DL-CSPF: deep-learning-based cell segmentation with a physical framework for digital holographic microscopy

Zhuoshi Li,[1,2,3,†] **Haojie Gu,**[1,2,3,†] **Linpeng Lu,**[1,2,3] **Qian Shen,**[1,2,3] **Jiasong Sun,**[1,2,3] **Qian Chen,**[3] **and Chao Zuo**[1,2,3,*]

[1]*Smart Computational Imaging Laboratory (SCILab), Nanjing University of Science and Technology, Nanjing, Jiangsu 210094, China*
[2]*Smart Computational Imaging Research Institute (SCIRI) of Nanjing University of Science and Technology, Nanjing, Jiangsu 210019, China*
[3]*Jiangsu Key Laboratory of Spectral Imaging & Intelligent Sense, Nanjing, Jiangsu 210094, China*
[†]*The authors contributed equally to this work.*
*zuochao@njust.edu.cn

Digital holographic microscopy (DHM) offers label-free, full-field imaging of live-cell samples by capturing optical path differences to produce quantitative phase images. Accurate cell segmentation from phase images is crucial for long-term quantitative analysis. However, complicated cellular states (e.g., cell adhesion, proliferation, and apoptosis) and imaging conditions (e.g., noise and magnification) pose significant challenge to the accuracy of cell segmentation. Here, we introduce DL-CSPF, a deep-learning-based cell segmentation method with a physical framework designed for high-precision live-cell analysis. DL-CSPF utilizes two neural networks for foreground-background segmentation and cell detection, generating foreground edges and "seed points." These features serve as input for a marker-controlled watershed algorithm to segment cells. By focusing on foreground edges and "seed points", which have lower information entropy than complete cell contours, DL-CSPF achieves accurate segmentation with a reduced dataset and without manual parameter tuning. We validated the feasibility and generalization of DL-CSPF using various open-source and DHM-collected datasets, including HeLa, pollen, and COS-7 cells. Long-term live-cell imaging results further demonstrate that DL-CSPF reliably characterized and quantitatively analyzed the morphological metrics across the cellular lifecycle, rendering it a promising tool for biomedical research. © 2024 Optica Publishing Group.

1. INTRODUCTION

In biomedical images analysis, cell segmentation is a fundamental task that is universally required across a diversity of experimental setups and imaging techniques [1–5]. Accurate cell segmentation is a crucial prerequisite for cellular quantitative analysis, holding significant implications for the investigations of cell kinetics [6], cell proliferation [7], cancer treatment [8], and drug release monitoring *in vitro* [9]. However, complex cell distribution states such as cell adhesion and growth states including proliferation and apoptosis typically present significant challenges for segmentation. Moreover, for cell samples with weak absorption features, the amplitude of the light does not change when passing through samples but a substantial phase delay is introduced, which is difficult for human eyes and photodetectors to capture. Fluorescence microscopy uses fluorescent molecules to label target proteins in cells to provide specificity for imaging, yet invasive imaging is incapable of enabling long-term observation of live cells [10–15].

Quantitative phase imaging (QPI) utilizes the RI as an endogenous contrast agent to generate cell- or subcellular-specific quantitative maps, achieving label-free non-destructive live-cell imaging [16–24]. As a classical QPI technique, digital holographic microscopy (DHM) interferometrically encodes the complex field information into intensity modulation (i.e., interferogram or hologram), and performs the quantitative analysis of wave-matter interactions by decoding phase delay [25–29]. However, due to high-coherence sources typically used in DHM, the imaging quality is plagued by speckle noise, thereby enhancing the difficulty of segmentation. Therefore, stable and accurate cell segmentation on phase images under complex conditions (e.g., cell adhesion, proliferation/apoptosis, and noise) is pivotal for long-term dynamic live-cell analysis with DHM, which is of significant importance for the application of QPI technology in biomedicine.

Traditional approaches to cell segmentation rely on manually crafted feature definitions that allow the algorithmic recognition

of cellular regions and borders. These algorithms can be systematically classified into the following categories: threshold-based methods (e.g., Otsu [30] and maximum entropy algorithms [31]), edge-based methods (Canny edge detection [32], Laplacian of Gaussian [33], Sobel and Prewitt operators [34], etc.), region-based methods (such as region growing [35], region splitting/merging, and watershed algorithms [36]), and clustering and graph-based methods (e.g., K-means clustering [37]). Unfortunately, manual feature definitions are usually highly context-specific and require task-dependent and experience-dependent parameters tuning to work well [38,39]. Switching to a different cell type (with different morphological features) or confronting diverse cellular states (cell adhesion, cell proliferation, and apoptosis, etc.) typically requires a redesign and/or reoptimization of the segmentation algorithm, as well as fairly tedious parameters tuning.

In recent years, deep learning (DL) techniques for computer vision are increasingly being utilized for a variety of tasks in biological image analysis, providing a new paradigm for cell segmentation tasks [40–47]. Unlike the traditional "physics-based" approach, DL-enabled segmentation methods are a kind of "data-driven" approach, which is developed from a segmentation scheme based on a fully convolutional network (e.g., U-Net [48], mask R-CNN [49], and U2-Net [50]), to the methods incorporating attention mechanisms [51,52]. Furthermore, an instance segmentation approach such as YOLACT [53] demonstrates decent real-time segmentation performance. The Segment Anything Model [54,55], as an advanced image segmentation framework based on vision Transformers and prompt-driven mechanisms, exhibits high generalization capabilities and flexible interactive segmentation functionality. Alternatively, Cellpose is a Transformer-based DL segmentation method, featuring a universal cross-modality cell segmentation framework and a graphical user interface (GUI) [56,57]. Additionally, users can enhance generalization by adopting the "human-in-the-loop" training approach with their own datasets. Notwithstanding, these methods all depend on extensive, high-quality, labeled paired datasets that are prohibitively laborious and time-intensive to acquire due to the heterogeneity and quality variability of biological images, as well as complex cellular states that make this process more challenging.

In this study, we propose a novel DL-based cell segmentation method with a physical framework (DL-CSPF), which introduces physical framework constraints into a conventional end-to-end DL model for high-accuracy cell segmentation in diverse situations, such as cell adhesion, proliferation, and apoptosis. Applied to DHM, DL-CSPF accurately segments and recognizes cells from phase images, thereby facilitating long-term, high-accuracy dynamic live-cell analysis. In DL-CSPF, inspired by physics-informed DL models, the introduction of a physical framework enables convolutional neural networks (CNNs) to conduct accurate segmentation with modestly sized datasets. Unlike off-the-shelf DL-based methods that directly extract effective information to learn features of individual cells, DL-CSPF employs two pre-trained networks for foreground-background segmentation and cell detection ("seed-point" extraction) from phase images. Subsequently, it utilizes the foreground as edge constraints and the "seed points" as feature guidance to achieve accurate segmentation through the

marker-controlled watershed algorithm (MCW) [58]. Due to the degradation of the learning target, which reduces the information entropy of the dataset required for training and thereby lowers the learning difficulty, only a small fraction of the dataset is sufficient to achieve accurate cell segmentation, and without parameter tuning. We conducted the experiment to validate the feasibility and generalization of DL-CSPF using open-source datasets and real datasets collected by DHM for Henrietta Lacks (HeLa) human cervical cancer cells, pollen cells, and COS-7 cells. Additionally, DL-CSPF was employed to segment and recognize HeLa cells over up to 9 h, accurately characterizing their area, dry mass (DM), and dry mass surface density (DMSD), and conducting quantitative morphological analysis of single-cell division behavior.

2. DEEP-LEARNING-BASED CELL SEGMENTATION METHOD WITH A PHYSICAL FRAMEWORK

A. Overview of Method

In DL-CSPF, the whole cell segmentation can be divided into three processes: collecting phase images by DHM to construct a well-matched paired dataset, designing DL models for foreground-background segmentation and cell detection, achieving accurate cell segmentation by using the obtained foreground boundaries and "seed points" as prior constraints for the MSW algorithm. Specifically, the foreground boundary represents the dividing line between the segmented object (cells) and the image background, without considering distinctions between individual cells. Meanwhile, "seed points" denote the recognition points (centroid points) of each cell. The specific process is shown in Fig. 1. In DHM, the object wave $O(x, y)$, which records the sample complex amplitude information, interferes with the reference wave $R(x, y)$, generating a hologram $I_H(x, y) = |O|^2 + |R|^2 + O^*R + R^*O$. Taking the Fourier transform (FT) on the hologram, the +1-order term (R^*O) can be obtained by band-pass filtering, and the phase information can be recovered through inverse Fourier transform (IFT), phase aberration correction [59], and unwrapping algorithms [60]. We collected the ground truth (GT) required for training by employing methods such as empirical gradient threshold (EGT) [61], distance transform (DT) [62], *Cellpose*, and *ImageJ* on the QPI results recovered by DHM, and the specific implementation is detailed in the next section.

In the original design of the physical framework, foreground-background segmentation and cell detection both depend on manually crafted feature recognition algorithms, e.g., EGT, DT, Otsu, and Laplacian of Gaussian (LoG) [63]. However, these algorithms generally require multi-parameters tuning, which relies on the experience of the user and is susceptible to diverse cellular types and states. Moreover, due to high-coherence sources typically used in DHM, the phase imaging quality is plagued by speckle noise caused by stray interferences from system imperfections, thereby impacting the robustness of the segmentation algorithm. Herein, we design the foreground-background segmentation model (FSNet) and cell detection model (CDNet) for foreground edges and "seed-point" extraction, as illustrated in Figs. 1(a)–1(c). DL-based
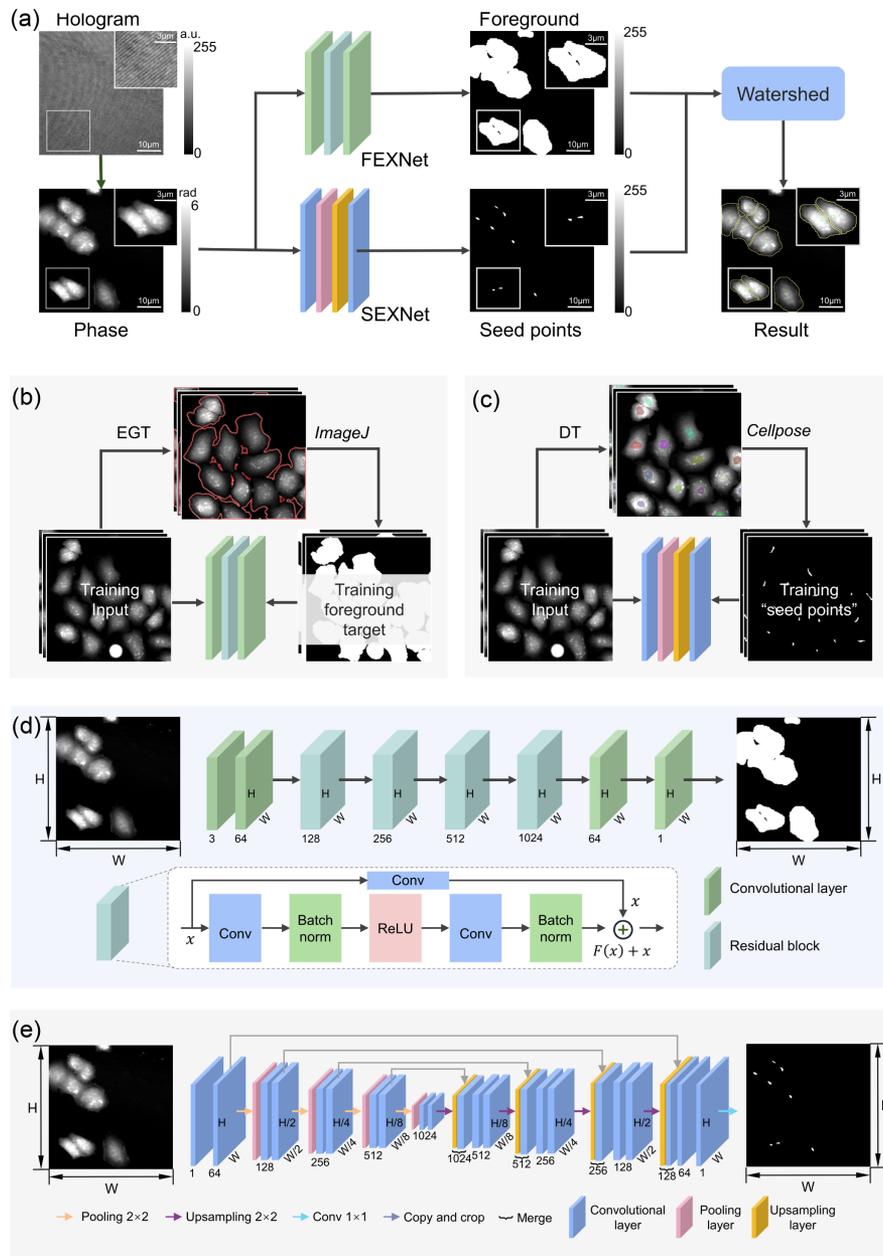
Fig. 1. Deep-learning-based cell segmentation method with a physical framework. (a) Overview of DL-CSPF. (b) Process of training for foreground-background segmentation task. (c) Process of training for cell detection task. (d) Overview of the FSNet architecture. (e) Overview of the CDNet architecture.

methods, through hierarchical nonlinear transformations, can automatically learn rich feature representations from data, encompassing multiple layers of features that range from low-level (e.g., edges, textures) to high-level (e.g., object parts, semantic information). The incorporation of the DL technique facilitates enhanced generalization in handling complex cellular states and types, even in the presence of noise, and eliminates the necessity for parameter tuning. DL-CSPF acquires foreground borders and "seed points" using properly trained FSNet and CDNet as constraints and guidance information, in which "seed points" $M = \{m_1, m_2, \ldots, m_k\}$ (corresponding to unique marker point L_i) can steer the MCW algorithm to segment adhesive cells from each other. Specifically, for each pixel P_i,

the topological distance to each marker point L_i is calculated, and the pixel is assigned to the marker point with the smallest topological distance, thereby forming the watershed lines. These watershed lines, in conjunction with the constraints provided by foreground borders, can isolate each pair of adhesive cells, facilitating accurate cell segmentation.

B. Foreground-Background Segmentation Model

In FSNet, we implement the foreground-background segmentation task by using a residual network (ResNet) structure [64]. As an effective scheme to address the degradation and gradient vanishing problems, ResNet introduces a deep residual

learning framework into CNN. Instead of expecting each group of stacked layers to directly fit a desired underlying mapping, ResNet explicitly enables these layers to fit a residual mapping. As shown in Fig. 1(d), the stacked nonlinear layers fit another mapping of $F(x) := H(x) - x$, in which $H(x)$ denotes the desired underlying mapping, and the original mapping is recast into $F(x) + x$. Unlike conventional cell segmentation tasks, foreground segmentation does not require focusing on the detailed boundaries between cells. Instead, it emphasizes the distinction between the object and the background. Therefore, FSNet based on the ResNet architecture is capable of recognizing smooth and accurate edges features, demonstrating strong generalization performance.

Specifically, in FSNet, we initially resize the phase images to $H \times W \times 64$ for the initial layer, where H and W are the height and width of the input images, respectively. This layer then passes through a set of residual blocks, with the dimensions increasing to $H \times W \times 128$, $H \times W \times 256$, and finally $H \times W \times 512$ after each set. The output is subsequently adjusted to $H \times W \times 1$ by a global average pooling, yielding a mask that delineates the cellular regions within the phase image. The model predicts a probability map for each pixel being part of the cell membrane area. Furthermore, the rectified linear unit (ReLU), as a commonly used activation function, introduces nonlinearity into CNN to mitigate the vanishing gradient problem. To effectively refine the loss function, we deploy the Adam optimizer, an algorithm renowned for its computational efficiency and robustness in converging towards the minimum of the loss function. By harnessing the power of backpropagation, we meticulously adjust the internal parameters of the FSNET through successive iterations. As a key factor that directly influences the model's optimization process and final performance, we adopt mean squared error loss (MSE-Loss) as the loss function for FSNet:

$$L_{\text{MSE}} = \frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2, \tag{1}$$

where y_i denotes the GT of the ith sample, and $\hat{y}_i$ is the predicted value of the ith sample.

C. Cell Detection Model

In cell detection, the centroid region of the cell, referred to as the "seed points", serves as the recognized target to acquire cellular distribution positions, as shown in Fig. 1(c). Accurate cell detection is a crucial prerequisite for addressing complicated cell states, such as cell adhesion and cell proliferation/apoptosis. Therefore, we design a cell detection model (CDNet) based on the U-Net, which enables the integration of high-resolution features from the encoder with the low-resolution features in the decoder; it effectively handles small objects and details in images. As a fully convolutional network, U-Net supplements a usual contracting network by successive layers, where pooling operators are replaced by upsampling operators. Moreover, in the upsampling part, U-Net also added a large number of feature channels, which allows the network to propagate context information to higher-resolution layers.

As shown in Fig. 1(e), the CDNet consists of a contracting path (left side) and an expansive path (right side), in which both of them follow the typical CNN architecture. Specifically, the

contracting path consists of the repeated application of two 3×3 convolutions (unpadded convolutions), each followed by a ReLU and a 2×2 max pooling operation with stride for downsampling. On the other hand, each step within the expansive path involves upsampling the feature map, applying a 2×2 convolution ("up-convolution") to halve the number of feature channels, concatenating with the correspondingly cropped feature map from the contracting path, and performing two 3×3 convolutions, each followed by a ReLU. In the cell detection task, we initially resize the phase images to $H \times W \times 64$, followed by the images of $H/16 \times W/16 \times 1024$ obtained through four rounds of downsampling, which involves consecutive convolutions and pooling operations to capture contextual information. Thereafter, the network performs four rounds of upsampling to achieve images of $H \times W \times 64$, which entails a gradual restoration of the spatial resolution of the feature maps through upsampling and convolutional operations, enabling precise localization. Between downsampling and upsampling, skip connections are utilized to fuse the feature maps, allowing the decoder to integrate low-level local features with high-level semantic information. Unlike the foreground-background segmentation task, CDNet utilizes the binary cross-entropy loss (BCE-Loss) [65] as a loss function for training:

$$L_{\text{BCE}} = -\frac{1}{N} \sum_{i=1}^{N} \left[y_i \log(p_i) + (1 - y_i) \log(1 - p_i) \right], \tag{2}$$

where y_i is the GT of the ith pixel, p_i denotes the predictive probability of the ith pixel, and N is the total number of pixels.

3. EXPERIMENT AND RESULTS

A. Experiment and Dataset Acquisition

In this section, we conduct the experiment on an open-source dataset [66] and a dataset constructed using DHM to verify the feasibility of the proposed method. We employed an off-the-shelf digital holographic microscope (DH-SCLM) developed by SCILab [67], on which a $20\times$, 0.5 NA objective lens and a CCD camera with 4.4 µm pixel size were equipped. A laser with a central wavelength of 532 nm was employed in DH-SCLM to capture holograms for QPI. We collected 710 holograms and reconstructed the corresponding phase images, and these data were further partitioned into training, validation, and test sets at ratios of 85%, 5%, and 10%, respectively. In the foreground-background segmentation stage, we initially adopted the EGT algorithm on the phase images to obtain preliminary results of foreground segmentation, followed by utilizing *ImageJ* for manual inspection and correction of any omissions or errors. Alternatively, for cell detection, we used the DT algorithm to preliminarily extract the "seed points" of the cells from the phase images. Subsequently, we used the *Cellpose2.0* to further supplement and revise omitted points and errors. The experiment was conducted on a workstation equipped with an Intel i9-10900K 3.70 GHz CPU and an NVIDIA GeForce RTX 3090 GPU. The proposed method was operated by Python 3.8 and PyTorch 1.12.1. During the training process, FSNet and CDNet were trained for 10 epochs and 25 epochs, respectively, taking 2 h and 4.5 h.
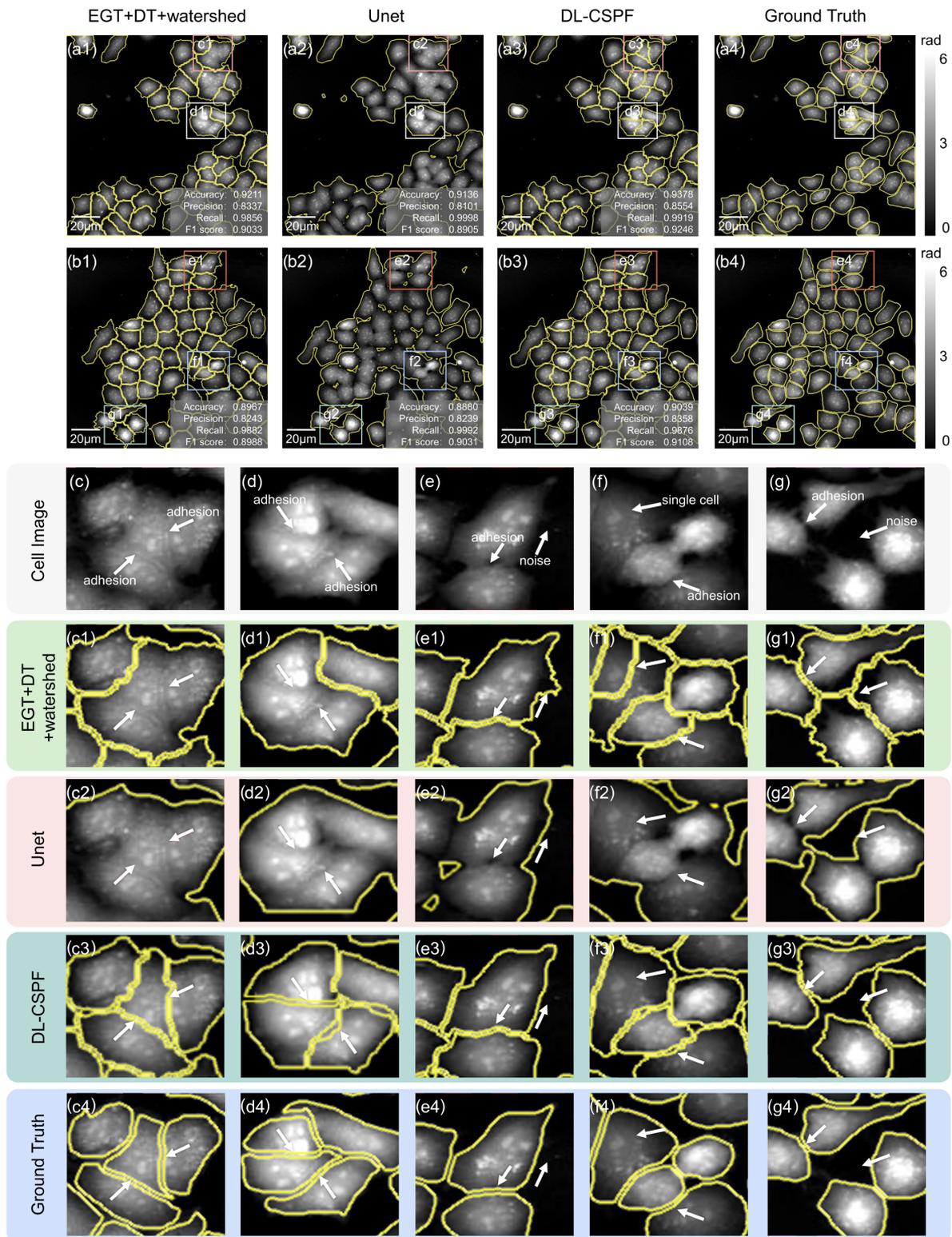
Fig. 2. Cell segmentation results and close-up views on an open-source dataset. (a1)–(a3), (b1)–(b3) Segmentation results on two sets of test samples using three different methods ("EGT + DT + watershed", U-Net, and DL-CSPF), respectively. (a4), (b4) Ground truth. (c)–(g) Original cell images, with red arrows emphasizing the complex scenarios encountered in segmentation, including cell adhesion, single-cell, and noise. (c1)–(g1), (c2)–(g2) Enlarged views of cell segmentation results by "EGT + DT + watershed" and U-Net exhibit rough edges, and some cells are either over-segmented or under-segmented. (c3)–(g3) The corresponding results based on DL-CSPF demonstrate accurate cell segmentation that is in good agreement with the GT, (c4)–(g4).

B. Segmentation Results and Discussion

Figure 2 showcases a comparison of cell segmentation results and close-up views on the open-source dataset using the conventional method ("EGT + DT + watershed") [63], the DL-based method (U-Net), and DL-CSPF. Among the various conventional segmentation schemes, the method that adopts EGT to segment foreground and background, extracting "seed points" using DT for cell detection, and applying the MCW to achieve cell segmentation, demonstrates better generalizability and segmentation accuracy, despite the need for multi-parameters tunings. As a typical DL-based cell segmentation method, U-Net demonstrates decent segmentation performance and representativeness. Therefore, we adopted "EGT + DT + watershed" and U-Net as subjects for comparative experiments. Figures 2(a1) and 2(b1) show the segmentation results using the "EGT + DT + watershed" method for two typical cell specimens selected from the test set; alternatively, the enlarged views are shown in Figs. 2(c1)–2(g1). It can be found that the results exhibit rough segmentation edges [Figs. 2(e1) and 2(g1)], with some cells being over-segmented [Fig. 2(f1)] or under-segmented [Figs. 2(c1)–2(d1)], which is attributed to the algorithm's limited recognition capabilities. The corresponding results based on the U-Net method and the zoomed-in views are showcased in Figs. 2(a2)–2(b2) and 2(c2)–2(g2), from which the cell adhesion issues cannot be resolved. The under-segmentation issue is due to the inadequate feature recognition of U-Net driven by a limited-sized dataset. In DL-CSPF, the introduction of a physical framework decomposes image features, thereby reducing information entropy, and sustains excellent cell segmentation performance when trained on a modestly sized dataset. The results are shown in Figs. 2(a3)–2(b3) and 2(c3)–2(g3), which are in good agreement with the GT [Figs. 2(c4)–2(g4)]. Despite challenges such as cell adhesion and noise, DL-CSPF can still achieve accurate cell segmentation, producing smoother segmentation edges and demonstrating more resilience to changes in image quality.

To validate the feasibility and generalization of DL-CSPF for real datasets with the complex live-cell samples, we implement the experiment on phase images obtained by DHM for HeLa cells, pollen cells, and COS-7 cells, as shown in Fig. 3. Despite only three types of cells being tested here, they are representative in terms of morphological characteristics, encompassing a wide range of parameters such as roundness, perimeter, and long/short axis ratio found in most live cells. Figures 3(a1)–3(c1) present the phase images reconstructed by DHM, and the close-up views are shown in Figs. 3(d1)–3(f1). Due to laser sources typically used in DHM, the imaging quality is susceptible to speckle noise, which significantly impacts the segmentation performance of "EGT + DT + watershed" (this is likewise a common problem encountered by traditional approaches), as shown in Figs. 3(a2)–3(c2). Additionally, in conventional approaches, multi-parameters tuning typically relies on the user's expertise and exhibits substantial variability with changes in cell types and imaging systems (e.g., magnification and resolution). Image quality problems caused by noise can adversely affect the cell edge determination and cell detection accuracy [Figs. 3(d2)–3(f2)]. In the U-Net-based DL segmentation method, we used the same number of datasets

as DL-CSPF for training, and the segmentation results are presented in Figs. 3(a3)–3(c3), in which the adherent cells cannot be recognized and isolated, as illustrated in Figs. 3(d3)–3(f3). The cross-section across adhered cells intuitively signifies the indeed existing intercellular boundaries in complex cell states, as depicted in Figs. 3(g)–3(i). In contrast, DL-CSPF demonstrates excellent segmentation performance across three cell types, features smooth cell edges and accurate cell identification, and exhibits strong generalizability, even though image quality is compromised by noise, as presented in Figs. 3(a4)–(c4) and 3(d4)–3(f4). Figures 3(a5)–3(c5) and 3(d5)–3(f5) are the GT and the corresponding enlarged views. Herein, we just discussed the segmentation of phase images recovered by DHM. Additionally, we tested phase results reconstructed based on differential phase contrast [68,69] and Fourier ptychographic microscopy methods [70,71] to validate the generalizability of DL-CSPF, still demonstrating favorable segmentation performance.

According to Table 1, our method outperforms the traditional DL method based on U-Net and the "EGR + DT + watershed" method across the board in terms of recall, accuracy, precision, and F1 score, mirroring the GT. Furthermore, when trained on an equal number of datasets, DL-CSPF exhibits higher segmentation accuracy than U-Net. Moreover, attention-mechanism-based DL cell segmentation models enable the model to dynamically attend to different areas of the image when handling each pixel or feature, thereby improving the precision and robustness of segmentation. However, it still has not detached from its fundamentally data-driven essence, and no specific comparative studies have been conducted in this work.

4. DL-CSPF FOR LONG-TERM LIVE-CELL ANALYSIS WITH DHM

In this section, we conducted the experiment of HeLa cells for up to 9 h using DHM. With the support of the DL-CSPF approach, we analyzed the morphological metrics of live HeLa cells, including area and DM. Additionally, DL-CSPF can also be integrated with DHM to form a live-cell analysis framework for studying cell circumference, volume, irregularity, length-to-short axis ratio, texture [72], etc. Herein, we particularly discussed the DMSD throughout the complete process of single-cell division, analyzing its biological behavior. The used HeLa cells were cultured in the DMEM medium with 10% fetal bovine serum under standard cell culture conditions (37.2°C in 5% CO_2 in a humidified incubator) for observation.

A. Cell Morphology Characterization

1. Area

Based on the position of the cellular edges, the number of pixels within each cell contour can be calculated and multiplied by the pixel area to obtain the area of a single cell. Given the presence of an objective lens in DHM, which imparts a system magnification, the formula for calculating the area of a single cell is defined as follows:

$$S = \sum_{\Omega} (\text{pixelsize}/\text{Ma})^2, \quad \text{(3)}$$
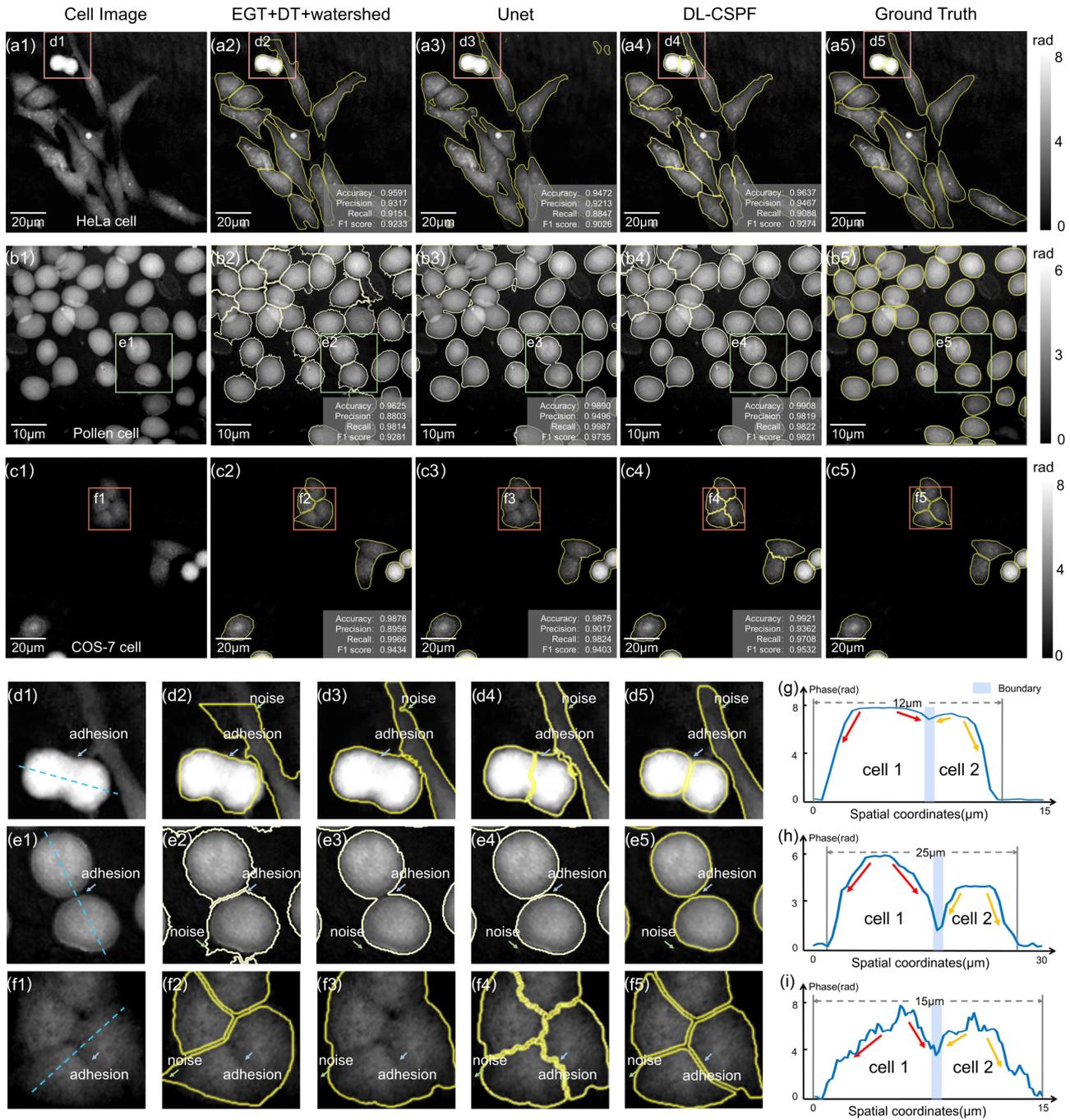
Fig. 3. Cell segmentation results for phase images by DHM on HeLa, pollen, COS-7 cells. (a1)–(c1) QPI results by DHM on three types of cells, the enlarged views of which are shown in (d1)–(f1). (a2)–(c2), (a3)–(c3), (a4)–(c4) Segmentation results using "EGT + DT + watershed", U-Net, and DL-CSPF methods, respectively. (c5)–(g5) Ground truth. (d2)–(f2), (d3)–(f3) The magnified views of the segmentation results using "EGT + DT + watershed" and U-Net demonstrate poor performance in handling cell adhesion and noise. (d4)–(f4) Corresponding results based on DL-CSPF, mirroring the GT, (d5)–(f5). (g)–(i) The cross-section over the adherent cells indicates the cell boundaries that indeed exist.

where pixelsize represents the pixel size of the CCD, Ma represents the system magnification, and Ω represents the contour position of an individual cell obtained after cell segmentation.

2. Dry Mass

Utilizing the fact that the refractive increments of most substances in cells are approximately the same and independent of composition, DHM is applicable to the measurement of cellular DM. The DMSD at each pixel (x, y) is calculated as

$$\rho(x, y) = \frac{\lambda}{2\pi\alpha}\varphi(x, y), \tag{4}$$

where α is a constant known as the specific refraction increment [73]. According to Ref. [74], we used the average value of this parameter of 0.2 ml/g for the DM calculations. Then the total

Table 1.   Quantitative Comparison Results of Three Methods

Method	Sample	Accuracy	Precision	Recall	F1 Score	Size of Data
EGT + DT + watershed	HeLa	0.9591	0.9317	0.9151	0.9233	NaN
	Pollen	0.9625	0.8803	0.9814	0.9281	NaN
	COS-7	0.9876	0.8956	0.9966	0.9434	NaN
U-Net	HeLa	0.9472	0.9213	0.8847	0.9026	710
	Pollen	0.9890	0.9496	0.9987	0.9735	710
	COS-7	0.9875	0.9017	0.9824	0.9403	710
DL-CSPF	HeLa	0.9637	0.9467	0.9088	0.9274	710
	Pollen	0.9908	0.9819	0.9822	0.9821	710
	COS-7	0.9921	0.9362	0.9708	0.9532	710
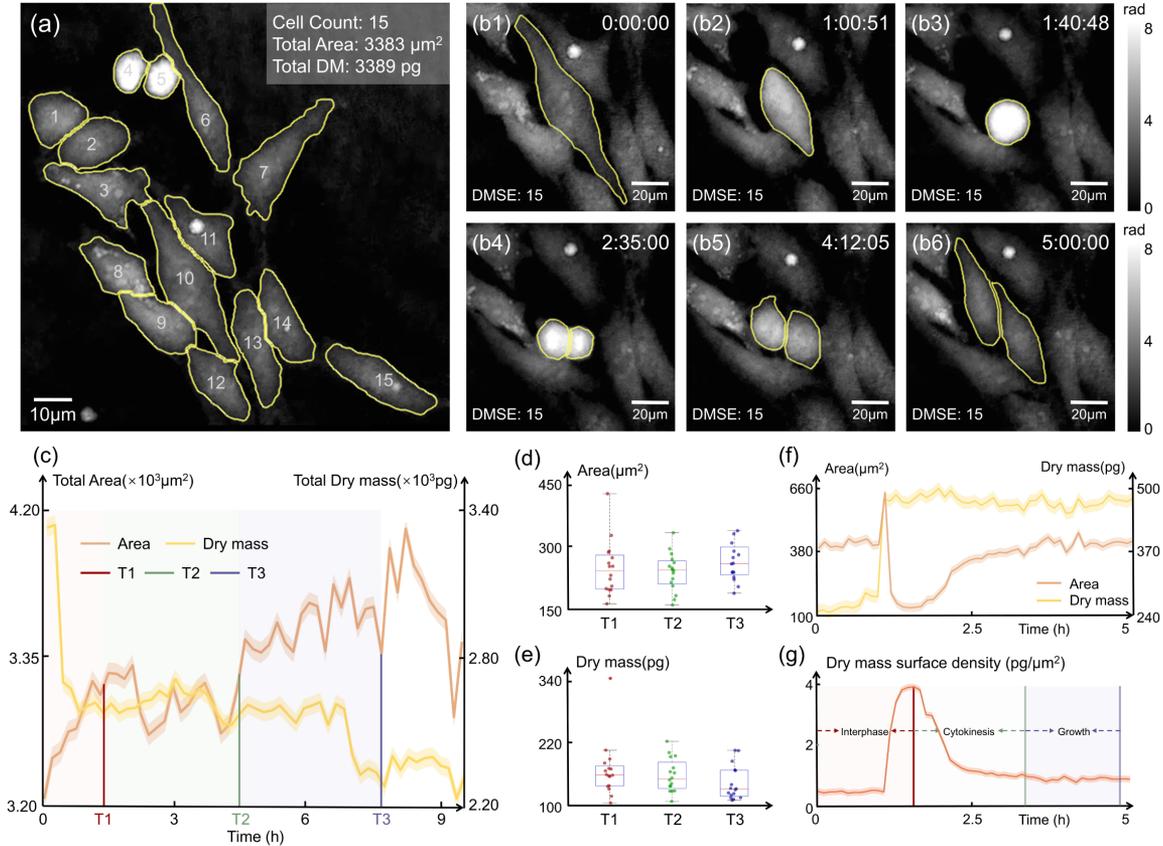
Fig. 4. Long-term dynamic live-cell analysis results. (a) Cell segmentation result at a certain time point. (b1)–(b6) Single-cell segmentation and recognition in the complete cell cycle. (c) Fluctuations in cellular total area and DM over the 9 h period. (d), (e) Box-scatter charts of the area and DM at time T_1, T_2, and T_3. (f) Single-cell variation of area and DM during the cell cycle. (g) Fluctuations in DMSD throughout the cell cycle.

DM is calculated by integrating the region of interest in the DMSD, and the expression is shown as follows:

$$\mathrm{DM} = \frac{\lambda}{2\pi\alpha} \int_S \Delta\varphi \, \mathrm{d}S. \tag{5}$$

B. Long-Term High-Accuracy Live-Cell Analysis Results

Figure 4 presents the results of DL-CSPF applied to long-term, high-precision live-cell analysis, in which Fig. 4(a) shows the segmentation result at a certain time point during the 9 h observation of HeLa cells using DHM. Furthermore, the fluctuations in cell area and DM over the 9 h period are illustrated

in Fig. 4(c), where the information at time points T_1, T_2, and T_3 is presented using a box-scatter chart [Figs. 4(d)–4(e)], demonstrating the individual distribution and statistical characteristics of cells at three time points, including the central tendency and dispersion of the distributions of cell area and DM metrics. As a fundamental cellular life activity, the investigation of cell proliferation phenomena is of particular importance. The complete process of single-cell division is tracked in Figs. 4(b1)–4(b6), which is accurately detected and recognized by DL-CSPF. In the cellular proliferation, the interphase constitutes the stage of the cell cycle characterized by cellular growth and DNA replication, typically occupying the majority of the cell cycle duration, which can be subdivided into the G_1, S, and G_2

phases [75]. Figure 4(f) illustrates the temporal fluctuations in single-cell area and DM throughout a cell cycle. Additionally, we introduced DMSD to analyze the variations during the division process, as shown in Fig. 4(g). Initially, the cell significantly increases its DM through protein synthesis and organelle replication, thereby providing the essential material foundation for subsequent DNA replication and cell division. Subsequently, the DM production remains constant and the surface growth pauses; the DMSD manifests a steep increase with a maximum value occurring a few minutes before the cytokinesis. As cells divide, the rapid surface increase and the septum degradation induce an abrupt decrease in DMSD. The accelerated dynamic cell segmentation and recognition results and are animated in Visualization 1.

5. CONCLUSION

In this work, we presented a novel DL-based cell segmentation method with a physical framework, in which the physical framework is introduced in a DL-based cell segmentation method for accurate and robust segmentation with diverse cell states (i.e., cell adhesion, proliferation, and apoptosis). Unlike the off-the-shelf end-to-end DL-based methods that directly extract effective information to learn features of individual cells, DL-CSPF conducts foreground-background segmentation and cell detection ("seed-point" extraction) from phase images using CNN. Due to the degradation of the information entropy of the dataset required for training, DL-CSPF achieves accurate cell segmentation while only using a small fraction of the dataset, demonstrating enhanced generalization. Moreover, compared with the traditional segmentation approaches that rely on manually crafted feature definitions, DL-CSPF exhibits superior segmentation performance in scenarios such as cell adhesion and noise without the need for task-dependent and experience-dependent parameters tuning. DL-CSPF is also employed in DHM for long-term dynamic live-cell analysis, and enables the morphological characterization and quantitative investigation of the cell division process by introducing DMSD. The feasibility of DL-CSPF was verified through experiments on an open-source dataset and a real dataset constructed by DHM for HeLa, pollen, and COS-7 cells. Further efforts will be made to improve the speed for achieving real-time high-dynamic cell segmentation. The algorithm will also be further engineered to incorporate plug-and-play functionality for seamless integration. By leveraging the advantages of integrating DL methodologies with a physical framework, DL-CSPF will also be further applied to the multimodality cell segmentation of fluorescence [76], phase contrast [69], DIC imaging [77] at the cellular scale, as well as to the segmentation of lipid droplets, mitochondria, nuclei, and other organelles at the organelle scale [78,79].

REFERENCES

1. J. Ma, R. Xie, S. Ayyadhury, *et al.*, "The multimodality cell segmentation challenge: toward universal solutions," Nat. Methods **21**, 1103–1113 (2024).
2. N. Hatipoglu and G. Bilgin, "Cell segmentation in histopathological images with deep learning algorithms by utilizing spatial relationships," Med. Biol. Eng. Comput. **55**, 1829–1848 (2017).
3. 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," Nat. Biotechnol. **40**, 555–565 (2022).
4. A. Wang, Q. Zhang, Y. Han, *et al.*, "A novel deep learning-based 3D cell segmentation framework for future image-based disease detection," Sci. Rep. **12**, 342 (2022).
5. W. Lin, Y. Chen, X. Wu, *et al.*, "STU-Net: Swin transformer U-Net for high-throughput live cell analysis with a lens-free on-chip digital holographic microscope," Opt. Eng. **63**, 111812 (2024).
6. M. Schreiber, T. Macháček, V. Vajs, *et al.*, "Suppression of the growth and metastasis of mouse melanoma by taenia crassiceps and mesocestoides corti tapeworms," Front. Immunol. **15**, 1376907 (2024).
7. K. D. Kovács, Z. Szittner, B. Magyaródi, *et al.*, "Optical sensor reveals the hidden influence of cell dissociation on adhesion measurements," Sci. Rep. **14**, 11719 (2024).
8. V. Del Vecchio, A. Rehman, S. K. Panda, *et al.*, "Mitochondrial transfer from adipose stem cells to breast cancer cells drives multi-drug resistance," J. Exp. Clin. Cancer Res. **43**, 166 (2024).
9. H. Gabai, M. Baranes-Zeevi, M. Zilberman, *et al.*, "Continuous wide-field characterization of drug release from skin substitute using off-axis interferometry," Opt. Lett. **38**, 3017–3020 (2013).
10. J. W. Lichtman and J.-A. Conchello, "Fluorescence microscopy," Nat. Methods **2**, 910–919 (2005).
11. Y. Liu, J. Zhou, S. Wen, *et al.*, "On-chip mirror enhanced multiphoton upconversion super-resolution microscopy," Nano Lett. **23**, 5514–5519 (2023).
12. J. Qian, Y. Cao, Y. Bi, *et al.*, "Structured illumination microscopy based on principal component analysis," eLight **3**, 4 (2023).
13. K. Zhao, X. Xu, W. Ren, *et al.*, "Two-photon minflux with doubled localization precision," eLight **2**, 5 (2022).
14. H. Lin and J.-X. Cheng, "Computational coherent Raman scattering imaging: breaking physical barriers by fusion of advanced instrumentation and data science," eLight **3**, 6 (2023).
15. Z. Wang, T. Zhao, H. Hao, *et al.*, "High-speed image reconstruction for optically sectioned, super resolution structured illumination microscopy," Adv. Photonics **4**, 026003 (2022).
16. A. Barty, K. Nugent, D. Paganin, *et al.*, "Quantitative optical phase microscopy," Opt. Lett. **23**, 817–819 (1998).
17. Y. Park, C. Depeursinge, and G. Popescu, "Quantitative phase imaging in biomedicine," Nat. Photonics **12**, 578–589 (2018).
18. J. Park, B. Bai, D. Ryu, *et al.*, "Artificial intelligence-enabled quantitative phase imaging methods for life sciences," Nat. Methods **20**, 1645–1660 (2023).
19. L. Lu, J. Li, Y. Shu, *et al.*, "Hybrid brightfield and darkfield transport of intensity approach for high-throughput quantitative phase microscopy," Adv. Photon. **4**, 056002 (2022).
20. Z. Li, J. Sun, Y. Fan, *et al.*, "Deep learning assisted variational Hilbert quantitative phase imaging," Opto-Electron. Sci. **2**, 220023 (2023).

21. Y. Shu, J. Sun, J. Lyu, *et al.*, "Adaptive optical quantitative phase imaging based on annular illumination Fourier ptychographic microscopy," PhotoniX **3**, 24 (2022).

22. E. R. Polanco, T. E. Moustafa, A. Butterfield, *et al.*, "Multiparametric quantitative phase imaging for real-time, single cell, drug screening in breast cancer," Commun. Biol. **5**, 794 (2022).

23. H. Zwirnmann, D. Knobbe, and S. Haddadin, "Towards end-to-end automated microscopy control using holotomography: workflow design and data management," IFAC-PapersOnLine **56**, 6477–6483 (2023).

24. Y. Jin, L. Lu, S. Zhou, *et al.*, "Neural-field-assisted trnasport-of-intensity phase microscopy: partially coherent quantitative phase imaging under unknown defocus distance," Photonics Res. **12**, 1494–1501 (2024).

25. Z. Huang and L. Cao, "Quantitative phase imaging based on holography: trends and new perspectives," Light Sci. Appl. **13**, 145 (2024).

26. J. Rosen, S. Alford, B. Allan, *et al.*, "Roadmap on computational methods in optical imaging and holography," Appl. Phys. B **130**, 166 (2024).

27. Y. Chen, X. Wu, L. Lu, *et al.*, "Pixel-super-resolved lens-free quantitative phase microscopy with partially coherent illumination," npj Nanophotonics **1**, 15 (2024).

28. Z. Li, Y. Chen, J. Sun, *et al.*, "High bandwidth-utilization digital holographic reconstruction using an untrained neural network," Appl. Sci. **12**, 10656 (2022).

29. X. Wu, N. Zhou, Y. Chen, *et al.*, "Lens-free on-chip 3D microscopy based on wavelength-scanning Fourier ptychographic diffraction tomography," Light Sci. Appl. **13**, 237 (2024).

30. J. Zhang and J. Hu, "Image segmentation based on 2D OTSU method with histogram analysis," in *International Conference on Computer Science and Software Engineering* (IEEE, 2008), Vol. **6**, pp. 105–108.

31. H. Zhang, Q. Zhu, and X.-F. Guan, "Probe into image segmentation based on Sobel operator and maximum entropy algorithm," in *International Conference on Computer Science and Service System* (IEEE, 2012), pp. 238–241.

32. P. Bao, L. Zhang, and X. Wu, "Canny edge detection enhancement by scale multiplication," IEEE Trans. Pattern Anal. Mach. Intell. **27**, 1485–1490 (2005).

33. H. Kong, H. C. Akakin, and S. E. Sarma, "A generalized Laplacian of Gaussian filter for blob detection and its applications," IEEE Trans. Cybern. **43**, 1719–1733 (2013).

34. A. S. Ahmed, "Comparative study among sobel, prewitt and canny edge detection operators used in image processing," J. Theor. Appl. Inf. Technol. **96**, 6517–6525 (2018).

35. R. Pohle and K. D. Toennies, "Segmentation of medical images using adaptive region growing," Proc. SPIE **4322**, 1337–1346 (2001).

36. K. Haris, S. N. Efstratiadis, N. Maglaveras, *et al.*, "Hybrid image segmentation using watersheds and fast region merging," IEEE Trans. Image Process. **7**, 1684–1699 (1998).

37. N. Dhanachandra, K. Manglem, and Y. J. Chanu, "Image segmentation using K-means clustering algorithm and subtractive clustering algorithm," Proc. Comput. Sci. **54**, 764–771 (2015).

38. P. Raj, S. K. Paidi, L. Conway, *et al.*, "CellSNAP: a fast, accurate algorithm for 3D cell segmentation in quantitative phase imaging," J. Biomed. Opt. **29**, S22706 (2024).

39. W. Cho, D. Choi, H. Lim, *et al.*, "Slice and conquer: a planar-to-3D framework for efficient interactive segmentation of volumetric images," in *IEEE/CVF Winter Conference on Applications of Computer Vision* (2024), pp. 7614–7623.

40. J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *IEEE Conference on Computer Vision and Pattern Recognition* (2015), pp. 3431–3440.

41. J. Lee, H. Kim, H. Cho, *et al.*, "Deep-learning-based label-free segmentation of cell nuclei in time-lapse refractive index tomograms," IEEE Access **7**, 83449–83460 (2019).

42. M. Lee, M. Kunzi, G. Neurohr, *et al.*, "Hybrid machine-learning framework for volumetric segmentation and quantification of vacuoles in individual yeast cells using holotomography," Biomed. Opt. Express **14**, 4567–4578 (2023).

43. M. Lee, Y.-H. Lee, J. Song, *et al.*, "Deep-learning-based three-dimensional label-free tracking and analysis of immunological synapses of car-t cells," Elife **9**, e49023 (2020).

44. C. Zuo, J. Qian, S. Feng, *et al.*, "Deep learning in optical metrology: a review," Light Sci. Appl. **11**, 39 (2022).

45. S. Feng, Q. Chen, G. Gu, *et al.*, "Fringe pattern analysis using deep learning," Adv. Photon. **1**, 025001 (2019).

46. W. Yin, Y. Che, X. Li, *et al.*, "Physics-informed deep learning for fringe pattern analysis," Opto-Electron. Adv. **7**, 230034 (2024).

47. A. Saba, C. Gigli, A. B. Ayoub, *et al.*, "Physics-informed neural networks for diffraction tomography," Adv. Photonics **4**, 066001 (2022).

48. O. Ronneberger, P. Fischer, and T. Brox, "U-Net: convolutional networks for biomedical image segmentation," in *18th International Conference Medical Image Computing and Computer-assisted Intervention (MICCAI), Part III 18*, (Springer, 2015), pp. 234–241.

49. K. He, G. Gkioxari, P. Dollár, *et al.*, "Mask R-CNN," in *IEEE International Conference on Computer Vision* (2017), pp. 2961–2969.

50. J. I. Orlando, P. Seeböck, H. Bogunović, *et al.*, "U2-Net: a Bayesian U-Net model with epistemic uncertainty feedback for photoreceptor layer segmentation in pathological OCT scans," in *IEEE 16th International Symposium on Biomedical Imaging (ISBI)* (IEEE, 2019), pp. 1441–1445.

51. O. Oktay, J. Schlemper, L. L. Folgoc, *et al.*, "Attention U-Net: learning where to look for the pancreas," arXiv (2018).

52. T. Prangemeier, C. Reich, and H. Koeppl, "Attention-based transformers for instance segmentation of cells in microstructures," in *IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* (IEEE, 2020), pp. 700–707.

53. D. Bolya, C. Zhou, F. Xiao, *et al.*, "YOLACT: real-time instance segmentation," in *IEEE/CVF International Conference on Computer Vision* (2019), pp. 9157–9166.

54. A. Kirillov, E. Mintun, N. Ravi, *et al.*, "Segment anything," in *IEEE/CVF International Conference on Computer Vision* (2023), pp. 4015–4026.

55. J. Ma, Y. He, F. Li, *et al.*, "Segment anything in medical images," Nat. Commun. **15**, 654 (2024).

56. C. Stringer, T. Wang, M. Michaelos, *et al.*, "Cellpose: a generalist algorithm for cellular segmentation," Nat. Methods **18**, 100–106 (2021).

57. M. Pachitariu and C. Stringer, "Cellpose 2.0: how to train your own model," Nat. Methods **19**, 1634–1641 (2022).

58. K. Parvati, B. Prakasa Rao, and M. Mariya Das, "Image segmentation using gray-scale morphology and marker-controlled watershed transformation," Discrete Dyn. Nat. Soc. **2008**, 384346 (2008).

59. C. Zuo, Q. Chen, W. Qu, *et al.*, "Phase aberration compensation in digital holographic microscopy based on principal component analysis," Opt. Lett. **38**, 1724–1726 (2013).

60. M. A. Herráez, D. R. Burton, M. J. Lalor, *et al.*, "Fast two-dimensional phase-unwrapping algorithm based on sorting by reliability following a noncontinuous path," Appl. Opt. **41**, 7437–7444 (2002).

61. J. Chalfoun, M. Majurski, A. Peskin, *et al.*, "Empirical gradient threshold technique for automated segmentation across image modalities and cell lines," J. Microsc. **260**, 86–99 (2015).

62. K. Thirusittampalam, M. J. Hossain, O. Ghita, *et al.*, "A novel framework for cellular tracking and mitosis detection in dense phase contrast microscopy images," IEEE J. Biomed. Health Inform. **17**, 642–653 (2013).

63. T. Vicar, J. Balvan, J. Jaros, *et al.*, "Cell segmentation methods for label-free contrast microscopy: review and comprehensive comparison," BMC Bioinf. **20**, 1–25 (2019).

64. K. He, X. Zhang, S. Ren, *et al.*, "Deep residual learning for image recognition," in *IEEE Conference on Computer Vision and Pattern Recognition* (2016), pp. 770–778.

65. Y. Ho and S. Wookey, "The real-world-weight cross-entropy loss function: modeling the costs of mislabeling," IEEE Access **8**, 4806–4813 (2019).

66. T. Vicar, J. Chmelik, R. Jakubicek, *et al.*, "Self-supervised pretraining for transferable quantitative phase image cell segmentation," Biomed. Opt. Express **12**, 6514–6528 (2021).

67. Y. Fan, J. Li, L. Lu, *et al.*, "Smart computational light microscopes (SCLMs) of smart computational imaging laboratory (SCILab)," PhotoniX **2**, 1–64 (2021).

68. L. Tian and L. Waller, "Quantitative differential phase contrast imaging in an LED array microscope," Opt. Express **23**, 11394–11403 (2015).

69. Y. Fan, J. Sun, Y. Shu, *et al.*, "Accurate quantitative phase imaging by differential phase contrast with partially coherent illumination: beyond weak object approximation," Photon. Res. **11**, 442–455 (2023).

70. G. Zheng, R. Horstmeyer, and C. Yang, "Wide-field, high-resolution Fourier ptychographic microscopy," Nat. Photonics **7**, 739–745 (2013).

71. C. Zuo, J. Sun, and Q. Chen, "Adaptive step-size strategy for noise-robust Fourier ptychographic microscopy," Opt. Express **24**, 20724–20744 (2016).

72. Q. Shen, Z. Li, J. Sun, *et al.*, "Live-cell analysis framework for quantitative phase imaging with slightly off-axis digital holographic microscopy," Front. Photon. **3**, 1083139 (2022).

73. M. Mir, Z. Wang, Z. Shen, *et al.*, "Optical measurement of cycle-dependent cell growth," Proc. Natl. Acad. Sci. USA **108**, 13124–13129 (2011).

74. R. Barer, "Interference microscopy and mass determination," Nature **169**, 366–367 (1952).

75. E. A. Nigg, "Cyclin-dependent protein kinases: key regulators of the eukaryotic cell cycle," Bioessays **17**, 471–480 (1995).

76. J. Qian, K. Xu, S. Feng, *et al.*, "Fast-structured illumination microscopy based on dichotomy–correlation parameter estimation (dCOR-SIM)," ACS Photon. **11**, 1887–1892 (2024).

77. M. R. Arnison, K. G. Larkin, C. J. Sheppard, *et al.*, "Linear phase imaging using differential interference contrast microscopy," J. Microsc. **214**, 7–12 (2004).

78. H. Kim, S. Oh, S. Lee, *et al.*, "Recent advances in label-free imaging and quantification techniques for the study of lipid droplets in cells," Curr. Opinion Cell Biol. **87**, 102342 (2024).

79. G. Onal, O. Kutlu, D. Gozuacik, *et al.*, "Lipid droplets in health and disease," Lipids Health Dis. **16**, 1–15 (2017).