\alpha\left(s\right)|X'(s)|^2+\beta\left(s\right)|X''(s)|^2 \tag{3}$$

The parameters α(s) and β(s) respectively control the elasticity and rigidity of the contour curve. The former restricts the stretching of the curve, while the latter limits the bending of the curve,

DOI: 10.54254/2753-8818/2025.AU28768

thereby ensuring the smoothness and continuity of the segmentation boundary [4]. For different cell types and imaging conditions, these parameters need to be finely adjusted. Even a slight deviation can lead to over-segmentation or under-segmentation, thereby reducing the robustness of the model.

Furthermore, the design of external energy also affects the model's performance. It depends on image gradients and primarily operates in strong edge regions, making the model highly sensitive to the initial contour. If the initial contour deviates from the true boundary, the model may fall into local minima, resulting in missed detections or edge drift. Iterative methods, like gradient descent or Euler-Lagrange equations, are computationally expensive and time-consuming, especially for high-resolution or high-throughput images. The GVF-Snake method addresses this by introducing gradient vector flow, enabling the force to act both on and beyond the edges, reducing dependence on the initial contour and improving the capture range. Despite the combination of prior shape constraints and machine learning features for complex biological structures, ACM struggles with low signal-to-noise ratio images, hence leading to contour leakage, and static optimization makes it difficult to track dynamic cell changes. Thus, ACM is more suitable for single-cell or overlapping multi-cell detection, but not for dynamic or high-throughput analysis.

2.3. Performance and limitations of traditional tracking strategies

By analyzing dynamic behaviors such as migration, division, and differentiation, cell tracking offers essential spatiotemporal data for embryonic development, tumor metastasis, and cell therapy studies. Non-deep learning methods mainly include detection and association, model evolution, and filtering and sampling. However, they still have performance bottlenecks in complex movements and diverse morphologies [2].

Cell tracking methods generally involve segmenting individual cells, extracting key features like centroid, area, and shape, and associating these features across frames to construct cell trajectories. Typical methods for segmentation and feature matching include thresholding, watershed algorithm, and Euclidean distance-based techniques.These tracking methods heavily depend on segmentation quality, with issues such as under- or over-segmentation due to cell adhesion, and noise or uneven illumination exacerbating errors. Moreover, they struggle with topological changes. For example, during cell division or when cells enter/exit the field of view, matching can become ambiguous. In addition, manual feature design struggles to adapt to cells with diverse shapes. Cell boundaries in model evolution-based methods, such as active contour models or level sets, are described via an energy function or geometric model. These methods iteratively minimize the energy to track cells across frames. Though effective for cells with stable, smooth contours, these approaches face high computational costs, are sensitive to initial contours and parameters, and are susceptible to getting stuck in local minima. Filtering and sampling-based methods, such as Kalman or particle filtering, model cell movement and update predictions by combining observations. While they offer some robustness to noise and motion prediction, their assumptions, like the linear Gaussian motion in Kalman filtering, often mismatch the nonlinear behavior of living cells, leading to prediction errors. Particle filtering, though effective for complex tracking scenarios, comes with high computational demands and risks divergence in low signal-to-noise environments. Furthermore, these methods rely on prior models and initialization, limiting their effectiveness in fully automated systems.

DOI: 10.54254/2753-8818/2025.AU28768

3. The deep learning-driven transformation in cell recognition and tracking

3.1. Cell image recognition with convolutional neural networks

In handling complex backgrounds and overlapping targets, cell image recognition based on CNNs outperforms traditional methods. Nevertheless, cross-scene adaptability and resource usage remain issues, requiring a balance between accuracy and efficiency based on needs.

The target detection and multi-target tracking of cell image sequences are research hotspots. The key challenges are cell morphology diversity, environmental variability, and overlaps with complex backgrounds. Traditional methods like active contour models or methods based on shape/texture features have limitations in complex scenarios. For example, the 2013 PhagoSight algorithm used the improved Otsu method with spatiotemporal features for 3D segmentation of phagocytic cells, but struggled with accuracy in dense areas and required low-computing-power, simple-background environments [1]. Subsequently, the introduction of deep learning methods brought breakthroughs, with convolutional neural network frameworks greatly improving the accuracy of three-dimensional positioning and tracking. However, these frameworks require high signal-to-noise ratio, resolution, and temporal continuity [2]. In scenarios of medium complexity with limited labeled data, the HOG-SVM algorithm, paired with direction gradient histograms, remains effective but struggles in complex multi-target situations [5]. For high-complexity and high-resolution images, the end-to-end instance segmentation model combining Mask R-CNN and feature pyramid network accurately detects and analyzes cell nuclei in pathological images, but it demands substantial computing power [6]. Ans these highlight the advantages of convolutional neural networks in cell image recognition. However, existing methods are often tailored to specific goals or conditions, lacking generalization across scenarios, leading to notable accuracy drops in new environments. In practical applications, it is essential to customize recognition strategies based on the morphological characteristics of the research object, imaging conditions, and task requirements to achieve a balance between accuracy and computational resources.

3.2. Deep learning-based algorithm for tracking cell movement trajectories

The dynamic analysis of cell movement depends on frame-by-frame identification and cross-frame association of static targets. The deep learning model outputs the cell's coordinates, area, brightness, and identifier, supporting trajectory construction by connecting discrete time points into continuous trajectories. Accordingly, the algorithm introduces three constraints: spatial position, morphological stability, and neighborhood structure [7]. To limit cell movement, the algorithm imposes a spatial constraint on the center coordinates of adjacent frame targets. Specifically, the center distance of a cell between two consecutive frames, V_i and V_{i+1} must satisfy the following threshold constraint:

$$\mathrm{Dist}(V_i, V_{i+1}) \leq d \tag{4}$$

where V_i is the center coordinates of a certain cell in the i-th frame. The threshold d is typically set to 0.8 times the average diameter to exclude unreasonable jump matching, while morphological constraints limit area changes to 15% between consecutive frames, improving pairing stability and preventing abnormal matching due to segmentation errors or cell adhesion. Besides, neighborhood structure constraints define neighboring cells (with an Euclidean distance within twice the diameter) and prioritize targets with stable neighbor counts, thereby enhancing trajectory consistency.

DOI: 10.54254/2753-8818/2025.AU28768

Based on the above rules, the system first constructs a spatial and directional feature set for each frame, and completes the initial pairing between adjacent frames. For targets that fail to be matched, the threshold is moderately relaxed and the matching is attempted again to fix the omissions caused by blurred boundaries or slight deformations. In the case of area abnormalities caused by adhesion, the large target is further divided into several sub-regions, and the secondary pairing is completed by combining the local extreme point information. At the temporal level, the trajectory angle θ is used for dynamic judgment. If the absolute value of the direction change $|\Delta\theta|$ in three consecutive frames is less than 30°, it is considered that the trajectory is continuous; otherwise, the abnormal connection is corrected through the multi-frame backtracking mechanism. This method combines spatial, morphological and neighborhood constraints as well as motion direction judgment, ensuring the integrity of the trajectory while controlling the false matching rate below 2%. Relying on the collaborative optimization of deep features and geometric rules, the algorithm has good robustness and is suitable for cell behavior tracking in various experimental scenarios.

3.3. Cross-sample generalization via transfer learning techniques

When training data is limited, common strategies to improve model generalization include transfer learning and data augmentation. Data augmentation enhances sample diversity, helping the model focus on common features. Furthermore, transfer learning leverages existing knowledge for quick adaptation, thereby boosting performance across different datasets. Specifically, data augmentation enlarges the training dataset by applying operations such as cropping, rotating, flipping, and scaling, simulating how cells appear from various angles, positions, scales, or orientations. For instance, in microscopic images, individual cells may shift due to imaging angles or field of view. By allowing the model to "experience" a wider variety of arrangements, augmentation improves its robustness in recognizing new samples. This method increases data diversity, reduces reliance on specific sample features, and helps prevent overfitting, thus addressing the issue of limited labeled data. In the case of transfer learning, pre-trained models are typically applied to related tasks to reduce training time and improve accuracy. For example, a ResNet model trained on large-scale natural images can be transferred to cell image classification tasks and fine-tuned to adapt to the new data characteristics. Similarly, Duari et al. used the scCamAge model, trained on yeast, and performed transfer learning on limited human fibroblast aging data. By fine-tuning the model with a small number of labeled samples, they achieved quick adaptation to new cell types while also maintaining high prediction accuracy and trajectory continuity [8]. These methods indicate that data augmentation and transfer learning provide significant advantages in improving a model's generalization ability, particularly when labeled data is limited. With these techniques, more efficient and accurate cell image analysis can be achieved across different experimental scenarios.

4. Future pathways for cell recognition and tracking

4.1. Construction of standardized cell image datasets

Though machine learning-based segmentation algorithms have made progress, these methods often depend on large amounts of labeled samples as training data [9]. At present, there is still a limited availability of publicly available real cell image datasets, which restricts the performance evaluation and further optimization of algorithms. Most existing studies rely on computer-simulated synthetic datasets for model pretraining. Despite the ability of synthetic data to partially alleviate issues such as labeling costs and data scarcity, they differ from real biological images in terms of imaging

DOI: 10.54254/2753-8818/2025.AU28768

principles, resolution, texture complexity, cell morphology, and noise distribution. This results in limited transferability of models trained on synthetic data to real-world data. Experiments show that deep learning models achieve high accuracy on synthetic data but tend to overfit in real scenarios, impairing their generalization ability. Moreover, subcellular structure recognition and cell tracking, as key tasks, currently lack a widely accepted standard dataset. Inconsistent evaluation benchmarks and annotation standards make it difficult to directly compare results from different studies. Thus, it is essential to boost multi-institution collaboration, facilitate data sharing, and enable collaborative annotation through open platforms, while improving annotation efficiency and consistency by using semi-automated tools. This will help create a diversified standard dataset covering different cell types, imaging modes, and experimental conditions. Additionally, establishing a unified algorithm performance evaluation system is equally important. Evaluation metrics should cover segmentation accuracy, tracking accuracy, computational efficiency, and model generalization ability to ensure a comprehensive assessment of algorithm performance. This helps standardize algorithm comparisons and accelerates the translation of cell image recognition and tracking technologies to clinical and biomedical applications.

4.2. Real-time online processing with intelligent microscopy

As cell image recognition and tracking technologies advance, integrating deep learning algorithms into microscope hardware for real-time processing has become a key priority for improving the efficiency and precision of biomedical research. Traditional methods typically depend on offline processing after image acquisition, resulting in extended processing times and delayed feedback. This limits real-time monitoring and hinders quick responses to dynamic cell behaviors. Achieving real-time online processing presents several technical challenges.

First, the algorithm must have efficient computing capabilities to perform image segmentation, target recognition, and tracking tasks quickly within the constraints of limited hardware resources. Solutions commonly employed include optimized convolutional neural networks, GPU acceleration, model pruning, and quantization, among other lightweight technologies. Second, the hardware and software of the microscope must be deeply integrated to ensure high bandwidth, low latency for data transmission, and synchronization between acquisition and processing. Besides, the system should be robust enough to adapt to various imaging conditions and handle complex changes in cell morphology. In practical applications, the intelligent microscope's real-time processing system can immediately provide feedback on dynamic cell behaviors, notably shortening the data processing cycle. This system supports disease diagnosis, drug screening, and basic research. For example, it enables real-time monitoring of cell migration, division, and drug reactions, hence revealing disease mechanisms and aiding the screening of effective treatments. Prior research has utilized lightweight deep learning models on embedded devices to enable real-time image analysis with microscopes, showcasing strong application potential [10]. In the future, as hardware performance improves and algorithms are further refined, intelligent microscope systems are expected to achieve high accuracy, speed, and integration, pushing forward real-time, accurate research in cell biology and medicine.

4.3. Exploration of multimodal imaging fusion techniques

By integrating various imaging approaches, multimodal imaging fusion enables the synchronous acquisition and dynamic monitoring of multi-dimensional cellular information, thus advancing cell biology and medical research with greater comprehensiveness and precision [11]. Common imaging techniques include optical microscopes, fluorescence imaging, short-wave infrared imaging (SWIR),

DOI: 10.54254/2753-8818/2025.AU28768

ultrasound, CT, and MRI. SWIR has high penetration and low scattering properties, allowing it to provide clear signals in deeper tissues. When paired with other techniques, it can overcome the limits of one method.

Multimodal fusion typically involves three steps. Firstly, in the technology combination stage, different imaging methods are used together, with their respective advantages complementing each other in terms of spatial resolution, penetration depth, and molecular specificity, thereby providing multi-level information on cell morphology, metabolic state, and molecular composition. Secondly, in the data alignment stage, signal synchronization, preprocessing, and feature extraction ensure consistency in space, time, and scale across images from different modalities. Additionally, in the information integration stage, a joint analysis algorithm is used to combine the information from various modalities, thus enabling multi-dimensional dynamic monitoring and analysis. In practical applications, multimodal fusion technology can track multiple biological processes within cells in real time, offering multi-dimensional evidence to reveal functional mechanisms and pathological changes. However, traditional methods have limitations in preserving the texture and features of the source image, leading to a decline in visual clarity and quantitative indicators such as structural similarity and noise robustness. To address this issue, the DMF-LP model adopts a dual-innovation design [12]. The LP-F pre-fusion module processes the source image before diffusion, enhancing texture features such as tissue edges and microvessels to avoid detail loss; the information entropy loss function quantifies information use, prioritizing high-entropy regions like the lesion-normal tissue boundary, improving pathological recognition and clarity in the fused image.

5. Conclusion

The results show that deep learning-driven cell image recognition and tracking technologies have overcome the performance limitations of traditional methods in dense cell scenarios. The core breakthrough lies in the ability to achieve high-precision modeling of complex shapes and dynamic behaviors through end-to-end feature learning, providing a reliable tool for cell dynamics research and clinical pathological analysis. However, this field still faces several key challenges, primarily due to the reliance on real data. Insufficient real-time processing capabilities limit its generalization ability, and the lack of real-time processing makes it difficult to apply in surgical settings. Besides, the absence of evaluation standards complicates algorithm comparisons. Future research should focus on building multimodal fusion frameworks, developing embedded lightweight systems, and establishing open-source collaborative ecosystems. These efforts will improve dynamic tracking, integrate imaging and analysis, and standardize evaluation.

References

[1] Henry, K.M., Pase, L., Ramos-Lopez, C.F., Lieschke, G.J., Renshaw, S.A., & Reyes-Aldasoro, C.C. (2013). PhagoSight: An open-source MATLAB® package for the analysis of fluorescent neutrophil and macrophage migration in a zebrafish model. PLOS ONE, 8(8), e72636.

[2] Liu, Y., Smith, J., Chen, X., & Johnson, M. (2018). MALDI-MSI of immunotherapy: Mapping the EGFR-targeting antibody cetuximab in 3D colon-cancer cell cultures. Analytical Chemistry, 90(24), 14156-14164.

[3] Luo, W.C., & Guo, W.B. (2000). A comparison and analysis of image thresholding segmentation methods. Modern Computer, (11), 22-25.

[4] Zhang, C., Tang, K.L., Zhang, H.Q, & Pan, S.H. (2021). A new image segmentation method based on active contour model. Journal of Chengdu University (Natural Science Edition), 40(01), 48-51.

[5] Liu, XY., & Wang, Y.M. (2019). Target tracking method based on the improved tracking-learning-detection algorithm using HOG-SVM. Science Technology and Engineering, 19(27), 266-271.

DOI: 10.54254/2753-8818/2025.AU28768

[6] Ramakrishnan, V., et al. (2024). Nuclei detection and segmentation of histopathological images using a feature pyramidal network variant of a Mask R-CNN. Bioengineering, 11(10), 994.
[7] Duari, S., Gautam, V., & Ahuja, G. (2025). Protocol for cellular age prediction in yeast and human single cells using transfer learning. STAR Protocols, 6(3), 104023.
[8] Sun, Y. (2022). Research on the identification of suspended cells and multi-cell tracking in microscopic images. Beijing University of Chemical Technology.
[9] Aleynick, N., Li, Y., Xie, Y., et al. (2023). Cross-platform dataset of multiplex fluorescent cellular object image annotations. Scientific Data, 10(1), 193.
[10] Doroshenko, O.V., et al. (2024). Automated assessment of wheat leaf disease spore concentration using a smart microscopy scanning system. Agronomy, 14(9), 1945.
[11] Wang, Q., Zhao, W., Pang, H., et al. (2025). Discussion on the value of multimodal image fusion technology. Advanced Journal of Nursing, 6(2).
[12] Wu, D., Tan, X., Wang, H., et al. (2025). DMF-LP: Enhancing image quality through multimodal medical image fusion using diffusion and Laplacian techniques. Biomedical Signal Processing and Control, 106, 107890.