An artificial intelligent platform for live cell identification and the detection of cross-contamination

Ruixin Wang[1#], Dongni Wang[1#], Dekai Kang[2#], Xusen Guo[2], Chong Guo[1], Meimei Dongye[1], Yi Zhu[1,3], Chuan Chen[1,3], Xiayin Zhang[1], Erping Long[1], Xiaohang Wu[1], Zhenzhen Liu[1], Duoru Lin[1], Jinghui Wang[1], Kai Huang[2], Haotian Lin[1]

[1]State Key Laboratory of Ophthalmology, Zhongshan Ophthalmic Center, Sun Yat-sen University, Guangzhou, China; [2]Key Laboratory of Machine Intelligence and Advanced Computing, Ministry of Education School of Data and Computer Science, Sun Yat-Sen University, Guangzhou, China; [3]Department of Molecular and Cellular Pharmacology, University of Miami Miller School of Medicine, Miami, FL, USA

Contributions: (I) Conception and design: H Lin, K Huang, R Wang, D Wang; (II) Administrative support: None; (III) Provision of study materials or patients: None; (IV) Collection and assembly of data: R Wang, D Wang, D Kang, X Guo, C Guo, X Zhang, E Long, M Dongye, X Wu, Z Liu, D Lin, J Wang; (V) Data analysis and interpretation: R Wang, D Kang, X Guo; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

[#]These authors contributed equally to this work.

Correspondence to: Prof. Haotian Lin. Director of Department of Artificial Intelligence and Big Data, State Key Laboratory of Ophthalmology, Zhongshan Ophthalmic Center, Sun Yat-sen University, Guangzhou, China. Email: gddlht@aliyun.com; Prof. Kai Huang. Key Laboratory of Machine Intelligence and Advanced Computing, Ministry of Education School of Data and Computer Science, Sun Yat-Sen University, Guangzhou, China. Email: huangk36@mail.sysu.edu.cn.

Background: About 30% of cell lines have been cellular cross-contaminated and misidentification, which can result in invalidated experimental results and unusable therapeutic products. Cell morphology under the microscope was observed routinely, and further DNA sequencing analysis was performed periodically to verify cell line identity, but the sequencing analysis was costly, time-consuming, and labor intensive. The purpose of this study was to construct a novel artificial intelligence (AI) technology for "cell face" recognition, in which can predict DNA-level identification labels only using cell images.

Methods: Seven commonly used cell lines were cultured and co-cultured in pairs (totally 8 categories) to simulated the situation of pure and cross-contaminated cells. The microscopy images were obtained and labeled of cell types by the result of short tandem repeat profiling. About 2 million patch images were used for model training and testing. AlexNet was used to demonstrate the effectiveness of convolutional neural network (CNN) in cell classification. To further improve the feasibility of detecting cross-contamination, the bilinear network for fine-grained identification was constructed. The specificity, sensitivity, and accuracy of the model were tested separately by external validation. Finally, the cell semantic segmentation was conducted by DilatedNet.

Results: The cell texture and density were the influencing factors that can be better recognized by the bilinear convolutional neural network (BCNN) comparing to AlexNet. The BCNN achieved 99.5% accuracy in identifying seven pure cell lines and 86.3% accuracy for detecting cross-contamination (mixing two of the seven cell lines). DilatedNet was applied to the semantic segment for analyzing in single-cell level and achieved an accuracy of 98.2%.

Conclusions: The deep CNN model proposed in this study has the ability to recognize small differences in cell morphology, and achieved high classification accuracy.

Keywords: Cell authentification; biomedical optical imaging; image classification; neural networks

Introduction

Cell line cross-contamination and misidentification have become serious problems since the first cross-contaminated HeLa cell line was reported in the 1950s (1,2). To date, one-fifth to one-third of cell lines have been cross-contaminated, which is commonly due to mislabeling, the repeated use of pipets, and the sharing of culture media (3-5). Cell line cross-contamination is often not readily detectable, resulting in inaccurate experimental results and unusable therapeutic products. The International Cell Line Authentication Committee, the National Institutes of Health, and many authoritative journals strongly urge researchers to authenticate cultured cells before carrying out experiments (6-8).

Various molecular approaches have been developed to distinguish different cell lines with respective strengths and limitations (3,9). Currently, the most frequently recommended method is short tandem repeat (STR) profiling, which detects variations in the number of STR sequences within microsatellite DNA (10). Each of the repeat regions is amplified and compared with the standardized cell line profiles. In general, 80% similarity is the threshold for declaring a match (10,11). However, the application of STR is limited by its high cost and availability only in specialized institutions. STR profiling is only suitable for distinguishing cell lines of a single species and is affected by genetic drift (12). Karyotyping (13) and polymerase chain reaction (14) can separate species as supplemental techniques. Single-nucleotide polymorphism and whole-genome sequencing are less affected by genetic drift but are very expensive (15). Moreover, complex expertise is required for each technology. In contrast, observation of morphology is a simple and direct technique to characterize cell lines. Cultured cells can be morphologically divided into three basic categories: fibroblast-like (elongated), epithelial-like (polygonal), and lymphoblast-like (spherical). However, whether cell morphology can be used to separate individual cell lines for cell identification is unclear.

Deep convolutional neural networks (CNNs) are widely used in modern artificial intelligence (AI), and have been successfully applied in medical image classification, including the identification of skin cancer (16) and retinopathy (17). Additionally, CNNs can analyze live cell images for mitosis detection (18) and hematopoietic lineage tracing (19). Moreover, their superiority in semantic segmentation that outputs pixel-wise predictions increase the precision of image recognition (20). CNNs provide new insights into individual cell lines identification and cross-contamination detection by morphological identification.

In this study, we sought to classify cell lines with similar morphologies and to identify cross-contaminated cells by a novel AI platform with three deep learning networks, AlexNet, bilinear CNN (BCNN), and DilatedNet. Specifically, we obtained microscopy images of seven commonly used cell lines to train these deep learning models, and applied eight categories of co-cultured cell lines (mixing two of the seven) for model testing.

Methods

Cell culture and image acquisition

Strict cell culture practices were utilized to prevent accidental co-culture among these seven cell lines, as follows: (I) handling one cell line at a time; (II) keeping each medium in a separate container; (III) carefully labeling each culture flask with the cell line name, passage number, and operation date; and (IV) discarding pipettes in a timely manner after each operation. The conditional medium that was used is provided in *Table S1*. Cells were seeded from low to intermediate density and were cultured in a humidified atmosphere of 5% CO_2 at 37 ℃. Continuous passage-culture was conducted for 10 passages. Images were acquired using an inverted phase-contrast microscope (Carl Zeiss Axio Vert A1). Each image was obtained at 50× optical magnification with a size of 1,040×1,388 pixels. The STR analysis was conducted by Sun Yat-sen University Forensic Medical Center (*Figure S1*). Eight types of cell mixtures were co-cultured as indicated in *Table S1*. The cells were seeded at a density in a mixture ratio of 1:10 to 1:1,000. Images were taken every day.

Preprocessing

The preprocessing of cell microscopy images mainly focused on improving the image quality and reducing the effect of noise. First, Gauss filtering was performed. Then, to avoid the impact of an imbalanced background light on the prediction results and to improve the robustness of the method, we employed gray normalization for brightness balance and contrast enhancement of the cell microscopy images using the following formula:

$$I_{out} = \frac{STD_{out}}{STD_{in}}(I_{in} - Mean_{in}) + Mean_{out} \qquad [1]$$

Where I_{in} is the gray value of the pixel in the input image, I_{out} is the gray value of the pixel in the output image, $Mean_{in}$ and STD_{in} are the mean and standard deviation, respectively, of the input image, and $Mean_{out}$ and STD_{out} are the mean and standard deviation of the image after normalization. In this work, we set $Mean_{out}$ and STD_{out} to 180 and 20, respectively.

Data augmentation

Data augmentation is essential for teaching the network the desired invariance and robustness properties. For microscopy images, we primarily need cell size invariance and robustness against gray value variations. We employed two distinct forms of data augmentation. The first form of data augmentation is cell image scaling, which can ensure that the model obtains a certain degree of cell size invariance. The second form of data augmentation is gamma correction, as shown in the following formula; the model can learn the illumination invariance by adjusting the brightness of the original training set image to different ranges.

$$s = cr^{\gamma} \tag{2}$$

Where r is the input image gray level, s is the output image gray level, and c and γ are constants.

Cell density effect

Cell morphology will differ when the density of cells varies. This variation could affect the performance of the model. Thus, an effective method for measuring cell density is necessary. For the gray-normalized cell microscopy images, we employed the adaptive threshold processing method and morphological operations to generate a cell region mask image. Each pixel had a value of 0 or 1, with 0 indicating that the pixel was the background and 1 indicating the cell region. The cell density was then calculated by the proportion of the cell regions as the following formula.

$$Density = \frac{S_{cell}}{S_{total}} \tag{3}$$

To explore the effect of cell density on the performance of the model, we partitioned the dataset using a fixed cell density threshold. We then, chose the same numbers of high- and low-density cell images to make up the training dataset. We used high- and low-density cell microscopy images to train the models, with the other parameters

remaining exactly the same. We then tested the performance of the high- and low-density models on test sets with all densities. The effect of cell density on the accuracy of the model was evaluated.

AlexNet

We used AlexNet as the CNN model for initial cell classification. AlexNet is a pioneering deep CNN that won the ILSVRC-2012 with a top-5 test accuracy of 84.6%. The architecture is relatively simple and consists of 5 convolutional layers, max-pooling layers, rectified linear units (ReLUs) as non-linearities, three fully connected layers, and a dropout. *Figure S2* shows the CNN architecture.

Bilinear CNN model

The BCNN (21) for image classification consisted of a quadruple $B = (f_A; f_B; P; C)$. Here, f_A and f_B are feature functions based on CNNs, P is a pooling function, and C is a classification function. A feature function is a mapping, which takes an image and a location L and outputs a feature of size $K \times D$. The location can include position and scale. The feature outputs are combined at each location using the matrix outer product. That is, the bilinear combination of f_A and f_B at location l is given by

$$bilinear\,(l,I,f_A,f_B) = f_A(l, I)^T f_B(l, I) \tag{4}$$

In this work, we used VGG16 truncated at relu5_3 (22) as the feature function. The input image block size was 192×192 pixels, and proceeding through VGG16 produced 12×12×512-pixel features. Then, the feature map was reshaped to 144×512 pixels. The feature vector was multiplied by its transposition using the outer product at each location of the cell image to obtain the bilinear vector. The bilinear vectors were fed to a two-layer fully connected layer of dimension 1,024 and 7 (cell categories) respectively.

Dilated network

Dilated (23) convolutions are a generalization of Kronecker-factored convolutional filters, which support exponentially expanding receptive fields without losing resolution.

Let $F: Z^2 \rightarrow R$ be a discrete function. Let $\Omega_r = [-r, r]^2 \cap Z^2$ and let $k: \Omega_r \rightarrow R$ be a discrete filter of size $(2r+1)^2$. The discrete convolution operator * can be defined as

Page 4 of 13

Wang et al. Cell face recognition: a technology for cellular cross-contamination

$$(F * k)(p) = \sum_{s+t=p} F(s)k(t) \tag{5}$$

We now generalize this operator. Let l be a dilation factor, and let $*l$ be defined as

$$(F *_l k)(p) = \sum_{s+lt=p} F(s)k(t) \tag{6}$$

We will refer to $*l$ as a dilated convolution or a $l-$ dilated convolution. The familiar discrete convolution $*$ is simply the $1-$ dilated convolution.

Results

Image acquistion, preprocessing and augmentation

Seven commonly used cell lines, including epithelial cells (SRA01/04, ARPE19, and Hkb20), fibroblasts (HFL1), nerve cells (661W), and tumor cells (HuH-7 and T24), were cultured and verified by STR analysis (*Figure S1*). Pure-cell images were taken from independently cultured cell lines (a total of 1,823 images) for model training, which was conducted by six-fold cross-validation. Pure-cell images as well as images from eight categories of co-cultured cell lines (mixing two of the seven and a total of 848 images) were used for model testing (*Table S1*). All images were captured with 50× optical magnification at a size of 1,040×1,388 pixels.

To improve image quality and reduce noise, we performed gray normalization and contrast enhancement (see Methods). Given the potential variations in size and background brightness, we performed gamma correction and imaging scaling on the captured pure-cell images before model training (*Figures 1A,B*). After preprocessing and data augmentation, we used a sliding window (192×192 pixels, with a 64-pixel striding step) to scan the pure-cell images and segment them into patches for model training (*Figure 1C*). The patches were classified and assigned a corresponding color. To determine the final classification of the raw image data, we returned the pseudo-colored patches to their original locations and generated a heat map. The most prevalent patch was regarded as the final classification of the entire image (*Figure 1D*). This process increased the number of training sets to 333,862–362,247 pieces per cell and to a total of 2,464,090 pieces for all seven cell lines (*Figure 1E*). The risk of overfitting was also reduced by this data expansion approach (24).

Classification of pure-cell images using AlexNet

The application of AlexNet (25) using the CNN method has increased the accuracy of traditional machine learning models from 70% to 80%. Compared to the other networks with hundreds to thousands layers, AlexNet contains only 8 layers, resulting in the network's ease of operation and minimal calculation requirements (*Figure S2*).

First, we trained the AlexNet classifier to identify the seven pure cell lines. The agreement between the ground truth and the test results was 98.9% in the test set (the mean of all classes), and the confusion matrix is shown in *Figure S3*. After returning the pseudo-colored patches to their original locations, most pure-cell images for testing resulted in a homochromatic heat map with their corresponding colors. The classification result of an entire image is shown in *Figure 2* and *Figure S4*.

The shape and size of the cells can be affected by cell density. To test whether cell density affects the accuracy of the model, we implemented pure cell line images with low- and high-cell density for AlexNet-based model training. The low-density cell model performed well with the low-cell-density (0–50% confluence) part of the validation set (average accuracy of 94.74%), and the high-density model performed well in the high-cell-density (50–100% confluence) part of the validation set (average accuracy of 97.23%). The accuracy of the high-density model on cells with lower than 20% confluence (accuracy of 58.77%) was higher than that of the low-density model on cells with higher than 80% confluence (accuracy of 25.3%) (*Figure 3A*). These results suggest that high- and low-density cell images had different features, that images of both cell densities are necessary for model training, and that AlexNet performed better with high-cell-density images. This was likely because high-density images contained more cells in each patch. Moreover, high-density cells were not simply closer together but rather presented a texture that may be a key feature the AI extracted (26,27).

The size of the sliding window can also affect the accuracy of the model, as a larger window will contain a larger number of cells with a more complex texture of patches. Therefore, we tested the model classification accuracy with window sizes of 64×64, 128×128, 192×192, 256×256, and 320×320 pixels. The accuracy of the model was greater than 97% with a window size of 192×192 pixels, and 99% with a window size of 320×320 pixels, indicating
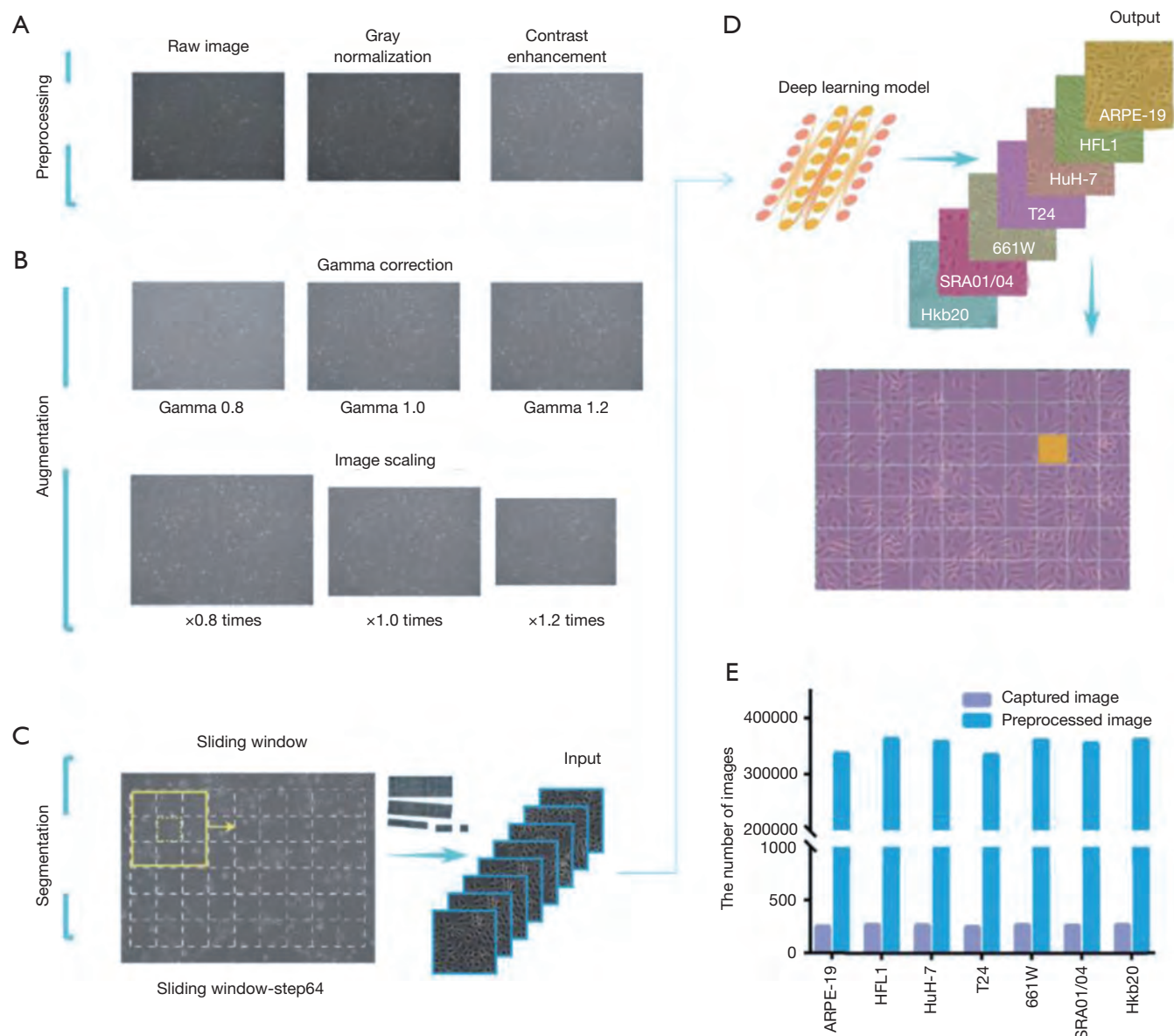
Figure 1 Training set augmentation and segmentation. (A) Raw images of seven cell lines were captured by the microscopy imaging system, with an image size of 1,388×1,040 pixels, each with 50-fold optical enlargement. Before training, gray normalization and contrast enhancement were conducted as preprocessing steps. (B) Gamma correction and image scaling were then used to augment the images. (C) A sliding window was applied to segment entire images into small patches, which were then ready to train a deep learning model. (D) The output patches with corresponding colors were combined to produce a heat map as the classification result of the entire image. (E) After augmentation and segmentation, the number of original training images was extended from 247–268 pieces to 333,862–362,247 pieces per cell.

Page 6 of 13

Wang et al. Cell face recognition: a technology for cellular cross-contamination
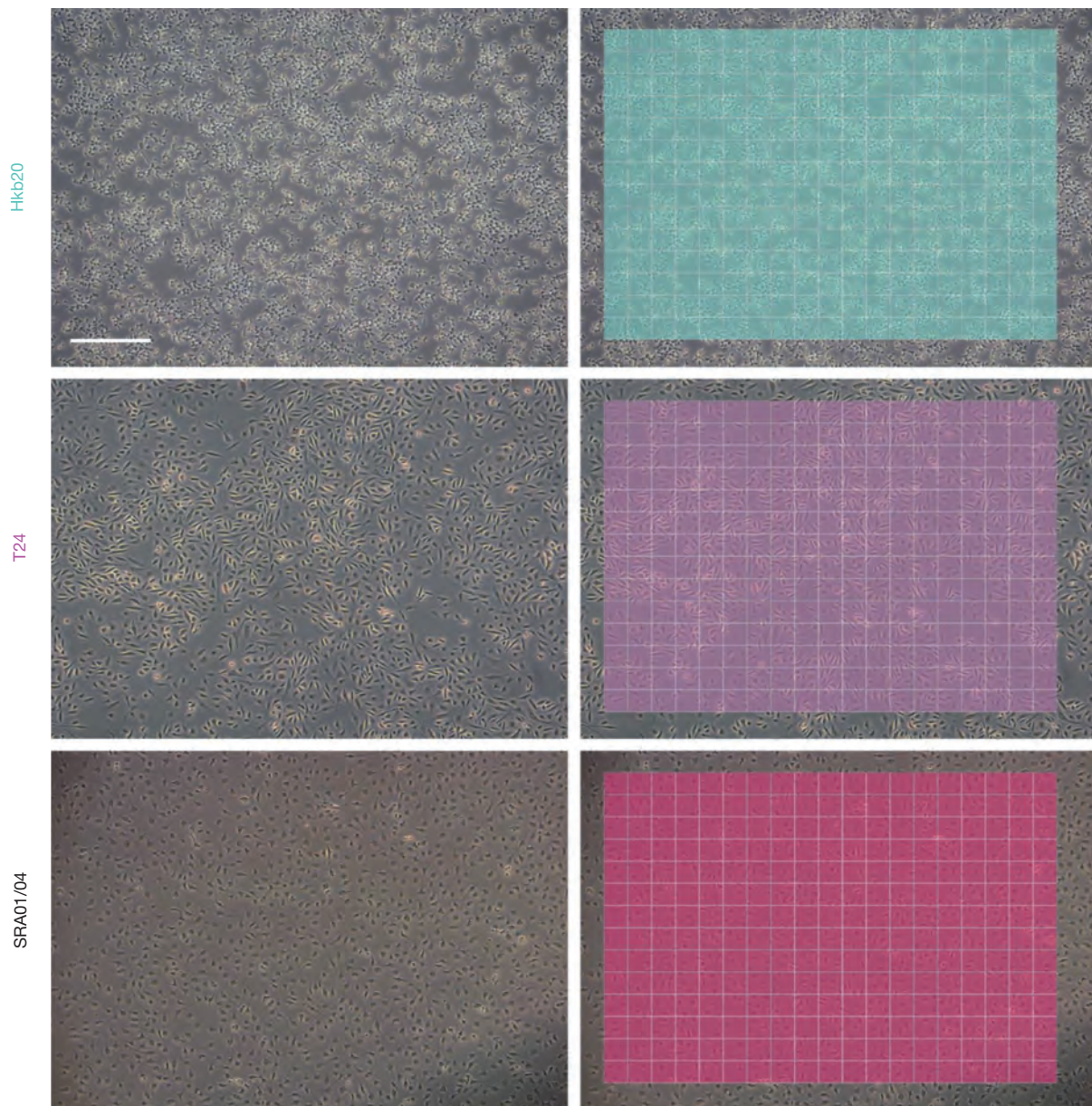
Figure 2 Using AlexNet for pure-cell image classification. Pure-cell images were classified using a convolutional network called AlexNet. The result was an entire image consisting of small patches of certain colors, which corresponded to the cell type that AlexNet identified in that patch. The accuracy of pure-cell patches for AlexNet was 98.9%. Scale bar, 100 μm. See also *Figure S4*.

that the accuracy of classification improved as the window size increased (*Figure 3B,C*).

Detection of cross-contamination and identification of cell types using multiple models

Eight data sets of images with two different co-cultured

cell types were obtained. The first step for the AI model was to determine whether the cells were pure or mixed. After scanning, the entire image was divided into patches (a heat map). We took the maximum proportion possessed by any patch type as the probability of a pure cell population (*Figure 4A*). To get an appropriate threshold, we tested on both pure-cell and contaminated-cell datasets and drew
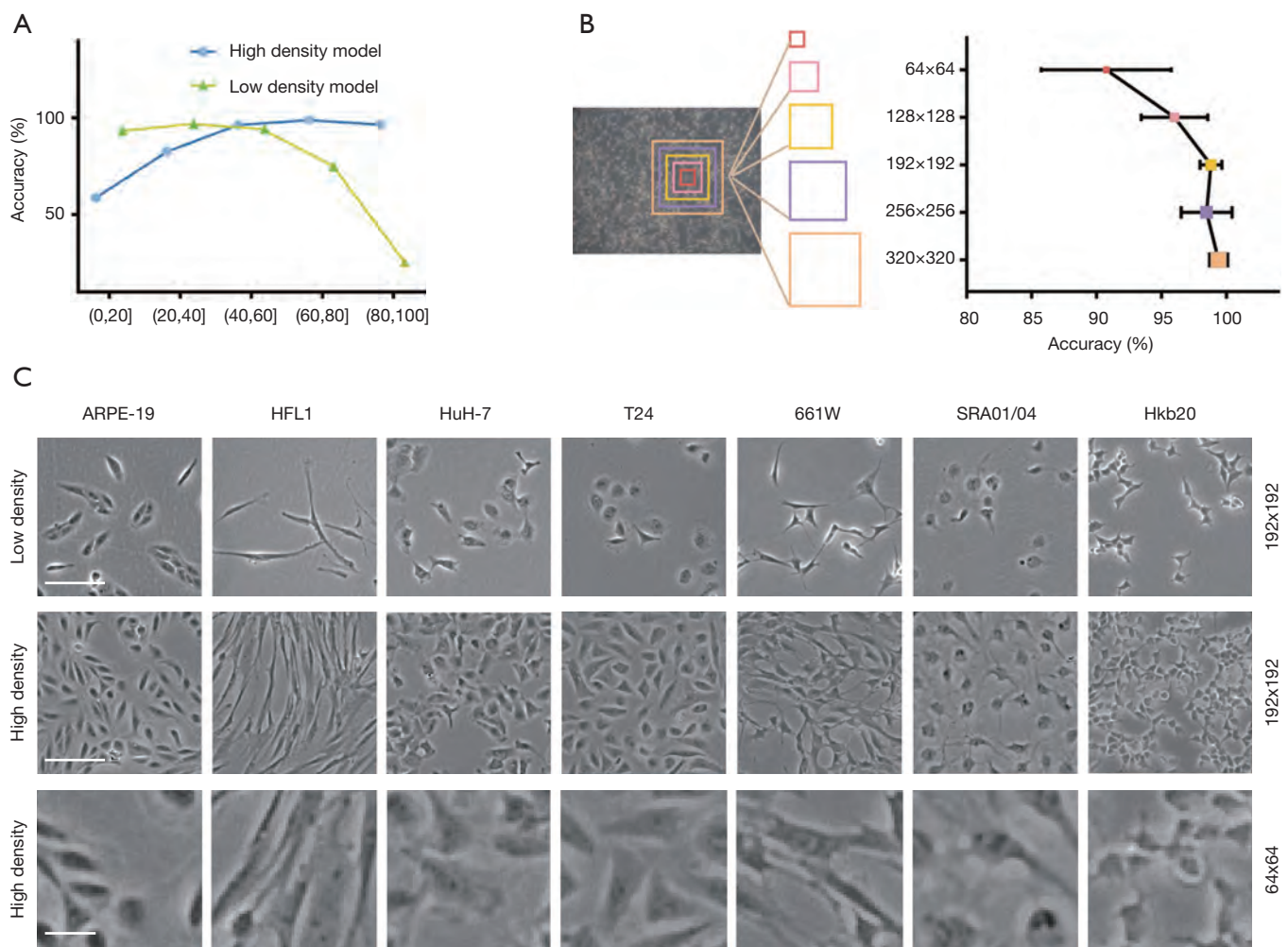
Figure 3 Exploring the potential factors influencing the accuracy of AlexNet. (A) We selected only low- and high-cell-density images to train the AlexNet model separately. We tested the model using the whole-density dataset. The results showed that the low-density cell model performed well for the low-density (0–50% confluence) portion of the validation set, whereas the high-cell-density model performed well for the high-density (50–100% confluence) portion of the validation set. The high-density model was less accurate for the 0–10%-confluent portions than the low-density model was for the 90–100%-confluent portions. (B) Different sizes of the segmented window were used to crop the captured images; thus, a gradient of image sizes was input into the classifier. Finally, the results showed that increasing the window size improved the accuracy. (C) By comparing the middle row with the upper or lower row, high-density images were found to have more features, such as complex textures and more cells, which was consistent with the large-window patches. Scale bar, 25 μm (up and middle row), 8 μm (low row).

a receiver operating characteristic (ROC) map, which showed that the classifier had excellent performance, with an area under the ROC curve (AUC) of 0.9872. According to the ROC curve, we set the classification threshold of 97.37%, with the true positive rate of 90.86% and the false positive rate of 2.54% (*Figure 4A*). If the maximum proportion of the resulting color was greater than the threshold, we directly used the maximum patch result as the cell identity and compared this result with the original

cell tag to determine whether the label was correct. If the sample was contaminated, the potential type of candidate cell was suggested according to the classification result. We calculated the identification accuracy as A or B in the mixed-cell (A + B) test set (*Figure S7*) and found an average accuracy of 67.66%, indicating that AlexNet could not accurately differentiate mixed cells.

Since the identification of mixed cells was based on the classification of pure cells, we further improved the
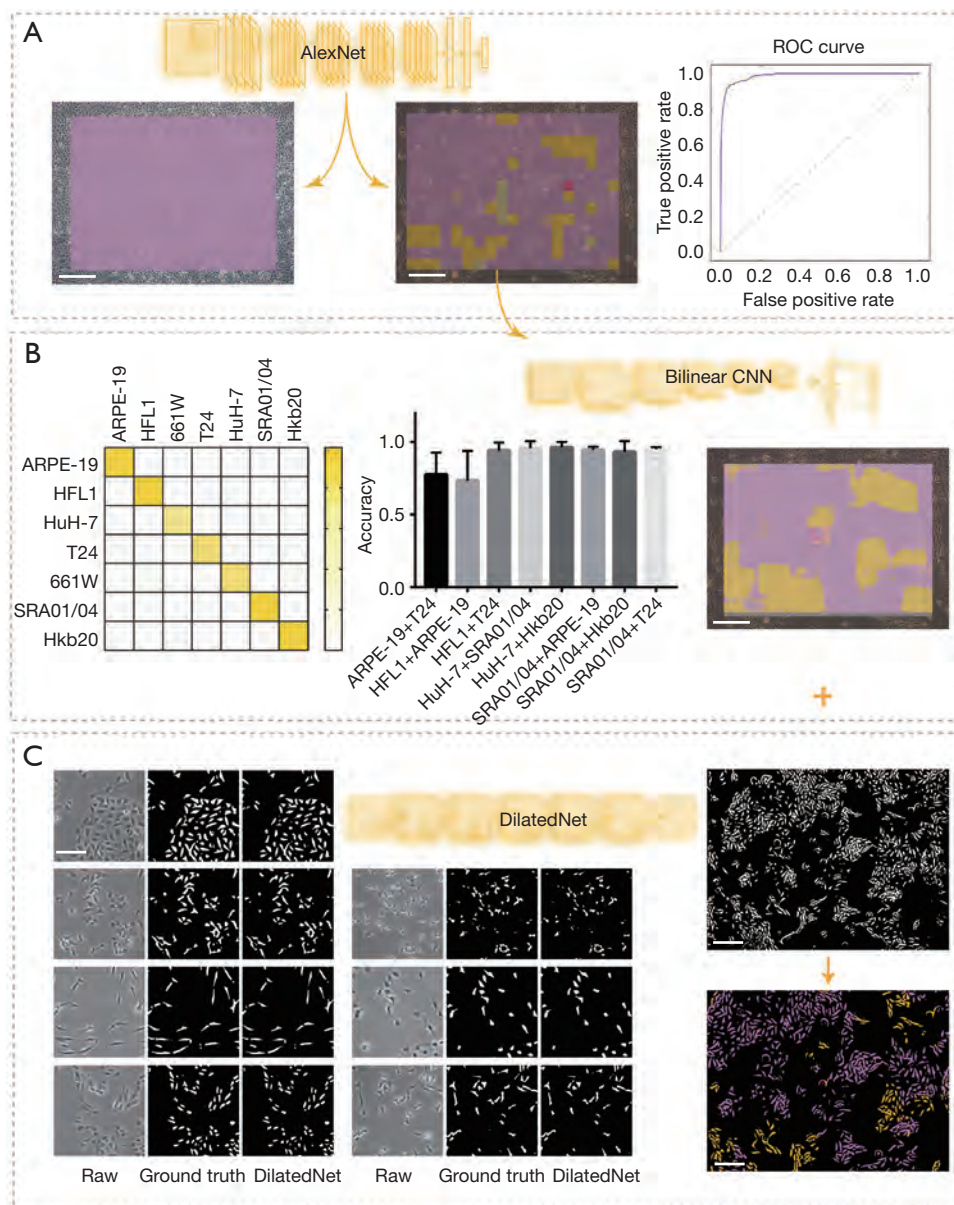
Page 8 of 13

Wang et al. Cell face recognition: a technology for cellular cross-contamination

Figure 4 Detection of contamination and identification of cell types using multiple models. (A) A brief network architecture diagram is shown as the yellow background of each model. For more details, see also *Figures S2,S5,S6*. To determine whether a cell image was contaminated, we took the maximum proportion possessed by a patch type as the probability of a pure-cell population (orange of left panel, purple of right panel). Right panel, receiver operating characteristic (ROC) curve for pure cells and mixed cells showed that the classifier had excellent performance, with an area under the ROC curve (AUC) of 0.9872. The plot of the ROC curve represents a true positive rate of 90.86%, and a false positive rate of 2.54%. Scale bar, 100 μm. (B) The bilinear convolutional network was applied to further differentiate mixed cells. Left panel, the confusion matrix displays the matches between the ground truth and the test results for the pure-cell test dataset, with an average accuracy of 99.5%±0.7% (mean ± SD), and the accuracy of testing for each category of mixed-cell images is shown in the middle panel. The output of bilinear convolutional neural network (BCNN) in the testing of mixed-cell images significantly reduced the areas of misidentification (compare to a-middle panel). Right panel, sliding step 1 was set to improve the visualization resolution. Scale bar, 100 μm. (C) DilatedNet was applied to segment cells into single cells and separate them from background. Left and middle panels, the size of the patches was 256×256, which was zoomed to 512×512. The output of DilatedNet is considerably similar to the ground truth. Right upper panel, the patches were spliced together to produce an image representing a single-cell location. Right lower panel, after merging the classification result and segmented result, the cross-contamination could be detected. Scale bar, 25 μm (left and middle row), 100 μm (right row).

accuracy of the model in recognizing the seven pure-cell types. Actually, cell type classification is a type of fine-grained image classification, which aims at recognizing subcategories within a same basic category, such as granite and marble. The texture features of images are the key factors in fine-grained identification (28), which is also consistent with our analysis of image features by density and window-size testing above. Recently, Lin *et al.* proposed an end-to-end architecture for fine-grained visual recognition called BCNN, which achieved the best classification accuracy for the CUB200-2011 dataset among weak supervised fine-grained classification models (21). The extracted features can be presented as an outer product of two features, f(A) and f(B), extracted from a standard CNN.

A pre-trained visual geometry group network (VGG-16) was taken as a standard CNN. Before the fully connected layer, a specially designed bilinear vector was obtained via multiplication using the outer product from the concatenation of the last feature maps and its transposition. The architecture of BCNN is based on Tensorflow (*Figure S5*). Training was performed by stochastic gradient descent (SGD) with a mini-batch size of 16 and momentum of 0.9. In total, 100,000 iterations were performed with a learning rate of 10-4. We performed tests with the pure-cell dataset and the mixed-cell dataset. The average accuracy of the BCNN model (pre-trained VGG-16 with a bilinear vector) for the pure-cell test dataset was 99.5%, and the relative confusion matrix for the pure-cell test set is shown in *Figure 4B*. The BCNN model achieved a mean accuracy of 86.26% for the mixed-cell test dataset, which was significantly higher than that of AlexNet (*Figure 4B, Figure S8*). Moreover, the pre-trained VGG-16 achieved 74.51% mean accuracy on the mixed-cell dataset. To improve the visualization resolution, we modified the step size of the sliding window from 64 to 1 to locate the edges, which represent the output with 64×64 or 1×1 in the center of each sliding window; however, the average accuracy did not significantly change (*Figure 4B, Figure S9*).

Smantic segmentation algorithm was applied to distinguish cells from background and divide them into single cells with clear boundaries. Conventional semantic segmentation networks such as fully convolutional networks (FCN) (29), SegNet (30) and U-net (31) integrate multi-scale contextual information via successive pooling and upsampling layers that reduce resolution. However, in our case, the features of edges, which are small but important for cell segmentation, might be lost because of pooling and upsampling. Dilated convolution has an unchanged kernel with an exponentially increased receptive field. We constructed a 10-layer neural network called DilatedNet based on dilation convolution that only contained convolution layers without pooling or upsampling (*Figures S6,S10*). We then implemented DilatedNet using Tensorflow and Keras and labeled 1,200 single cell images with 512×512 pixels. The cell images and labels were divided into a training set, a validation set, and a test set in a 4:1:1 ratio. Model training was performed by Adam (32) with a mini batch size of 5. In total, 100,000 iterations were performed, with a learning rate of 1e-4. DilatedNet achieved an accuracy of 98.2% on the test set. The segmentation result is shown in *Figure 4C*. Then, we merged the cell types (the classification result using BCNN) and cell locations (the segmentation result using DilatedNet) as the final representation of the mixed-cell data set (*Figure 5, Figure S5*). Therefore, cross-contamination can be shown at the individual-cell level.

Discussion

This study is the first to detect cross-contaminated or misidentified cell lines using a deep learning technique combining AlexNet, BCNN and DilatedNet. AlexNet achieved an AUC of 0.9872 for detecting contamination. BCNN achieved an accuracy of the model for classification of 99.5% for the seven pure cell lines and an accuracy of 86.3% for the eight categories of co-cultured cell lines (mixing two of the seven). DilatedNet was applied to the semantic segment for each single cell with precise edges and achieved an accuracy of 98.2% on the test set. We also characterized how CNN differentiated cell images with similar morphologies by the density test and the window-size test. Our results suggest that despite the single-cell size and shape characteristics, textures formed by growth are also important morphological features. Our findings also help improve the classification through the fine-grained identification network.

Cell lines undergo genetic drift with continued passaging in culture due to changes in allele frequencies that occur in random sampling of the parent. Genetic drift may lead to deletion or mutation of certain alleles and may be accentuated by overpassaging or overdiluting, especially for malignant cell lines (3). This cell line instability can lead to changes in STR markers such as satellites, reducing the specificity of STR analysis. As a result, STR sets a threshold of 80% to allow a certain degree of genetic drift. Several studies have found that cell lines at high passage numbers
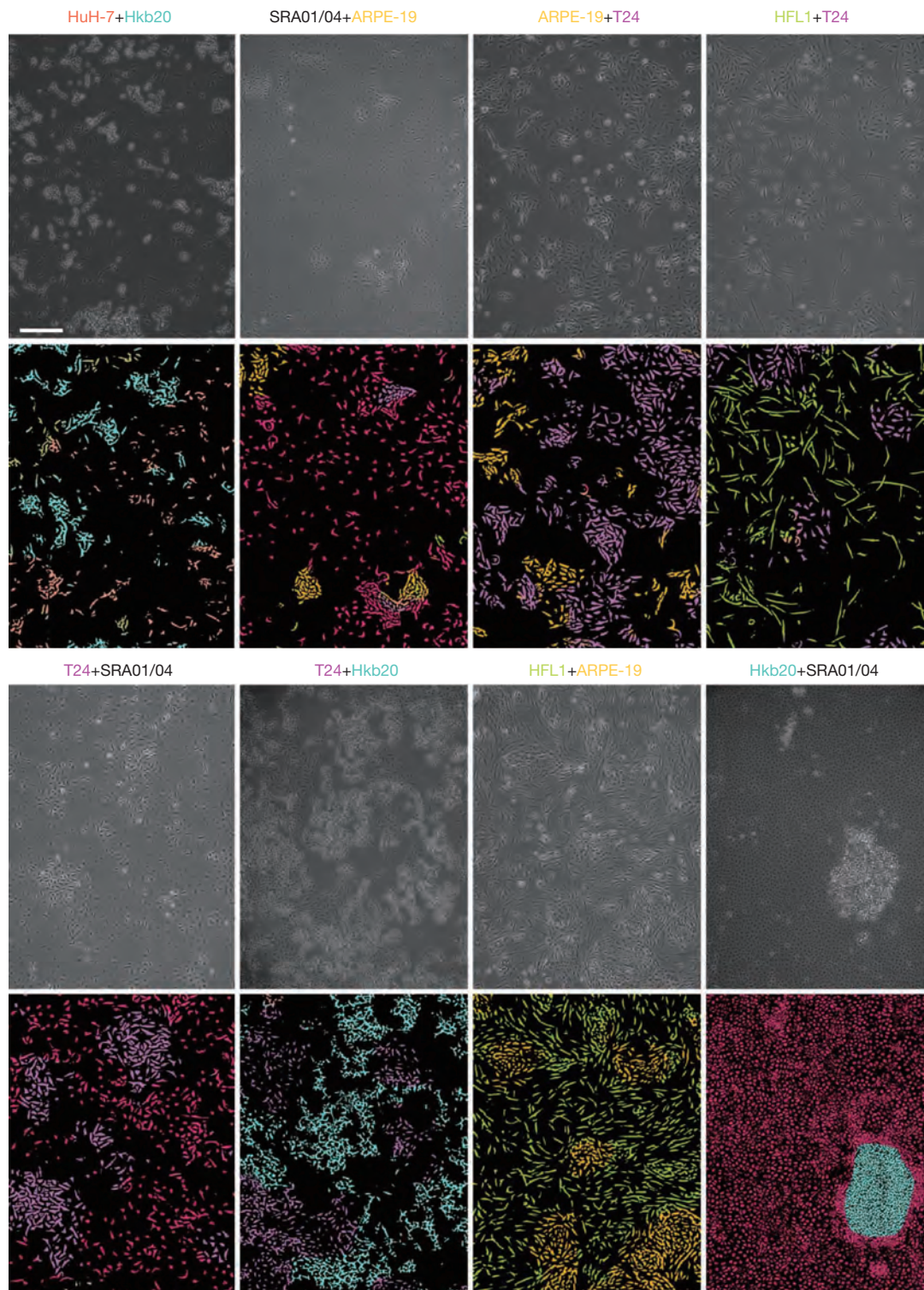
Page 10 of 13

Wang et al. Cell face recognition: a technology for cellular cross-contamination

Figure 5 Combining bilinear convolutional neural network (BCNN) and DilatedNet for mixed-cell image identification. The text colors represent the true types of mixed cells. The pseudo images display the final results for mixed-cell image identification, based on the combination of BCNN and DilatedNet for both classification and segmentation. The AI platform successfully output the truest labels at the single-cell level. Scale bar, 100 μm. See also *Figures S8,S11*.

experience morphological changes (33). The SNP test, which is less affected by genetic drift (34), can be applied to determine whether a cell line will spontaneously change its identity during long-term culture. Subsequently, we can collect cell images before and after an identity change to train a more accurate AI platform.

Studying one disease *in vitro* often involves only a few cell lines. This study successfully completed the test of seven types of cell lines that are suitable for specialist laboratories. In the future, modules that are classified according to anatomical systems can be built and tested in specific research areas. In addition, this AI system has basically covered all current ophthalmic cell lines.

Online detection and cloud updates may provide better user experience and system optimization. After uploading the images to be detected to the website, the test result can be directly obtained, and the uploaded images will be used for a new round of training. Moreover, the exclusive AI identification system can be customized, especially in the case of newly established or purchased cell lines. It was important to note that new cell types need to be certified by STR before training the customized model. Because cell images have similar underlying features, custom models will be trained more efficiently by transfer-learning methods based on existing models. For cell banks that manage thousands of cells, completing a large number of classifications at the same time may require more powerful hardware to support the computation of large data, which requires more cooperation to achieve. In the other side, collecting thousands of types of cell images is an irreplaceable advantage of the cell banks.

We demonstrated that BCNN is particularly useful for fine-grained classification on when testing mixed-cell images, which can model local pairwise feature interactions in a transnationally invariant manner. In contrast, the traditional classification neural networks, such as AlexNet, are designed for classification of large objects (e.g., birds, flowers, and dogs) and may not be suitable for small objects with distinct textural features. The main idea of BCNNs is the bilinear vector, the possible explanation of its "black box" is two mutually coordinated components: f(A) targeted the location of objects, whereas f(B) extracts their features. In addition, BCNN was based on a pre-trained VGG-16, one of the most widely used CNNs. Compared with AlexNet, VGG has a deeper architecture. Based on transfer learning, a pre-trained VGG-16 can be used rather than training a completely blank network.

Proximity to the image input layer corresponds to more extracted features shared in common, whereas the deeper layers extract more advanced, abstract, and task-specific features. Therefore, we can keep the parameters of the shallow layers unchanged and retrain the deeper layers to solve our own classification tasks. With this fine-tuning approach, we can save more computational power and achieve decent prediction accuracy while using a larger network. In this study, the pre-trained VGG-16 increased the accuracy for AlexNet significantly by 6.85%. Based on the pre-trained VGG, the bilinear vector contributed an additional 12% accuracy.

The classification results of mixed-cell images could not be shown as individual cells using the BCNN, so a dilated convolutional network was used to separate the cells. In addition, we tried some semantic segmentation networks used for natural scenes (U-net, Seg-net, and FCN) to achieve segmentation and classification in one step. However, these semantic segmentation networks did not perform well in our dataset, exhibiting a relatively high loss rate. Their purpose is to achieve segmentation and classification in one step, but they did not perform well in this dataset. The probable reason was that these networks often show a loss of edge in objects. When using individual cells for training, loss of edges is not acceptable for such tiny cells.

In summary, we successfully demonstrated that CNNs can be trained by images of pure cell lines to determine misidentification and cross-contamination of live cultured cell lines. This AI system requires no reagents and can be performed in house and in real time. The system is readily available, allowing faster popularization, and can provide wider coverage with powerful efficacy. However, further work is necessary to determine the feasibility of applying this model to authenticate thousands of cell lines.

Acknowledgments

Footnote

Provenance and Peer Review: This article was commissioned by the Guest Editors (Haotian Lin and Limin Yu) for the series "Medical Artificial Intelligent Research" published in *Annals of Translational Medicine*. The article was sent for external peer review organized by the Guest Editors and the editorial office.

Page 12 of 13

Wang et al. Cell face recognition: a technology for cellular cross-contamination

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at http://dx.doi.org/10.21037/atm.2019.07.105). The series "Medical Artificial Intelligent Research" was commissioned by the editorial office without any funding or sponsorship. HL served as the unpaid Guest Editor of the series. The other authors have no other conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: https://creativecommons.org/licenses/by-nc-nd/4.0/.

References

1. Gartler SM. Apparent HeLa Cell Contamination of Human Heteroploid Cell Lines. Nature 1968;217:750-1.
2. Lande R. Natural Selection and Random Genetic Drift in Phenotypic Evolution. Evolution 1976;30:314-34.
3. Capes-Davis A, Theodosopoulos G, Atkin I, et al. Check your cultures! A list of cross-contaminated or misidentified cell lines. Int J Cancer 2010;127:1-8.
4. Neimark J. Line of attack. Science 2015;347:938-40.
5. Horbach SPJM, Halffman W. The Ghosts of HeLa: How cell line misidentification contaminates the scientific literature. PLoS One 2017;12:e0186281.
6. Lorsch JR, Collins FS, Lippincott-Schwartz J. Fixing problems with cell lines. Science 2014;346:1452-3.
7. Masters JR. Cell-line authentication: End the scandal of false cell lines. Nature 2012;492:186.
8. Nardone RM. Eradication of cross-contaminated cell lines: A call for action. Cell Biol Toxicol 2007;23:367-72.
9. Almeida JL, Cole KD, Plant AL. Standards for Cell Line Authentication and Beyond. Plos Biol 2016;14:e1002476.
10. Masters JR, Thomson JA, Daly-Burns B, et al. From the Cover: Short tandem repeat profiling provides an international reference standard for human cell lines. Proc Natl Acad Sci 2001;98:8012-7.
11. Lorenzi PL, Reinhold WC, Varma S, et al. DNA fingerprinting of the NCI-60 cell line panel. Mol Cancer Ther 2009;8:713-24.
12. Poetsch M, Petersmann A, Woenckhaus C, et al. Evaluation of allelic alterations in short tandem repeats in different kinds of solid tumors—possible pitfalls in forensic casework. Forensic Sci Int 2004;145:1-6.
13. MacLeod RA, Kaufmann M, Drexler HG. Cytogenetic harvesting of commonly used tumor cell lines. Nat Protoc 2007;2:372-82.
14. Liu M, Liu H, Tang X, et al. Rapid identification and authentication of closely related animal cell culture by polymerase chain reaction. In Vitro Cell Dev Biol Anim 2008;44:224-7.
15. Didion JP, Buus RJ, Naghashfar Z, et al. SNP array profiling of mouse cell lines identifies their strains of origin and reveals cross-contamination and widespread aneuploidy. BMC Genomics 2014;15:847.
16. Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature 2017;542:115-8.
17. Gulshan V, Peng L, Coram M, et al. Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. JAMA 2016;316:2402-10.
18. Held M, Schmitz MHA, Fischer B, et al. CellCognition: time-resolved phenotype annotation in high-throughput live cell imaging. Nat Methods 2010;7:747-54.
19. Buggenthin F, Buettner F, Hoppe PS, et al. Prospective identification of hematopoietic lineage choice by deep learning. Nat Methods 2017;14:403-6.
20. Hao X, Zhang G, Ma S. Deep Learning. IEEE Trans Pattern Anal Mach Intell 2016;10:417-39.
21. Lin TY, Roychowdhury A, Maji S. Bilinear CNNs for Fine-grained Visual Recognition. In ICCV, 2015.
22. Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition. In ICLR, 2015.
23. Noh H, Hong S, Han B. Learning deconvolution network for semantic segmentation. Proc IEEE Int Conf Comput Vis 2015;1520-8.
24. Guo J, Gould S. Deep CNN ensemble with data augmentation for object detection. arXiv preprint arXiv:1506.07224, 2015.
25. Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks. In NIPS, 2012.
26. Haralick RM, Shanmugam K, Dinstein IH. Textural

Features for Image Classification. IEEE Trans Syst Man and Cybernet 1973;3:610-21.

27. Xie Y, Wang J. Study on the identification of the wood surface defects based on texture features. Optik - Int J Light Electron Opt 2015;126:2231-5.

28. Lowe DG. Object recognition from local scale-invariant features. ICCV 1999;99:1150-7.

29. Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. In CVPR, 2015: 3431-40.

30. Badrinarayanan V, Kendall A, Cipolla R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE Trans Pattern Anal Mach Intell 2017;39:2481-95.

31. Ronneberger O, Fischer P, Brox T. editors. U-Net: Convolutional Networks for Biomedical Image Segmentation. International Conference on Medical Image Computing & Computer-assisted Intervention; 2015.

32. Kingma D, Ba J. Adam: A Method for Stochastic Optimization. Comput Sci 2014;43:87-4.

33. Hughes P, Marshall D, Reid Y, et al. The costs of using unauthenticated, over-passaged cell lines: how much more data do we need? Biotechniques 2007;43:575, 577-8, 581-2 passim.

34. Castro F, Dirks WG, Fähnrich S, et al. High-throughput SNP-based authentication of human cell lines. Int J Cancer 2013;132:308-14.

Table S1 Cell-culture conditions and the numbers of images

Cell category	Culture media	FBS (%)	SP (%)	No. of pictures
ARPE-19	DMEM	10	1	249
HFL1	F12k	10	1	268
HuH-7	DMEM	10	1	264
T24	1640	10	1	247
661W	DMEM	10	1	266
SRA01/04	DMEM	10	1	262
Hkb20	DMEM	10	1	267
ARPE-19+T24	DMEM	10	1	154
HFL1+ARPE-19	DMEM	10	1	152
HFL1+T24	DMEM	10	1	156
HuH-7+SRA01/04	DMEM	10	1	134
HuH-7+Hrb20	DMEM	10	1	182
SRA01/04+ARPE-19	DMEM	10	1	157
SRA01/04+Hkb20	DMEM	10	1	112
SRA01/04+T24	DMEM	10	1	108

ARPE-19, adult retinal pigment epithelial cell line; HFL1, fetal lung fibroblast cell line; HuH-7, hepatocellular carcinoma cell line; T24, bladder carcinoma cell line; 661W, retinal photoreceptor derived cell line; SRA01/04, lens epithelial cells; Hkb20, embryonic kidney cell line; DMEM, Dulbecco's modified eagle medium; F12K, Kaighn's modification of Ham's F-12 medium; FBS, fetal bovine serum.
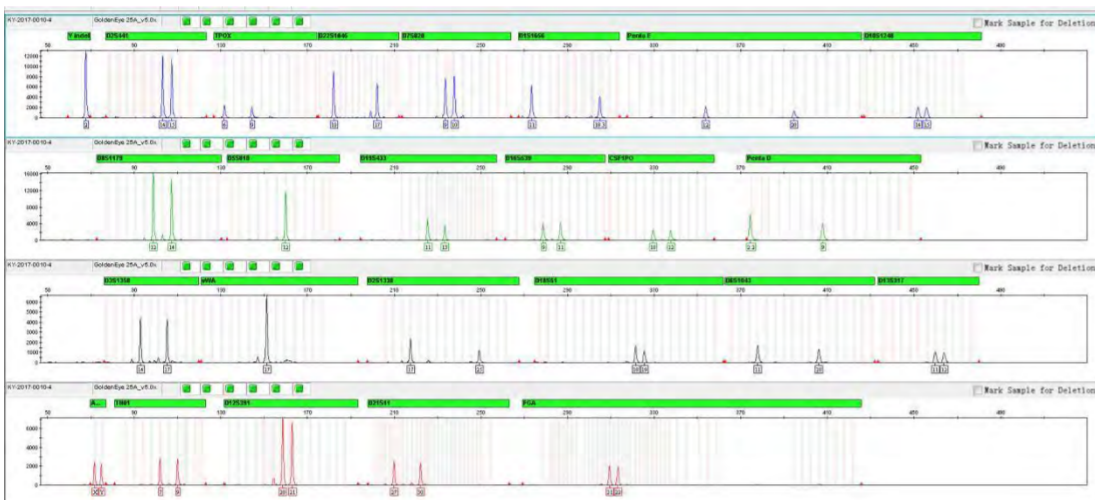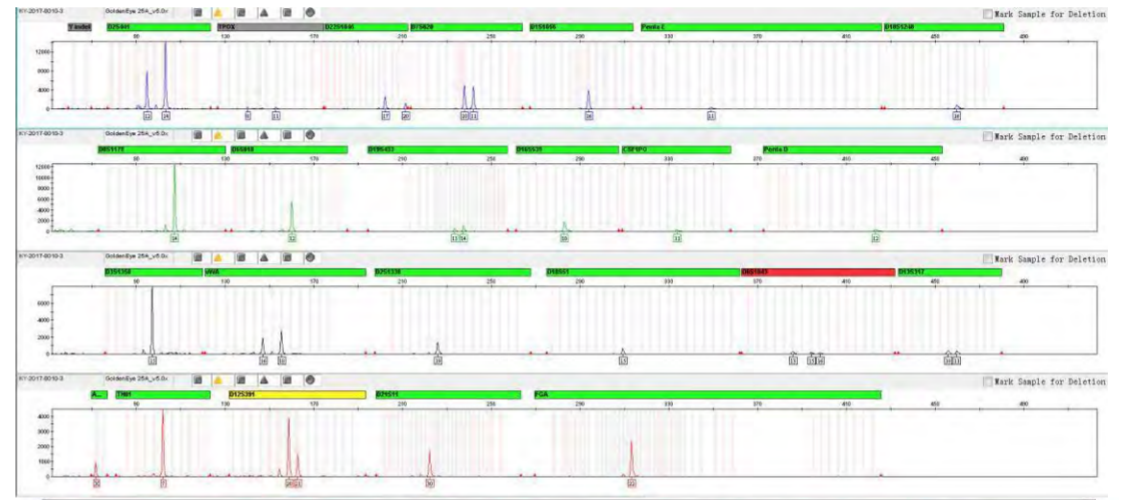
Hkb20
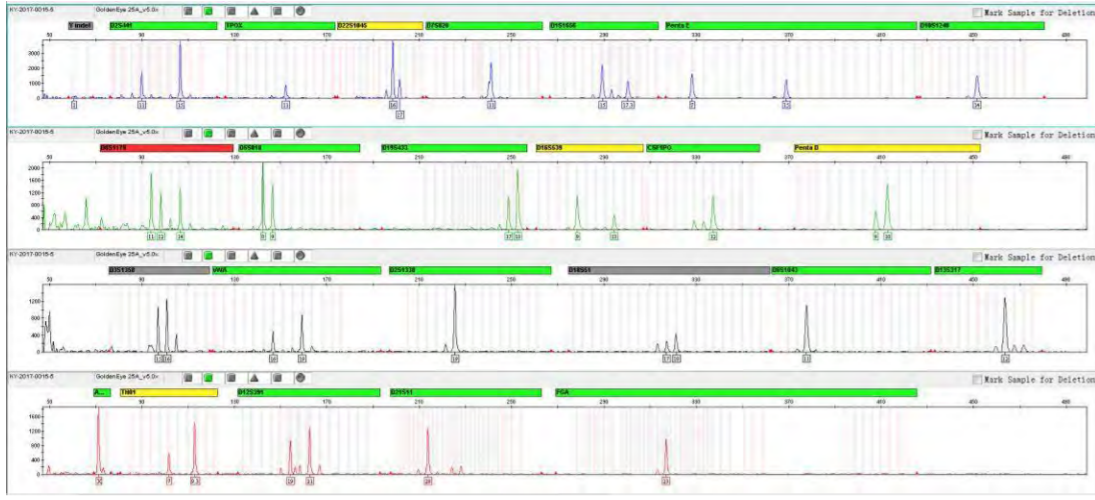
ARPE-19

Result of STR matching analysis by your data.
— DSMZ Profile Database —

A graphical presentation is shown at the bottom of this page.

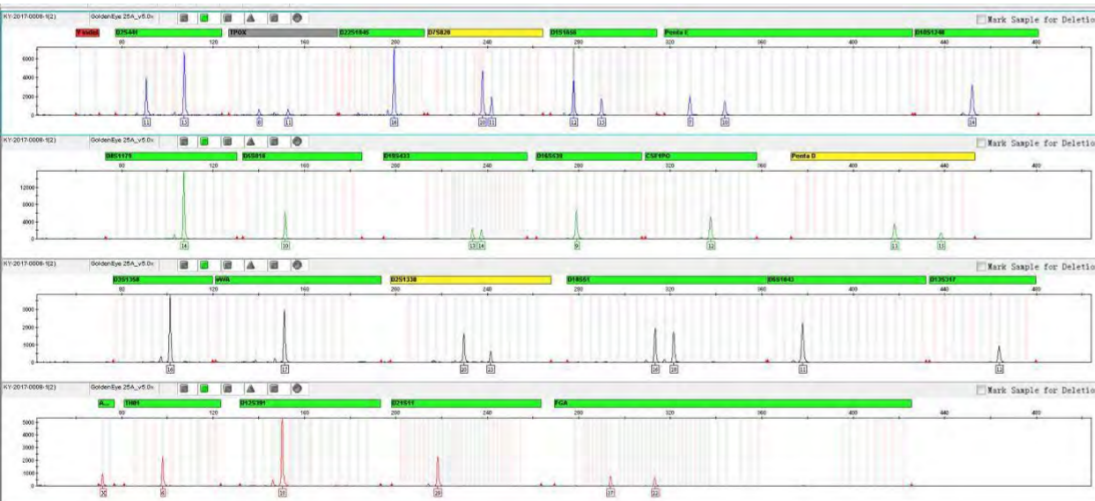| EV | Cell No. | Cell name | Locus names | | | | | | | | | Figures |
			D5S818	D13S317	D7S820	D16S539	VWA	TH01	AM	TPOX	CSF1PO	
	Query (Your Cell)		8,9	12,12	11,11	9,13	16,19	7,9.3	x,x	11,11	12,12	
0.94(34/36)	RCB2253	Hkb20	8,9	12,12	11,11	9,13	16,19	7,9.3	X,X	11,11	12,12	-
0.89(32/36)	635	293T	8,9	12,14	11,11	9,13	16,19	7,9.3	X,X	11,11	11,12	-
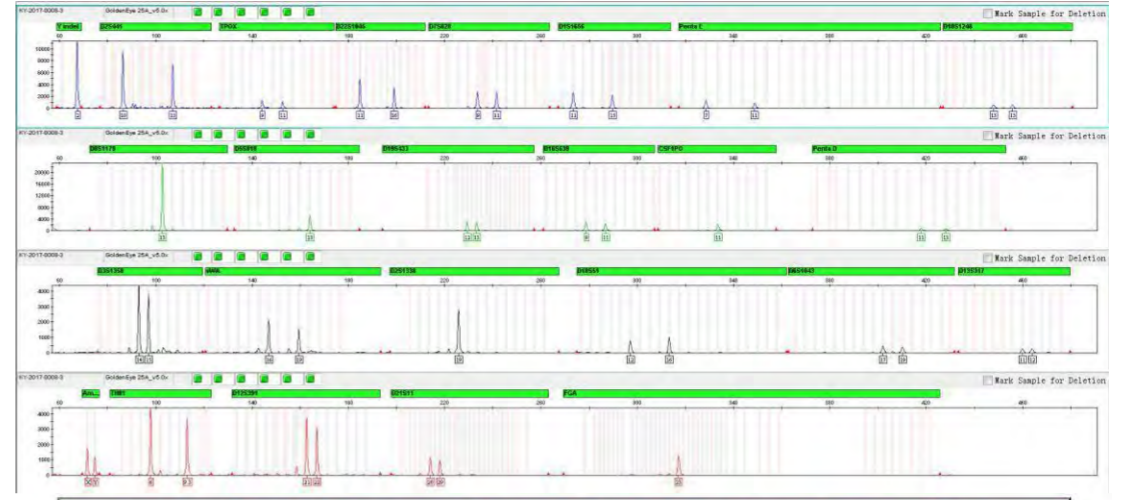
Result of STR matching analysis by your data.
— DSMZ Profile Database —

A graphical presentation is shown at the bottom of this page.

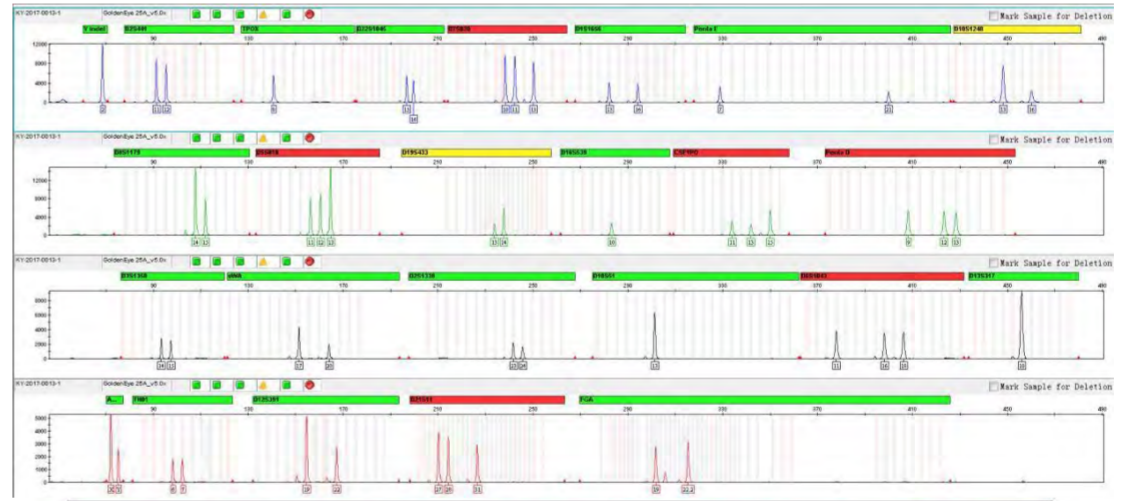| EV | Cell No. | Cell name | Locus names | | | | | | | | | Figures |
			D5S818	D13S317	D7S820	D16S539	VWA	TH01	AM	TPOX	CSF1PO	
	Query (Your Cell)		13,13	11,12	9,11	9,11	16,19	6,9.3	x,y	9,11	11,11	
1.00(36/36)	CRL-2302	ARPE-19	13,13	11,12	9,11	9,11	16,19	6,9.3	X,Y	9,11	11,11	-
1.00(36/36)	CRL-2502	ARPE-19/HPV-16	13,13	11,12	9,11	9,11	16,19	6,9.3	X,Y	9,11	11,11	-

HFL1

HuH-7

Result of STR matching analysis by your data.
— DSMZ Profile Database —

A graphical presentation is shown at the bottom of this page.

EV	Cell No.	Cell name	Locus names									Figures
			D5S818	D13S317	D7S820	D16S539	VWA	TH01	AM	TPOX	CSF1PO	
	Query (Your Cell)		12,12	11,12	9,10	9,11	17,17	7,9	x,y	6,9	10,12	
1.00(36/36)	CCL-153	HFL1	12,12	11,12	9,10	9,11	17,17	7,9	X,Y	6,9	10,12	-
1.00(36/36)	IFO50074	HFL1	12,12	11,12	9,10	9,11	17,17	7,9	X,Y	6,9	10,12	-

Result of STR matching analysis by your data.
— DSMZ Profile Database —

A graphical presentation is shown at the bottom of this page.

EV	Cell No.	Cell name	Locus names									Figures
			D5S818	D13S317	D7S820	D16S539	VWA	TH01	AM	TPOX	CSF1PO	
	Query (Your Cell)		12,12	10,11	10,11	10,10	16,18	7,7	x,x	8,11	11,11	
0.94(34/36)	JCRB0403	HuH-7	12,12	10,11	11,11	10,10	16,18	7,7	X,X	8,11	11,11	-
0.89(32/36)	RCB1366	HuH-7	12,12	10,10	11,11	10,10	16,18	7,7	X,X	8,11	11,11	-

T24

SRA01/04

Result of STR matching analysis by your data.
— DSMZ Profile Database —

A graphical presentation is shown at the bottom of this page.

EV	Cell No.	Cell name	Locus names									Figures
			D5S818	D13S317	D7S820	D16S539	VWA	TH01	AM	TPOX	CSF1PO	
	Query (Your Cell)		10,10	12,12	10,11	9,9	17,17	6,6	x,x	8,11	12,12	
1.00(36/36)	310	ECV-304	10,10	12,12	10,11	9,9	17,17	6,6	X,X	8,11	12,12	-
1.00(36/36)	CRL-1998	ECV-304	10,10	12,12	10,11	9,9	17,17	6,6	X,X	8,11	12,12	-
1.00(36/36)	CRL-3023	HBMEC-2	10,10	12,12	10,11	9,9	17,17	6,6	X,X	8,11	12,12	-
1.00(36/36)	JCRB0744	ECV304	10,10	12,12	10,11	9,9	17,17	6,6	X,X	8,11	12,12	-
1.00(36/36)	RCB0431	T24	10,10	12,12	10,11	9,9	17,17	6,6	X,X	8,11	12,12	-
0.89(32/36)	HTB-4	T24	10,12	12,12	10,11	9,9	17,17	6,6	X,X	8,11	10,12	-

Result of STR matching analysis by your data.
— DSMZ Profile Database —

A graphical presentation is shown at the bottom of this page.

EV	Cell No.	Cell name	Locus names									Figures
			D5S818	D13S317	D7S820	D16S539	VWA	TH01	AM	TPOX	CSF1PO	
	Query (Your Cell)		11,12,13	10,10	10,11,13	10,10	17,20	6,7	x,y	8,8	11,13,15	
1.00(42/42)	RCB1591	SRA 01/04	11,12,13	10,10	10,11,13	10,10	17,20	6,7	X,Y	8,8	11,13,15	-
0.67(26/39)	642	U-2973	11,12	10,11	10,11	11,13	17,20	7,9.3	X,Y	8,8	11,11	-

Figure S1 Short tandem repeat (STR) analysis of human cell lines. The STR analysis was conducted by Sun Yat-sen University Forensic Medical Center. All of the EVs are ≥0.94.
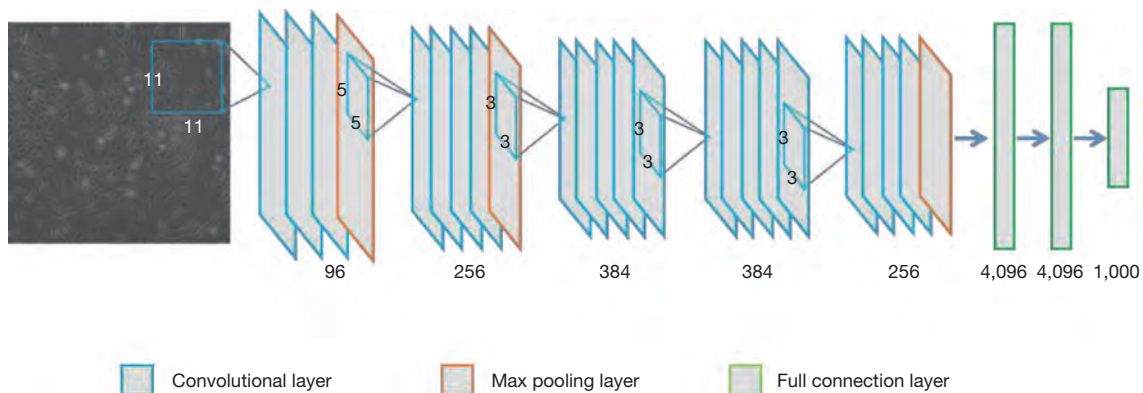
Figure S2 AlexNet framework. We used AlexNet as a neural network model for cell classification. Its structure includes five convolutions, the largest pooling layer and linear rectifier unit (ReLUs), three fully connected layers and a drop (dropout) layer (25).
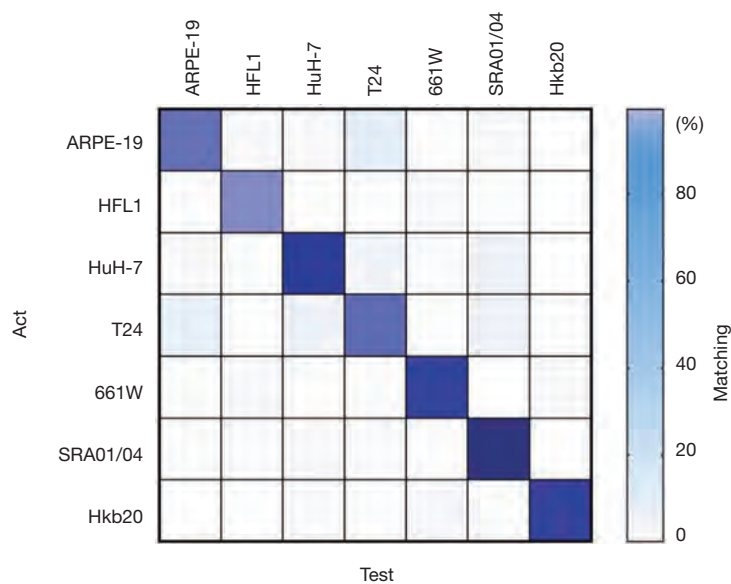

Figure S3 Confusion matrix of AlexNet. The segmented patches were used to train an AlexNet. The confusion matrix displayed the matches between the ground truth and the test results in the test dataset, with values ranging from 97.76% to 99.59% and an accuracy of 98.9%.
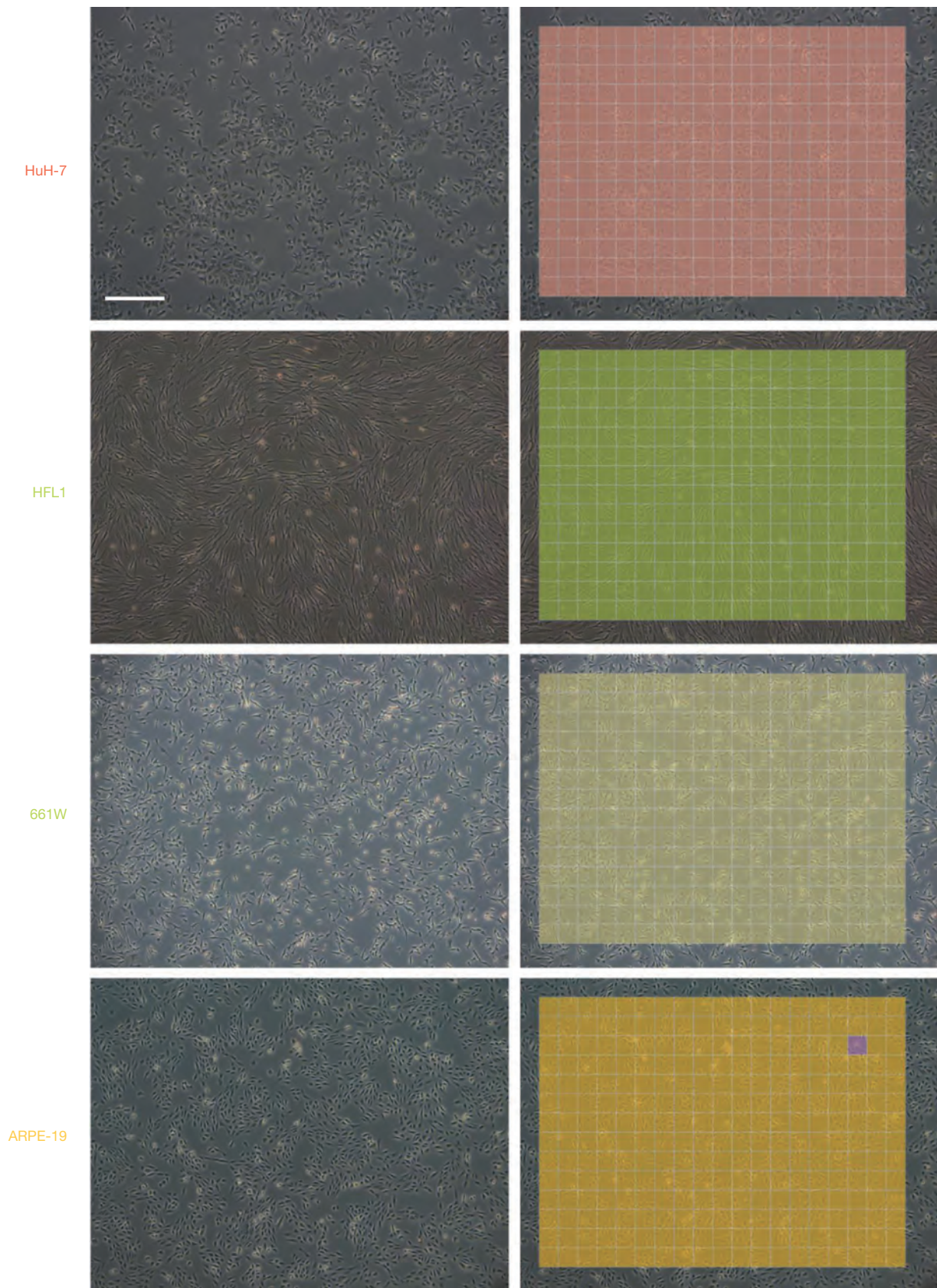
Figure S4 Using Alexnet for pure cell images classification. AlexNet was used to identify seven pure cells. The final result of one image was consistent with each patch result with a distinct color tag. Scale bar, 100 μm.
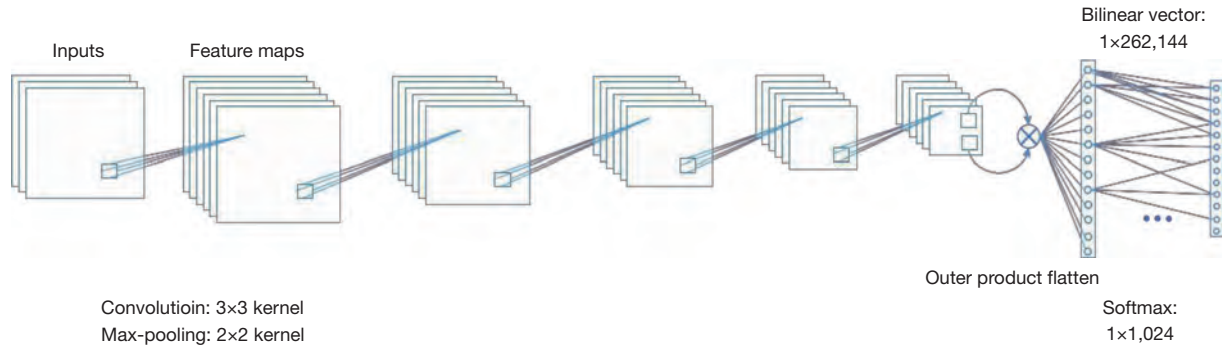
Figure S5 The architecture of a bilinear convolutional neural network (BCNN). A cell image is passed through one CNN, and its outputs multiply with its transpose using the outer product at each location of the image and are pooled to obtain the bilinear vector. This is passed through a classification layer to obtain predictions.
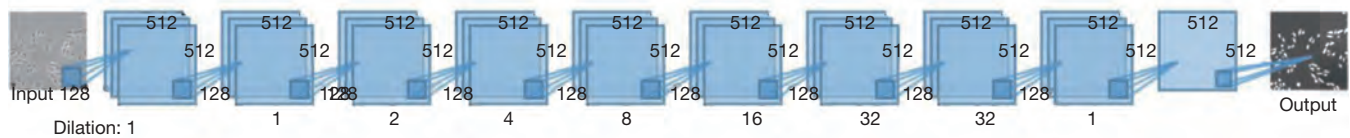
Figure S6 DilatedNet architecture diagram We constructed a 10-layer neural network based on dilation convolution to identify the cell profile. Each layer contained 128 filters with a size of 5×5. The dilation rate of each layer is indicated in the bottom. DilatedNet receives the 512×512-pixel input image block and then produces a 512×512-pixel cell probability heat map.
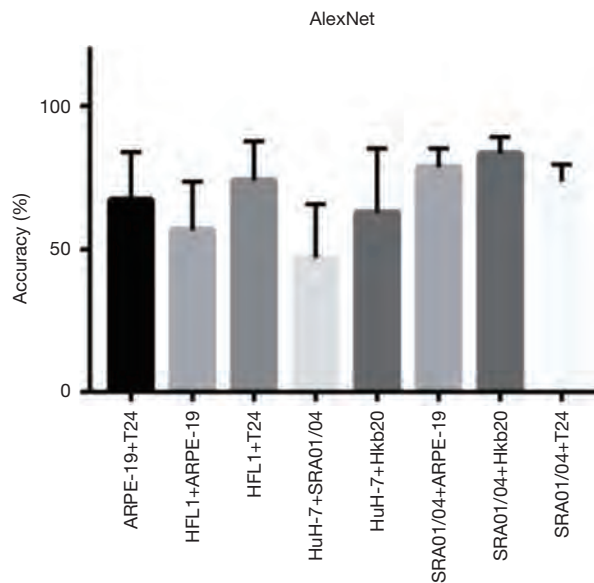
Figure S7 Using Alexnet for mixed-cell images classification. AlexNet achieved an average accuracy of 67.66% in mixed-cell images.
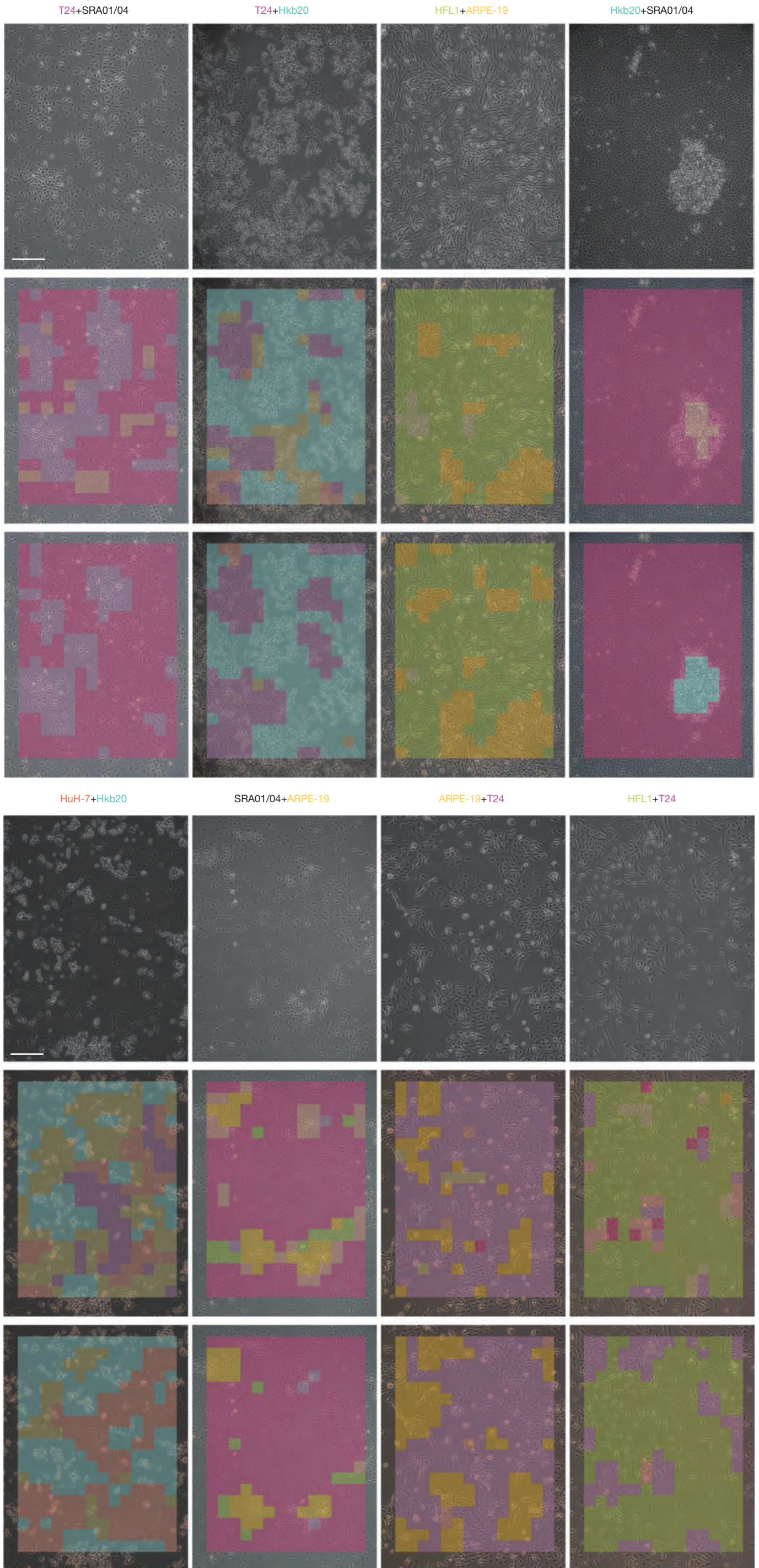
Figure S8 Bilinear convolutional neural network (BCNN) performed better than AlexNet for identification of a mixed-cell dataset. Compared to AlexNet, BCNN could detect a greater portion of the correct color tag and performed better in fine-grained recognition of small contaminated areas. The text colors represent the true types of mixed cells. Scale bar, 100 μm.
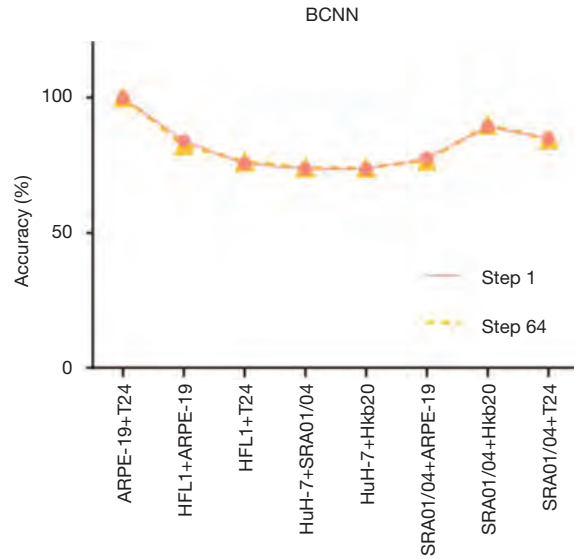
Figure S9 Comparing the accuracy of step1 and step 64 for a bilinear CNN One image was randomly selected in each category of mix-cell images for the accuracy test with steps of 1 pixel and 64 pixels, respectively. The result showed that the average accuracy did not change significantly. BCNN, bilinear convolutional neural network.
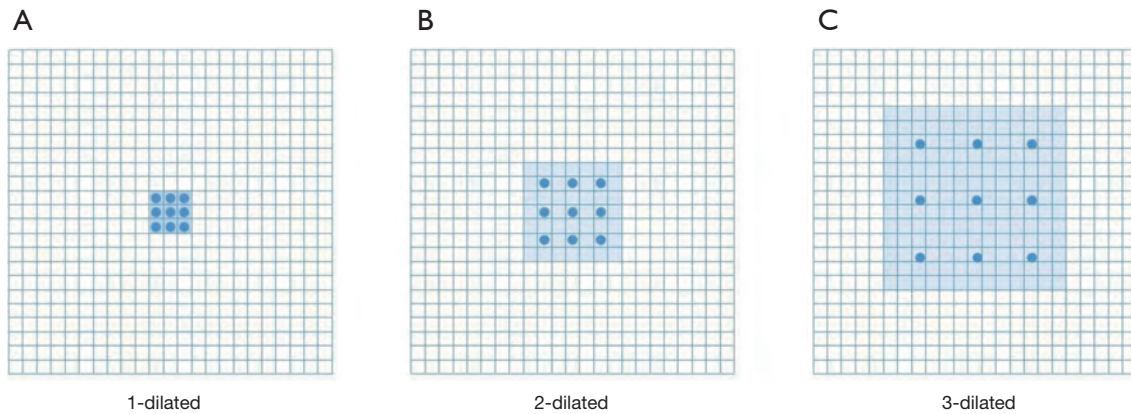

Figure S10 Systematic dilation supports exponential expansion of the receptive field without loss of resolution or coverage. (A) F1 is produced from F0 by a 1-dilated convolution; each element in F1 has a receptive field of 3×3. (B) F2 is produced from F1 by a 2-dilated convolution; each element in F2 has a receptive field of 7×7. (C) F3 is produced from F2 by a 4-dilated convolution; each element in F3 has a receptive field of 15×15. The number of parameters associated with each layer is identical. The receptive field grows exponentially, whereas the number of parameters grows linearly.
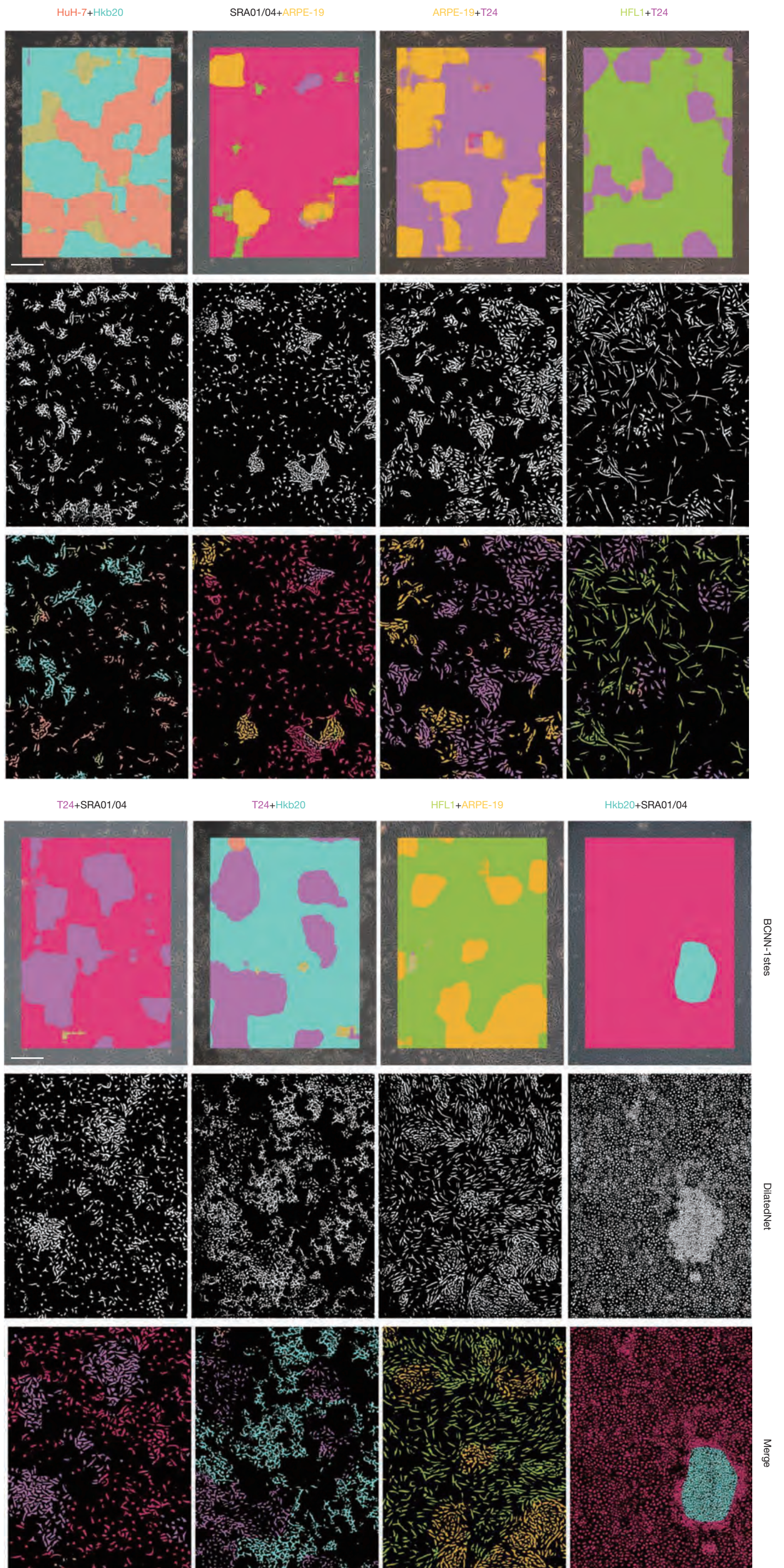
Figure S11 Combining bilinear convolutional neural network (BCNN) and DilatedNet for mixed-cell image identification. We combined BCNN and DilatedNet for both classification and location determination to address mixed-cell image identification. First, we used BCNN (sliding window with a step size of 1) to classify mixed-cell images, and DilatedNet was applied to locate the edge of cells. Then, both results were merged as showed in this figure. The text colors represent the true types of mixed cells. Scale bar, 100 μm.