# Deep Learning and Computer Vision Strategies for Automated Gene Editing with a Single-Cell Electroporation Platform

SLAS Technology I–II © Society for Laboratory Automation and Screening 2020 DOI: 10.1177/2472630320982320 journals.sagepub.com/home/jla **©SAGE** 

Cesar A. Patino<sup>1,3</sup>, Prithvijit Mukherjee<sup>1,2,3</sup>, Vincent Lemaitre<sup>3</sup>, Nibir Pathak<sup>1,2</sup>, and Horacio D. Espinosa<sup>1,2,3</sup>

#### Abstract

Single-cell delivery platforms like microinjection and nanoprobe electroporation enable unparalleled control over cell manipulation tasks but are generally limited in throughput. Here, we present an automated single-cell electroporation system capable of automatically detecting cells with artificial intelligence (AI) software and delivering exogenous cargoes of different sizes with uniform dosage. We implemented a fully convolutional network (FCN) architecture to precisely locate the nuclei and cytosol of six cell types with various shapes and sizes, using phase contrast microscopy. Nuclear staining or reporter fluorescence was used along with phase contrast images of cells within the same field of view to facilitate the manual annotation process. Furthermore, we leveraged the near-human inference capabilities of the FCN network in detecting stained nuclei to automatically generate ground-truth labels of thousands of cells within seconds, and observed no statistically significant difference in performance compared to training with manual annotations. The average detection sensitivity and precision of the FCN network were  $95 \pm 1.7\%$  and  $90 \pm 1.8\%$ , respectively, outperforming a traditional image-processing algorithm ( $72\pm7.2\%$  and  $72\pm5.5\%$ ) used for comparison. To test the platform, we delivered fluorescentlabeled proteins into adhered cells and measured a delivery efficiency of 90%. As a demonstration, we used the automated single-cell electroporation platform to deliver Cas9-guide RNA (gRNA) complexes into an induced pluripotent stem cell (iPSC) line to knock out a green fluorescent protein–encoding gene in a population of ~200 cells. The results demonstrate that automated single-cell delivery is a useful cell manipulation tool for applications that demand throughput, control, and precision.

#### **Keywords**

deep learning, computer vision, electroporation, single-cell, CRISPR-Cas9

## Introduction

Precise delivery of exogenous materials into cells is essential to performing a variety of cell manipulation tasks, including CRISPR-Cas9 gene editing,1,2 plasmid transfection,<sup>3,4</sup> and small interfering RNA (siRNA) gene silencing.5 To this end, Espinosa and coworkers developed a nanofountain probe electroporation (NFP-E) system capable of delivering molecules of different sizes (i.e., DNA plasmids, molecular beacons, proteins, nanoparticles, and CRISPR-Cas9 complexes) into a variety of cell types.<sup>2,6–8</sup> The localized-electroporation mechanism used by NFP-E is a gentle process in which only a small area of the cell membrane, in contact with the nanoprobe (~500 nm diameter), is permeabilized. Conversely, bulk physical membrane disruption methods (i.e., bulk-electroporation and mechanoporation systems) subject the entire cell surface to high electric fields or mechanical stresses that may lead to an undesirable intracellular stress response or reduced viability.<sup>9–12</sup> Other commonly used delivery methods, like lipid-vesicle carriers and viral vectors, are limited by low delivery efficiency,<sup>13</sup> lack of dosage uniformity,<sup>2</sup> and cytotoxicity.<sup>14</sup> The nondestructive mechanism and precise dosage control enabled by

Evanston, IL, USA

<sup>2</sup>Theoretical and Applied Mechanics Program, Northwestern University, Evanston, IL, USA

<sup>3</sup>iNfinitesimal LLC, Skokie, IL, USA

Received Aug 19, 2020, and in revised form Nov 23, 2020. Accepted for publication Nov 24, 2020.

Supplemental material is available online with this article.

#### **Corresponding Author:**

Horacio D. Espinosa, Department of Mechanical Engineering, Northwestern University, 2145 Sheridan Rd., Evanston, IL 60208, USA. Email: espinosa@northwestern.edu

<sup>&</sup>lt;sup>1</sup>Department of Mechanical Engineering, Northwestern University,



**Figure I.** Schematic of automated single-cell delivery workflow using Infinitesimal's nanofountain probe electroporator (NFP-E): (a) Computer vision software recognition of cell nuclei in the field of view, (b) image of the NFP-E system, and (c) phase contrast micrograph of nanoprobe engagement on the detected cell (scale bar: 50  $\mu$ m).

the NFP-E system are advantageous for applications that require uniform delivery, temporal precision, high cell viability, and single-cell control (i.e., single-cell gene editing and monoclonal cell-line generation).<sup>15</sup> These applications, however, require the delivery of molecular cargoes into hundreds to thousands of cells, which necessitates a high-throughput workflow.

To increase throughput of single-cell delivery systems and expand the range of applications, automation of the experimental workflow is essential.<sup>16,17</sup> Single-cell transfection systems perform calibration routines prior to delivery and use live imaging and manual selection of cells within a field of view (FOV). Consequently, these operations require user intervention and result in extended experimental times. To automate this process, computer vision algorithms must be integrated in the experimental workflow to precisely locate the positions of cells and nanoprobes in a FOV prior to probe positioning and molecular delivery (Fig. 1). Moreover, the computer vision routine must be able to distinguish cells in both brightfield- and fluorescence imaging modalities to enable label-free transfection. Traditional image-processing workflows, like watershed and mathematical morphology operations, have been implemented to detect cells but with limited success. For example, the watershed algorithm is well suited for segmentation of fluorescent images when there is a sharp intensity gradient between adjacent objects. The contrast at the edges of cells and their nuclei in phase contrast micrographs is not, however, as clear as in fluorescence imaging. One approach, the ellipse-fit algorithm, was developed to segment the nuclei of brightfield images based on the identification of high-contrast internal features through a series of mathematical morphology and edge detection operations.<sup>18</sup> This method, however, is limited to specific cell types possessing clearly discernable nuclei features.

These limitations have been overcome with the rapidly growing field of artificial intelligence (AI), in which deep learning algorithms vastly outperform traditional imageprocessing workflows for executing computer vision tasks, in some cases exceeding human inference performance.<sup>19–21</sup> Deep learning convolutional neural networks (CNNs; e.g., VGG16<sup>22</sup> and ImageNet<sup>21</sup>) are particularly advantageous for the task of object recognition due to their ability to extract features (like edges and color patterns) from local regions in an image. Fully convolutional networks (FCNs), an encoder-decoder adaptation of CNNs, provide pixellevel object classification, thereby enabling the identification and segmentation of numerous individual objects in an image.<sup>23</sup> The instance-aware segmentation capability of FCNs is particularly advantageous for segmenting closely bound objects, such as cells, and has been widely used for cell recognition tasks.<sup>24</sup> For example, cell nuclei segmentation was the objective of a computer vision competition, Kaggle's 2018 Data Science Bowl,<sup>25</sup> where high-performing deep learning networks included FCNs, mask-RCNNs (region CNNs),<sup>26,27</sup> feature pyramid networks (FPNs),<sup>28</sup> and ensemble networks. Building on such investigations, here we use FCNs and a transfer-learning approach to accurately detect cells in phase contrast micrographs and enable highthroughput automated delivery of exogenous cargoes into single cells. Furthermore, we demonstrate an automatic ground-truth (GT) labeling approach to expedite the annotation process and provide a more rapid avenue for laboratories to use their own datasets for training deep networks.

## **Materials and Methods**

## Automated Single-Cell Electroporation

An NFP-E system (Infinitesimal LLC, Skokie, IL) was used to deliver molecular cargoes into adhered cells via application of electrical pulses. Delivery reagents were loaded into a nanopipette (Eppendorf, Hamburg, Germany) and connected to the NFP-E instrument via a pipette holder containing a Pt-wire electrode. Culture dishes with adhered cells were then placed on a motorized microscope stage ( $\pm 1 \, \mu m$ resolution) for XY positioning and inverted microscopy imaging (Nikon TE 2000; Nikon, Tokyo, Japan). Next, the nanopipette was immersed in the culture dish, focused using a normalized variance autofocus routine, and calibrated using a template-matching algorithm (Suppl. Fig. S1). Once the pipette-tip position was calibrated, the cells were brought into focus using the autofocus routine. Phase contrast images ( $20 \times$  objective; Nikon) were acquired with CCD (Andor, Oxford Instruments, Abingdon, UK) and machine vision cameras (FLIR, Wilsonville, OR), and cells in the FOV were automatically identified using the AI machine vision routines. The calibrated pipette tip was moved to each cell using the three-axis piezo drive. Once positioned on the nuclei centroid coordinate, the nanopipette tip was lowered while simultaneously acquiring circuit resistance measurements using Infinitesimal LLC proprietary electronics. Using Infinitesimal's contact detection algorithm,<sup>29</sup> based on a resistance measurement profile, the nanopipette motion was stopped to achieve a good seal while avoiding cell puncture. Once in contact with the cell, a pulsed voltage history, selected via a graphic user interface (GUI), was applied to open pores in the membrane and delivered the reagents. The following pulse parameters were adjusted prior to electroporation: pulse shape (i.e., square, bi-level, or exponential), supplied voltage  $[V_s(V)]$ , pulse duration (s), frequency (Hz), and number of pulses.

## Cell Detection Using Fully Convolutional Networks

A FCN containing 20 hidden layers (Fig. 2A) was developed (using TensorFlow, www.tensorflow.org; and Keras, https://keras.io) to detect cells in phase contrast and fluorescence imaging modalities. The FCN architecture consists of an encoder-decoder structure designed to extract local image features and construct a multiclass probability output that maps to each pixel in the image. Convolution layers  $(3 \times 3 \text{ kernel size})$  were connected to a ReLu nonlinear activation function, and downsampled with a max-pooling layer ( $2 \times 2$  kernel size, stride 2). Upsampling ( $2 \times 2$  kernel size, stride 2) and concatenation steps were implemented in the decoder portion of the network. A weighted soft-max cross-entropy loss function was used to classify the cell interior, outlines, and exterior with a weight ratio of 5:10:1, respectively. Image datasets, containing image input and labeled ground truth, were divided for cross-validation with a 5:1 training-to-validation ratio. Augmentation generators (image flipping, zooming, intensity variation, and rotation) were applied to transform the image and GT inputs. Training was performed on a graphic processing unit (GPU; NVIDIA RTX 2080; NVIDIA, Santa Clara, CA) equipped with network acceleration libraries (NVIDIA cuDNN). The FCN network was trained in batches (batch size: 6), in which the network weights were optimized using stochastic gradient descent (learning rate: 1e-4; and momentum: 0.9). Training was stopped when the moving average of the validation loss reached a minimum value.

The image sets used for training the FCN network were obtained by capturing phase contrast and fluorescence micrographs, in the same FOV, of cells stained with fluorescent dyes to enable accurate localization of the nuclei (Hoechst) or cytosol (Calcein AM), respectively. The fluorescent micrographs were used to label the outlines of the nuclei or cytosol using an image manipulation software (GIMP, www.gimp.org), and were processed with an image analysis pipeline (CellProfiler, https://cellprofiler.org) to generate the three classes (interior, boundary, and exterior) for each image.<sup>30</sup> After labeling the images, a general fluorescence FCN model was trained with the fluorescence images (including all cell types in our database) and the manually annotated GT labels as inputs (see Fig. 2Bi) to enable nuclei or cytosol detection in fluorescence micrographs. Since the phase contrast and fluorescence micrographs were captured for each FOV, the same GT labels generated using the fluorescent images were used as inputs for training a FCN network to enable nuclei or cytosol detection in phase contrast micrographs (see Fig. 2Bii). In addition to a general phase contrast FCN model (trained with all of the cell types in our database), we fine-tuned separate FCN models for each individual cell type to increase performance. To expedite the GT labeling process, an automated labeling approach was incorporated (see Fig. **2Biii**) by using the FCN model used for fluorescence predictions (Fig. 2Bi) to generate the GT labels on new image sets (fluorescence and phase images captured in the same FOV) automatically.

After training the FCN models for cell and nuclei detection for various cell types, the models were optimized for inference acceleration by using an inference engine optimization software development toolkit (Intel OpenVINO; Intel, Santa Clara, CA). The optimized models were integrated into the NFP-E software (C# language) to enable nuclei and cytosol detection from live images and subsequent nanoprobe electroporation.

A postprocessing pipeline was developed in the NFP-E software (OpenCV, https://opencv.org) to measure the shape and intensity profiles of the detected nuclei and cells. The following shape features were measured for each object: area, perimeter, circularity, solidity, eccentricity, major axis, and minor axis (**Suppl. Table S1**). To obtain the intensity-based features (mean intensity, median intensity, and intensity variance), the FCN-predicted objects were overlaid with the original input image for measurement of intensity



Figure 2. Fully convolutional network (FCN) architecture and training scheme. (a) Encoderdecoder architecture of a deep FCN containing 20 hidden layers connected with a series of convolution, activation, pooling, upconvolution, and concatenation steps (see colored arrows). The input image is downsampled in the encoder portion of the network and decoded to provide pixel-level predictions of three object classes (interior, boundary, and exterior). (b) Training schemes: (i) FCN training with fluorescence image and manual annotation inputs, (ii) transfer learning of phase contrast image using manual annotations and pre-trained model from step (i), and (iii) transfer learning of phase contrast image using automatic annotations and pre-trained model from step (i).

at the location of the objects. Statistics were obtained for each of the shape and intensity features for each cell type (nuclei and cytosol) measured from the GT labels and used for determining thresholds for filtering objects in the live images. Area filters were used to remove small artifacts and large clumps, if present, determined using the following:  $|A_{object} - \mu_{area}| > \alpha * \sigma_{area}$ , where  $A_{object}$  is the area of the segmented object in the live image,  $\mu_{area}$  is the mean area of the GT dataset,  $\alpha$  is a user-defined multiplication factor (  $\alpha$  > 0), and  $\sigma_{\rm area}$  is the standard deviation of the area of the GT dataset. A one-sample *t* test was used to measure statistically significant differences between the segmented image and the GT dataset. If the average object area in the live image is significantly greater (P < 0.05) than the GT dataset mean area, the image was categorized as undersegmented. In the case of an undersegmented image, a watershed algorithm was used to further segment the clumped objects. Moreover, a GUI was developed to enable the adjustment of filters, based on the shape and intensity measurements of interest, to automatically select cells of interest for transfection.

## **Performance Metrics**

The segmentation performance of the FCN algorithm was evaluated by measuring the overlap of the output predictions with the GT annotations. The intersection over union (IOU) was calculated from the following equation:  $IOU = \frac{|A \cap B|_{area}}{|A \cup B|_{area}}$ , where A represents the annotated image GT, and B represents the predicted label outputs.

A threshold IOU value was used to determine true positives (TP), false positives (FP), and false negatives (FN). The precision (P) and sensitivity (S) were calculated from these values according to the following equations: P = TP / (TP + FP) and S = TP / (TP + FN). The harmonic mean of the P and S values was used to calculate a segmentation quality score, F1, at specific IOU thresholds:  $Fl = 2P \times S / (P + S)$ .

#### Protein Delivery Experiments

Breast cancer cells (MDA-MB231) and kidney cells (HEK293) were plated in separate 35 mm culture dishes (Falcon, **[AQ: 1]**) 24 h prior to electroporation. A 2.5 mg/ mL solution of bovine serum albumin (BSA) protein tagged with Alexa Fluor 488 (BSA-488; Thermo Fisher Scientific, Waltham, MA) in 1× Dulbecco's phosphate-buffered saline (DPBS; Gibco, Carlsbad, CA) was loaded into a nanopipette (Eppendorf) and mounted on the NFP-E. The cells were electroporated with the following pulse parameters:

bi-level profile,  $V_{S1} = 15$  (V),  $V_{S2} = 10$  (V),  $t_1 = .5$  (ms),  $t_2 = 2.5$  (ms), f = 20 (Hz), and  $N_{pulses} = 200$ . Cells were imaged post electroporation in an inverted microscope equipped with fluorescence and phase contrast imaging modules (Nikon).

Human induced pluripotent stem cells (iPSCs) encoding for a copGFP gene (copGFP-iPSCs) were plated on a gridded dish (~25,000 cells) and incubated for 24 h prior to electroporation. A stock solution of red fluorescing BSA-AF647 (Thermo Fisher Scientific) in 1× DPBS (Gibco) was loaded into a nanopipette. A bi-level pulse profile [V<sub>S1</sub> = 15 (V), V<sub>S2</sub> = 5 (V), t<sub>1</sub> = .5 (ms), t<sub>2</sub> = 2.5 (ms), f = 20 (Hz), and N<sub>pulses</sub> = 200] was applied to electroporate the cells.

## Cas9-gRNA Delivery Experiments

A guide RNA (gRNA) was engineered to target the sequence of the copGFP gene of an iPSC line. A final concentration of 1  $\mu$ M gRNA and 1  $\mu$ M Cas9 nuclease in DPBS was prepared and placed in an incubator (5% CO<sub>2</sub>, 37 °C, and 98% relative humidity) for 10 min to synthesize the Cas9 and gRNA.

The Cas9–gRNA ribonucleoprotein complex was loaded into a nanopipette and placed in the NFP-E holder. The automated electroporation protocol was performed to deliver the gene-editing molecules into the iPSCs (~1000) plated on a gridded dish (Ibidi) using the following electroporation parameters: The bi-level profile is  $V_{s1} = 15$  (V),  $V_{s2} = 10$  (V),  $t_1 = .5$  (ms),  $t_2 = 2.5$  (ms), f = 20 (Hz), and  $N_{pulses} = 200$ . The cells were treated with rho-kinase inhibitors post electroporation and allowed to expand for 4 days. Loss of green fluorescence in the electroporated cells indicated the absence of copGFP protein and consequently the knockout of the copGFP gene.

#### Cell Culture

Cancer cell lines (HEK293, MDA-MB231, and HeLa) were cultured with Dulbecco's modified Eagle medium (DMEM; Gibco) with 4% fetal bovine serum (FBS) and 1% antibiotic. The cells were passaged with 0.25% Trypsin-EDTA (Life Technologies, Carlsbad, CA), transferred to a 35 mm culture dish (Falcon), and placed in an incubator (5%  $CO_2$ , 37°C, and 98% relative humidity) for 24h prior to electroporation.

Primary fibroblasts were cultured in culture dishes containing fibroblast basal medium with a low-serum growth kit (ATCC PCS-201-041; ATCC, Manassas, VA). The cells were passaged with 0.05% Trypsin-EDTA (Life Technologies), transferred to a 35 mm culture dish (Falcon), and placed in an incubator (5% CO<sub>2</sub>, 37 °C, and 98% relative humidity) for 24 h prior to electroporation.

iPSCs engineered with green-fluorescing copGFP genes were cultured in Matrigel-coated (Corning, Corning, NY) culture dishes containing Essential 8 Medium (Thermo Fisher Scientific).

#### Fluorescence Staining

A Hoechst nuclear dye (Life Technologies) solution of 0.1 mg/ml in phosphate-buffered saline (PBS) was prepared. The Hoechst solution was added to culture dishes plated with cells after removal of the culture media. Culture dishes with Hoechst were placed in an incubator for 10 min at 5%  $CO_2$ , 37 °C, and 98% relative humidity. After incubation, the Hoechst solution was removed and washed thrice in PBS prior to imaging.

#### Statistical Analysis

Statistics and performance metric calculations were performed in Python. A two-tailed t test was used to calculate P values to assess statistical significance. Standard error bars were computed from experiments with triplicate data at minimum. Spearman's correlation matrix was used to generate the heat map of extracted cell features.

## **Results and Discussion**

## Phase Contrast Cell Nuclei and Cytosol Detection

Precise detection of cell nuclei in brightfield images is essential for the automated label-free delivery of exogenous materials, including plasmids and CRISPR-Cas9 complexes, into a large number of cells using the NFP-E system. We obtained phase contrast and fluorescence images of various cell types stained with a nuclei marker [including HeLa cells, iPSCs, HEK293 cells, fibroblasts, and neural stem cells (NSCs)], each comprising a unique morphology, and trained the FCN algorithm to detect the nuclei in the phase contrast images using the manually annotated GT labels obtained from the fluorescent images. The segmentation outputs overlaid with the image inputs are shown in Figure 3A. Nuclei of all of the cell types were clearly identified, including images in which cells were seeded inside micropatterned substrates used for the generation of monoclonal cell lines. Segmentation performance was measured from the IOU overlap of predicted nuclei outputs and GT label inputs. A threshold IOU value was used to discriminate true positives from false positives, and in turn used to calculate segmentation precision and sensitivities. A segmentation score (F1), calculated from the harmonic mean of precision and sensitivity values, was used to measure the segmentation performance for each of the cell types in our dataset using the fine-tuned FCN models (Fig. 3B). Cells containing clearly distinct nuclei outlines in the phase contrast images, like



**Figure 3.** Fully convolutional network (FCN) nuclei segmentation. (a) Composite images overlaying FCN segmentation labels (magenta) on phase contrast micrographs of various cell types. Image of detected induced pluripotent stem cells (iPSCs) in microwell (350  $\mu$ m diameter) is shown in bottom right. Scale bars = 100  $\mu$ m. (b) Bar plot containing segmentation performance metrics [FI score at 50% intersection over union (IOU)] measured for the various cell types from (a). (c) Line graph of FCN training and validation loss during the course of 2000 iterations, with and without zoom augmentation. (d) Segmentation performance line plot for a range of IOU thresholds, with and without zoom augmentation.

iPSCs, were segmented with the highest precision and accuracy (F1<sub>iPSC</sub> = 0.92), while highly merged cells, like HEK293, resulted in the lowest segmentation score  $(F1_{HEK203} = .48)$ . The other cell types, including NSCs, fibroblasts, and HeLa cells, contained clearly separated nuclei and were segmented with high segmentation scores  $(F_{NSC} = .75, F1_{fibroblast} = .69, and F1_{HeLa} = .66, respectively).$  The average segmentation scores gathered for all of the cell types were in the range of the high-performing algorithms submitted to the Kaggle nuclei segmentation competition, in which the dataset used comprised a high number of fluorescence-labeled nuclei.<sup>25</sup> Segmentation performance of fluorescence micrographs is higher than that of brightfield imaging modalities due to the sharp intensity gradients at the edges of the stained nuclei compared to phase contrast. Using transfer learning from a large database (~10,000 nuclei) facilitated the training process and made the network

more robust to overfitting. Furthermore, we implemented image augmentation (zoom, flip, and rotation) during the training to increase the size of the training data and to generate a robust FCN model capable of detecting cells at regions of interest (ROIs) of different sizes. Consequently, the augmentation steps enabled the FCN algorithm to be trained for more iterations while avoiding overfitting (**Fig. 3C and Suppl. Fig. S2**). Accordingly, the segmentation performance was higher with zoom augmentation for the entire range of IOU thresholds (**Fig. 3D**). Training the algorithm with zoomed regions enables NFP-E users to transfect cells of interest within a specified ROI while maintaining high detection accuracy and precision.

To test the performance of a network trained with automatically generated GT labels, we collected images (phase contrast and fluorescence) of HeLa cells containing a blue nuclei stain (Hoechst) and automatically generated labels by passing the fluorescence images through a pre-trained FCN network. We compared the automatically generated labels to human annotations and measured the segmentation scores. The segmentation performance of the automatically generated labels approached the precision of human annotations  $[F1_{auto} (70\% \text{ IOU}) = .94]$  compared to the fluorescence GT labels. Next, we used the automatically generated labels to train a FCN network to identify cell nuclei in phase contrast images. The nuclei segmentation outputs from two independent FCN networks, trained with manual and automatically generated GT labels respectively, coincide closely with the fluorescent nuclei staining, as illustrated in the composite images (Fig. 4A). Furthermore, the F1 segmentation performance was similar for both cases, with no statistically significant difference in the segmentation scores (Fig. 4B). The capability to rapidly annotate image data while maintaining high performance is essential to expediting the workflow from image acquisition to software deployment.

To segment the exterior outlines of the cells, we used the automatic GT labeling workflow to train FCN models with images of two engineered cell types: (1) td-Tomatoexpressing MDA-MB231 breast cancer cells, and (2) enhanced green fluorescent protein (eGFP)-expressing iPSCs (see Fig. 4C). In this manner, we automatically generated approximately 3000 labels in a matter of minutes. We measured the overlap between the automatically generated GT labels with manual annotation and observed high segmentation performance (F1<sub>fluorescence</sub> = 0.82) (Fig. 4D). Likewise, the segmentation performance of the phase contrast predictions was high and similar for both cell types  $(F1_{ph-MDA-MB231} = 0.69 \text{ and } F1_{ph-iPSCs} = 0.70)$ . The high performance of FCN algorithms for segmenting cell outlines and nuclei is uniquely advantageous for the task of recognizing various types of cells for automated single-cell delivery. Importantly, high performance was maintained while processing phase contrast images, which is an essential capability for label-free cell manipulation protocols.

### Feature Extraction and Image Analysis

Automated extraction of features (e.g., shape, intensity, and texture features) from accurately segmented images enables a process for high-throughput analysis and screening of large populations of cells. Consequently, we implemented a feature extraction pipeline to measure various shape and intensity features from single cells. To test this capability, we quantified and compared the normalized distributions of distinct shape (i.e., area, circularity, and solidity) and intensity features (i.e., mean intensity) of the td-Tomato-MDA-MB231 cells and iPSCs (**Fig. 5A and Table S2**). Accordingly, both cell types have unique, distinct morphology and intensity characteristics. For example, MDA-MB231 cells are more circular and have a higher number of large cells

relative to the average compared to iPSCs. To visualize the relationship of each variable, we used a Spearman's correlation matrix and generated a heatmap to visualize the results (Fig. 5B). The correlation heatmap facilitates the analysis of single-cell data, in which one can reduce the parameter dimension by selecting one of the highly correlated features (i.e., mean and median intensity).<sup>31</sup> In one approach, known as cell painting, thousands of features are extracted from a single cell labeled with a variety of dyes, with each color targeting a particular intracellular region.<sup>32</sup> Computational image cytology methods provide a powerful framework for single-cell analysis. In the context of systems biology, for example, the NFP-E system can be used to study the inputoutput relationship of signal transduction, in which morphogenetic changes, induced from the delivery of exogenous cargoes, can be measured and analyzed with an object detection and feature extraction pipeline.<sup>33</sup>

To selectively transfect cells of interest in experiments, we implemented a GUI to filter cells based on intensity and shape features. For example, we found that electroporation efficiency is increased for well-adhered cells compared to rounded cells. Using this filtering scheme, we successfully selected cells with high transfection probability automatically. Moreover, to increase the system reliability and prevent engagement of pipette tip on false-positive locations, we filtered the segmented nuclei centroid with a mask generated from the cell outline segmentations. In this manner, we were able to measure detection performance (sensitivity and precision) and filter out potential false-positive locations. We compared the detection performance prior to filtering, and compared it to an ellipse-fit image-processing pipeline. Higher sensitivity and precision values were obtained with the FCN network for all cell types (Suppl. Table S1). Consistent with the literature reports, deep learning algorithms outperform traditional image-processing techniques in the field of computer vision.

To integrate the detection and feature extraction pipelines in a deployable manner that can be used without the need of specialized GPUs, we optimized the FCN models using inference engine acceleration routines. The inference engine optimization step stripped the trained FCN model of layers used exclusively for training (i.e., batch normalization) and incorporated libraries optimized for acceleration. In this manner, we were able to rapidly detect cells (in 1.25 s; Suppl. Fig. S5), extract features, and apply filters prior to single-cell electroporation. In comparison, the manual selection time of cells in a FOV (~80 cells) is in the range of 60-100s. This translates to approximately a 1h reduction in the electroporation workflow for processing 5000 cells. As a result of the increased throughput provided from the AI-driven automation platform, the NFP-E can be used for manipulating large numbers of cells, which is an essential capability for cell-engineering applications like gene editing and cell line generation.



**Figure 4.** Automatic label generation performance and cytosol segmentation. (a) Composite images of phase contrast (ph-c) HeLa cells with (**left**) Hoechst-stained nuclei micrograph, (**middle**) fully convolutional network (FCN) output predictions obtained using manual ground truth (GT) label inputs, and (**right**) FCN output predictions obtained using automatically generated GT label inputs. Scale bars = 20  $\mu$ m. (b) Bar graph of F1 segmentation scores [50% intersection over union (IOU)] calculated using fluorescence (Hoechst) inference outputs and ph-c predictions using manual or automated GT labels. (c) (**Left**) fluorescence micrographs [of td-Tomato-expressing MDA-MB231 cells (red) and enhanced green fluorescent protein (eGFP)-encoding induced pluripotent stem cell (iPSCs)] were fine-tuned to generate output labels of the cell cytosol (**Middle**: fluorescence predictions), and subsequently used as input labels for training ph-c micrographs. (**Right**) ph-c Predictions: Composite images of the input ph-c micrographs overlapped with the output labels (cyan) are shown. Scale bars = 100  $\mu$ m. (d) Bar graph of F1 segmentation scores (50% IOU) for the fluorescence and ph-c label outputs. The fluorescence model was trained using both cell types, and ph-c models were fine-tuned to specific cell types accordingly. Standard error: n = 3 for each case.



Figure 5. Cell feature extraction. (a) Violin plots of features extracted from the cytosol outlines of individual cells. (b) A Spearman's correlation matrix heatmap of shape and intensity features measured from individual cells (cytosol outlines).



**Figure 6.** (a) Delivery of fluorescent proteins into different cell types with automated nanofountain probe electroporation (NFP-E). (**Top**) Bovine serum albumin (BSA) tagged with red fluorophore (AF647) delivery into copGFP induced pluripotent stem cells (iPSCs); and (**middle and bottom**) BSA tagged with green fluorophore (AF488) delivered into (middle) MDA-MB231 and (bottom) HEK293 cells . (**b–d**) Automated delivery of Cas9–guide RNA (gRNA) complex into adhered iPSCs encoding for a green fluorescent protein (cop-GFP). (**b**, **Top**) Schematic of electroporation delivery and subsequent binding of gRNA to target copGFP gene and cleavage with Cas9 nuclease; (**middle**) fluorescence micrograph of copGFP iPSCs prior to electroporation; and (**bottom**) phase contrast micrograph used for fully convolutional network (FCN) detection. (**c**, **Top**) Schematic of copGFP gene knockout; (**middle**) fluorescence micrograph of negative control group 5 days after incubation; and (**bottom**) corresponding phase contrast micrograph. (**e**) Bar plot of delivery (BSA AF647) and knockout efficiencies in copGFP-iPSCs: standard error bars [N<sub>images</sub> = 4 (delivery) and 8 (knockout)]. All scale bars: 100 μm.

# Single-Cell Delivery and Cas9–gRNA Gene Knockout

Automated delivery of exogenous materials into single cells is the central feature of the NFP-E system. To measure the efficiency of delivery following automatic detection, we delivered a red fluorescent protein (BSA-AF647) into an engineered iPSC line containing a gene that encodes for a green fluorescent protein (copGFP), and green fluorescent BSA protein (AF488-BSA) into breast cancer and kidney cells (MDA-MB231 and HEK293) (Fig. 6A). Following optimization of the delivery conditions, we delivered Cas9gRNA complexes into copGFP-iPSCs to test our system's capabilities for gene-editing applications. The gRNA was designed to target the location of the copGFP gene in the iPSC chromosomes to enable specific removal of the gene with the Cas9 nuclease via the nonhomologous end-joining (NHEJ) mechanism (Fig. 6B-6D). iPSCs in the FOV were automatically detected with the AI software and electroporated with the NFP-E system, allowing for the efficient delivery of Cas9-gRNA complexes into populations of cells adhered on a gridded dish (~30 cells per grid FOV; see **Fig. 6B**). We used the NFP-E to deliver gene-editing complexes into ~200 cells in an automated manner. The cells were incubated post electroporation for 5 days and imaged. Successful knockout was observed in cells that no longer expressed the green fluorescent cop-GFP protein (**Fig. 6C**). No cells were edited in the control experiments (**Fig. 6D**). Using the NFP-E system, we successfully delivered proteins into 90% of cells and knocked out the copGFP gene in ~29% of cells (**Fig. 6E**). The editing efficiency was consistent throughout the different regions of the culture dish, indicating uniformity in the delivery conditions.

The uniform delivery and precise dosage capabilities provided by the NFP-E systems are ideally suited for the implementation and optimization of CRISPR-Cas9 editing workflows. Researchers have leveraged CRISPR-Cas9 technologies to edit genes via homology-directed repair (HDR), or knock out genes through NHEJ,<sup>34,35</sup> to edit pathogenic genes,<sup>36</sup> engineer novel cell lines,<sup>2,37</sup> study intracellular signaling,<sup>6,38</sup> perform functional genomic screening,<sup>39</sup> and develop cellular and animal models.<sup>40</sup> The efficiency of gene editing for both HDR and NHEJ remains low, however, due to complexities in the biochemical processes (e.g., dependency on cell cycle stage and off-target binding)<sup>35,41</sup> and inefficiencies in the methods used to deliver the gene-editing cargoes into the cell.<sup>13</sup> Consequently, increasing the precision, timing, and dosage control of the delivery mechanism is imperative to reduce the technical noise of the gene-editing workflow and elucidate the biological aspects of the experiments.<sup>8</sup> The automated NFP-E system provides these advantages. Furthermore, the nondestructive localized electroporation process allows for repeated editing, while the automated AI segmentation and feature extraction pipelines enable the systematic tracking of phenotypic changes incurred from the gene-editing treatments.

## Conclusion

Single-cell electroporation systems provide a gentle and efficient platform for executing cell manipulation and analysis tasks. Throughput is, however, a concern. When augmented by AI-enabled feature extraction algorithms, the system achieves the throughput needed in applications ranging from gene editing to cell line generation to cell analysis, in which high delivery efficiency, throughput, and dosage control are essential. It also allows for acquisition of a multitude of shape and intensity measurements from each cell, opening up new avenues for studies in the fields of systems biology, stem cell differentiation, and cell–cell communication. Lastly, the implemented cell-staining and FCN transferlearning approach can also be applied to other problems that necessitate the labeling of brightfield micrographs for subsequent deep learning training and cell phenotype recognition.

#### Author Contributions

H.D.E., C.A.P., P.M., and V.L. conceived the project. C.A.P. developed the computer vision routines. P.M. and C.A.P wrote the automation software. P.M., V.L., and N.P. performed the biological experiments. All authors analyzed and interpreted the data. H.D.E., C.A.P., and P.M. wrote the manuscript.

#### **Declaration of Conflicting Interests**

The authors declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: H.D.E. declares majority ownership of Infinitesimal LLC, a company commercializing biotools for cell manipulation and analysis.

#### Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Research reported in this publication was supported by the National Institutes of Health (NIH) SBIR Phase 1 award (no. 1R43GM128500-01) and NIH R21 award (no. 1R21GM132709-01). This work used the Argonne National Laboratory Center for Nanoscale Materials. Use of the Center for Nanoscale Materials, an Office of Science user facility, was supported by the US Department of Energy, Office of Science, Office of Basic Energy Sciences (under contract no. DE-AC02-06CH11357).

#### References

- Ran, F. A.; Hsu, P. D.; Wright, J.; et al. Genome Engineering Using the CRISPR-Cas9 System. *Nat. Protoc.* 2013, *8*, 2281–2308.
- Yang, R.; Lemaître, V.; Huang, C.; et al. Monoclonal Cell Line Generation and CRISPR/Cas9 Manipulation via Single-Cell Electroporation. *Small* 2018, 14, 1702495.
- Ishii, T.; Okahata, Y.; Sato, T. Mechanism of Cell Transfection with Plasmid/Chitosan Complexes. *BBA-Biomembranes* 2001, 1514, 51–64.
- Kang, W.; Yavari, F.; Minary-Jolandan, M.; et al. Nanofountain Probe Electroporation (NFP-E) of Single Cells. *Nano. Lett.* 2013, 13, 2448–2457.
- McNaughton, B. R.; Cronican, J. J.; Thompson, D. B.; et al. Mammalian Cell Penetration, siRNA Transfection, and DNA Transfection by Supercharged Proteins. *P. Natl. Acad. Sci.* USA 2009, 106, 6111–6116.
- Giraldo-Vela, J. P.; Kang, W.; McNaughton, R. L.; et al. Single-Cell Detection of mRNA Expression Using Nanofountain-Probe Electroporated Molecular Beacons. *Small* 2015, *11*, 2386–2391.
- Loh, O.; Lam, R.; Chen, M.; et al. Nanofountain-Probe-Based High-Resolution Patterning and Single-Cell Injection of Functionalized Nanodiamonds. *Small* 2009, *5*, 1667–1674.
- Kang, W.; McNaughton, R. L.; Espinosa, H. D. Micro- and Nanoscale Technologies for Delivery into Adherent Cells. *Trends. Biotechnol.* 2016, 34, 665–678.
- Mukherjee, P.; Nathamgari, S. S. P.; Kessler, J. A.; et al. Combined Numerical and Experimental Investigation of Localized Electroporation-Based Cell Transfection and Sampling. ACS Nano 2018, 12, 12118–12128.
- Canatella, P. J.; Karr, J. F.; Petros, J. A.; et al. Quantitative Study of Electroporation-Mediated Molecular Uptake and Cell Viability. *Biophys. J.* 2001, *80*, 755–764.
- Sharei, A.; Zoldan, J.; Adamo, A.; et al. A Vector-Free Microfluidic Platform for Intracellular Delivery. *P. Natl. Acad. Sci. USA* 2013, *110*, 2082–2087.
- Boukany, P. E.; Morss, A.; Liao, W.-C.; et al. Nanochannel Electroporation Delivers Precise Amounts of Biomolecules into Living Cells. *Nat. Nanotechnol.* 2011, *6*, 747–754.
- Liu, C.; Zhang, L.; Liu, H.; et al. Delivery Strategies of the CRISPR-Cas9 Gene-Editing System for Therapeutic Applications. J. Control. Rel. 2017, 266, 17–26.
- Stewart, M. P.; Langer, R.; Jensen, K. F. Intracellular Delivery by Membrane Disruption: Mechanisms, Strategies, and Concepts. *Chem. Rev.* 2018, *118*, 7409–7531.
- Yang, R.; Lemaître, V.; Huang, C.; et al. Monoclonal Cell Line Generation and CRISPR/Cas9 Manipulation via Single-Cell Electroporation. *Small* 2018, 14, 1702495.
- Sakaki, K.; Dechev, N.; Burke, R. D.; et al. Development of an Autonomous Biological Cell Manipulator with Single-Cell Electroporation and Visual Serving Capabilities. *IEEE T. Bio-Med. Eng.* 2009, 56, 2064–2074.
- Wang, W.; Liu, X.; Gelinas, D.; et al. A Fully Automated Robotic System for Microinjection of Zebrafish Embryos. *PLoS One* 2007, 2, e862–e862.

- Esmaeilsabzali, H.; Sakaki, K.; Dechev, N.; et al. Machine Vision-Based Localization of Nucleic and Cytoplasmic Injection Sites on Low-Contrast Adherent Cells. *Med. Biol. Eng. Comput.* 2012, 50, 11–21.
- McKinney, S. M.; Sieniek, M.; Godbole, V.; et al. International Evaluation of an AI System for Breast Cancer Screening. *Nature* 2020, 577, 89–94.
- Lawrence, S.; Giles, C. L.; Tsoi, A. C.; et al. Face Recognition: A Convolutional Neural-Network Approach. *IEEE T. Neural. Networ.* 1997, *8*, 98–113.
- Krizhevsky, A.; Sutskever, I.; Hinton, G. E. Imagenet Classification with Deep Convolutional Neural Networks. *Commun. ACM* 2017, 60, 84–90.
- Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv* 2014, preprint arXiv:1409.1556.
- Long, J.; Shelhamer, E.; Darrell, T. Fully Convolutional Networks for Semantic Segmentation. In *Proceedings of* the *IEEE Conference on Computer Vision and Pattern Recognition*, Boston, 2015; pp 3431–3440.
- Falk, T.; Mai, D.; Bensch, R.; et al. U-Net: Deep Learning for Cell Counting, Detection, and Morphometry. *Nat. Methods* 2019, *16*, 67.
- Caicedo, J. C.; Goodman, A.; Karhohs, K. W.; et al. Nucleus Segmentation across Imaging Experiments: The 2018 Data Science Bowl. *Nat. Methods* 2019, *16*, 1247–1253.
- Johnson, J. W. Adapting Mask-RCNN for Automatic Nucleus Segmentation. *arXiv* 2018, preprint arXiv:1805.00500.
- 27. He, K.; Gkioxari, G.; Dollár, P.; et al. Mask r-cnn. In *Proceedings of the IEEE International Conference on Computer Vision*, 2017; pp 2961–2969.
- Lin, T.-Y.; Dollár, P.; Girshick, R.; et al. Feature Pyramid Networks for Object Detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017; pp 2117–2125.
- Kohoutek, J. S.; Prinz, J. Electroporation System with Micromanipulator and Probe. Patent No. 20190388675. 2019.

- Carpenter, A. E.; Jones, T. R.; Lamprecht, M. R.; et al. CellProfiler: Image Analysis Software for Identifying and Quantifying Cell Phenotypes. *Genome Biol.* 2006, 7, R100.
- Caicedo, J. C.; Cooper, S.; Heigwer, F.; et al. Data-Analysis Strategies for Image-Based Cell Profiling. *Nat. Methods* 2017, 14, 849.
- Bray, M.-A.; Singh, S.; Han, H.; et al. Cell Painting, a High-Content Image-Based Assay for Morphological Profiling Using Multiplexed Fluorescent Dyes. *Nat. Protoc.* 2016, *11*, 1757.
- Handly, L. N.; Yao, J.; Wollman, R. Signal Transduction at the Single-Cell Level: Approaches to Study the Dynamic Nature of Signaling Networks. *J. Mol. Biol.* 2016, 428, 3669–3682.
- Cong, L.; Ran, F. A.; Cox, D.; et al. Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science* 2013, 339, 819–823.
- Lin, S.; Staahl, B. T.; Alla, R. K.; et al. Enhanced Homology-Directed Human Genome Engineering by Controlled Timing of CRISPR/Cas9 Delivery. *eLife* 2014, *3*, e04766.
- Ma, H.; Marti-Gutierrez, N.; Park, S.-W.; et al. Correction of a Pathogenic Gene Mutation in Human Embryos. *Nature* 2017, 548, 413–419.
- 37. Grobarczyk, B.; Franco, B.; Hanon, K.; et al. Generation of Isogenic Human iPS Cell Line Precisely Corrected by Genome Editing Using the CRISPR/Cas9 System. *Stem Cell Rev. Rep.* 2015, 11, 774–787.
- Chi, S.; Weiss, A.; Wang, H. A CRISPR-Based Toolbox for Studying T Cell Signal Transduction. *Biomed. Res. Int.* 2016, 2016, 5052369.
- Shalem, O.; Sanjana, N. E.; Hartenian, E.; et al. Genome-Scale CRISPR-Cas9 Knockout Screening in Human Cells. *Science* 2014, 343, 84–87.
- Hsu, P. D.; Lander, E. S.; Zhang, F. Development and Applications of CRISPR-Cas9 for Genome Engineering. *Cell* 2014, 157, 1262–1278.
- Doench, J. G.; Fusi, N.; Sullender, M.; et al. Optimized sgRNA Design to Maximize Activity and Minimize Off-Target Effects of CRISPR-Cas9. *Nat. Biotechnol.* 2016, *34*, 184–191.