World Scientific
www.worldscientific.com

Deep learning method for cell count from transmitted-light microscope

Mengyang Lu[*,**], Wei Shi[†,**], Zhengfen Jiang[‡], Boyi Li[*],
Dean Ta[*,§] and Xin Liu[*,¶,‖]

*Fudan University, Academy for Engineering and Technology
Shanghai, P. R. China*

†*Tianjin Center for Medical Device Evaluation and Inspection
Tianjin, P. R. China*

‡*Shanghai University
School of Communication & Information Engineering
Shanghai, P. R. China*

§*Fudan University
Center for Biomedical Engineering
Shanghai, P. R. China*

¶*Fudan University
State Key Laboratory of Medical Neurobiology
Institutes of Brain Science, Shanghai, P. R. China*
‖*xinliu.c@gmail.com*

Automatic cell counting provides an effective tool for medical research and diagnosis. Currently, cell counting can be completed by transmitted-light microscope, however, it requires expert knowledge and the counting accuracy which is unsatisfied for overlapped cells. Further, the image-translation-based detection method has been proposed and the potential has been shown to accomplish cell counting from transmitted-light microscope, automatically and effectively. In this work, a new deep-learning (DL)-based two-stage detection method (cGAN-YOLO) is designed to further enhance the performance of cell counting, which is achieved by combining a DL-based fluorescent image translation model and a DL-based cell detection model. The various results show that cGAN-YOLO can effectively detect and count some different types of cells from the acquired transmitted-light microscope images. Compared with the previously reported YOLO-based one-stage detection method, high recognition accuracy (RA) is achieved by the cGAN-YOLO method, with an improvement of 29.80%. Furthermore, we can also observe that

‖Corresponding author.
**Mengyang Lu and Wei Shi contributed equally to this work.

cGAN-YOLO obtains an improvement of 12.11% in RA compared with the previously reported image-translation-based detection method. In a word, cGAN-YOLO makes it possible to implement cell counting directly from the experimental acquired transmitted-light microscopy images with high flexibility and performance, which extends the applicability in clinical research.

Keywords: Automatic cell counting; transmitted-light microscope; deep-learning; fluorescent image translation.

1. Introduction

Cell counting is important in bio-medical research and clinical medical diagnosis. For example, an accurate complete blood count (CBC) is beneficial to diagnose diseases.[1,2] Currently, the cell counting methods mainly include manual counting methods, automatic counting methods based on image processing techniques, and deep-learning (DL) strategies. Among the above methods, the manual counting method is generally used due to its low-cost.[3] However, it is time-consuming and is greatly affected by user experiences.[2] The automatic counting method based on image processing techniques, such as Hough Transform,[4–6] Watershed Method,[7–9] etc., has been used to accomplish cell counting. However, the detection accuracy of these methods is limited when facing complex experimental conditions, especially facing high overlap cells.[2,10] It is noteworthy that the DL methods have been successfully used in object detection fields, e.g., face detection[11] and pedestrian detection,[12] it also shows tremendous superiority.[11–15] Considering this, some researchers have introduced DL-based methods to cell detection and counting.[16–23] As demonstrated in our previous work,[10] a modified YOLO method has been proposed for accurate blood cell count, which provides an efficient solution for indemnifying overlapping cells.

At present, cell counting is mainly implemented by fluorescent microscope images. In this way, cell counting can be completed with high accuracy, because cells in fluorescent images are generally not adhered to each other after fluorescent staining.[24] However, the acquisition of fluorescent images is time-consuming and expensive.[25] In addition, the processes of fluorescent staining may damage the cells.[26] Comparably, cell counting based on transmitted-light microscope images can be implemented more conveniently. However, due to the lack of contrast, the counting accuracy of transmitted-light images is unsatisfactory.[24] By combining a conventional image processing method (i.e., blob detection method) and a convolution neural network (CNN), in 2021, Zhang *et al.*[24] increase the detection accuracy of transmitted-light images.

To further increase the performance of cell counting in transmitted-light images, in this paper, a DL-based two-stage method termed as cGAN-YOLO is proposed. cGAN-YOLO is accomplished by incorporating a DL-based fluorescent image translation model and a DL-based cell detection model. Briefly, in stage 1, the fluorescent image translation model of conditional generative adversarial network (cGAN)[27,28] is used to predict virtual fluorescent images. In stage 2, the cell detection model of YOLO network[13] is utilized to detect cells from the generated virtual fluorescent images. A series of experimental data are analyzed to assess the performance of the cGAN-YOLO method. The various results show that cGAN-YOLO can achieve good performance for various cell types. The recognition accuracy (RA) of human umbilical vein endothelial cells (HUVEC) is 96.67%, which is improved by 10.42% compared with the YOLO-based one-stage detection method. The RA of Madin–Darby canine kidney cells (MDCK) under high-density conditions is 92.58%, which is improved by 29.80% compared with the YOLO-based one-stage detection method. Furthermore, compared with the cGAN-Watershed method, the RA of HUVECs and MDCKs obtained by the cGAN-YOLO method is improved by 18.75% and 12.11%, respectively.

This paper is arranged as follows. The structures of methods and datasets are described in Sec. 2. The various results are displayed and analyzed in Sec. 3. In Sec. 4, we discuss the related results and make some conclusions.

2. Materials and Methods

2.1. *cGAN-YOLO framework*

The proposed two-stage DL-based method (cGAN-YOLO) contains two sub-models, i.e., the

Fig. 1. The framework of the cGAN-YOLO method. cGAN-YOLO is achieved by incorporating a fluorescent image translation model based on cGAN and a cell detection model based on YOLOv3. In stage 1, the transmitted-light images are input into cGAN model to generate corresponding fluorescent images. In stage 2, YOLOv3 model can realize cell counting through these generated fluorescent images.

fluorescent image translation model based on cGAN and the cell detection model based on YOLOv3, respectively. Specifically, the fluorescent image translation model is able to predict the virtual fluorescent images from the experimentally acquired transmitted-light images. Sequentially, the cell detection model takes these generated fluorescent images as input for cell counting. Figure 1 shows the framework diagram of cGAN-YOLO.

2.1.1. *cGAN network structure*

In this work, the fluorescent image translation model is implemented by the cGAN network structure,[29] which is capable of producing high-quality images through the zero-sum game between the generator and the discriminator. The structure of the cGAN network is shown in Fig. 2. During the training phase, the generator aims to spawn images that look like real images, and the discriminator tries to classify the real images and the generated images. In this way, cGAN model can effectively learn the mappings between input and the ground truth to implement the fluorescent image translation of transmitted-light images. It is noteworthy that the proposed cGAN model can be implemented with paired or unpaired training datasets in order to

Fig. 2. The framework of cGAN model. (a) The training phase of cGAN. The generator G_Net translates the transmitted-light images into the virtual fluorescent images, and the discriminator D_Net distinguishes whether the input is ground truth (see Target in Fig. 2(a)) or the generated fluorescent image. (b) The testing process of cGAN. The trained G_Net can directly generate the virtual fluorescent images from the experimentally acquired transmitted-light images. The detailed structures of G_Net and D_Net are shown on the right. It is noted that this figure demonstrates the paired cGAN model, for an unpaired model, two generates and two discriminators with the same network structure are needed in the training phase.

flexibly achieve the fluorescent image translation. This alternative option can satisfy the needs of experiments in different scenarios. Briefly, to obtain high accuracy, the method in a paired way could be a good option. For convenient and flexible experiments, the method in an unpaired way could be helpful. Once transmitted-light images are input into the trained generator for the testing process, the corresponding fluorescent images could be obtained efficiently (see Fig. 2(b)).

In this model, an encoder–decoder framelet is chosen for the generator, which is conducive to extracting abstract features. Concretely, the whole structure includes an input layer, seven down-sampling layers, seven up-sampling layers, and an output layer. The input layer for extracting shallow features of the input contains a convolution, an instance normalization, and a rectified linear unit (ReLU). Subsequently, convolution with kernel size of 4 and stride size of 2, instance normalization, and ReLU are stacked sequentially in each down-sampling layer to extract useful features and reduce feature resolution. Then, these abstract features can be decoded to the original image size by the up-sampling layers including a deconvolution layer with stride size of 2, instance normalization, and ReLU. In addition, the down-sampling layer and the up-sampling layer of the same size are connected by the skip connection, which realizes the feature sharing and avoids the loss of important information in the down-sampling process.[30] Finally, the output layer containing a convolution layer can map the features to a virtual fluorescent image.

The discriminator based on the PatchGAN structure[27] is utilized. PatchGAN assumes that a pixel and its adjacent pixels at a certain distance are independent. When determining whether the pixel is real or fake, just pay attention to the patch of the image. In this way, it can enrich the details of the features and reduce the computational complexity of the discriminator. Specifically, inspired by Ref. 27, the patch size of 70 is adopted in the discriminator, and the discriminator consists of four down-sampling layers. Each layer includes a convolutional layer, an instance normalization, and a ReLU.

2.1.2. *YOLOv3 network structure*

Recently, many variations of YOLO have been proposed and shown great potential in object detection tasks.[31–33] Considering these, in this work, the cell detection model is achieved by the YOLOv3.[32] Figure 3 describes the brief network structure of YOLOv3 used in this work.

Briefly, the YOLOv3 network consists of a Darknet-53 feature extraction module and a multi-scale output module.[32] Here, the Darknet-53 feature extraction network contains several 3×3 convolution layers, 1×1 convolution layers, and residual blocks which help the network to extract features and avoid the gradient disappearance problem with the increasing depth of network layers.[34] The multi-scale output network learns the features at different scales. The detection, up-sampling, and feature fusion are implemented at three scales. In addition, feature fusion can benefit the network by learning
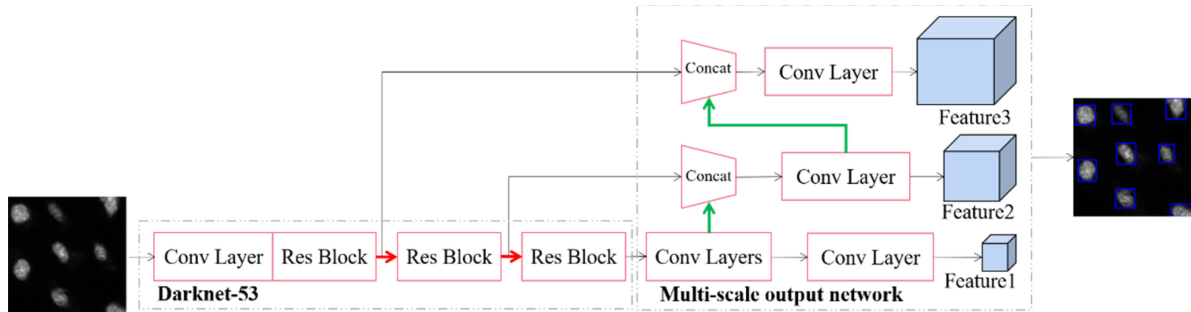

Fig. 3. YOLOv3 network structure diagram. The arrows with red color represent the down-sampling operations, and the arrows with green color mean the up-sampling operations. Briefly, the YOLOv3 network consists of a Darknet-53 module and a multi-scale output network. Conv Layer represents the basic convolution unit, which consists of the convolution, batch normalization, and leaky ReLU. Res Block contains several residual units to avoid the gradient disappearance problem due to the increasing depth of network layers. Concat represents feature fusion, which helps the network learn more features. In addition, Features 1–3, respectively, represent the extracted feature of YOLOv3 at different scales.

more features, and improving the detection accuracy of small targets.

2.2. *Loss function*

The loss function is significant to the network training convergence. Considering that the brightness and color can be unchanged by using l_1 loss, and the MS-SSIM loss integrates human subjective perception,[35,36] here, l_1 and MS-SSIM are added into the cGAN model. The formula of loss function for cGAN can be described as follows:

$$L_{\text{stage1}} = \arg \min_G \max_D \ell_{c\text{GAN}}(G, D) + \lambda_1 \|y - G(x)\|_1 \\ + \lambda_2 [1 - \text{MS\_SSIM}(y, G(x))], \quad (1)$$

where G denotes the generator used to transform the experimentally acquired transmitted-light images to the corresponding fluorescent images. y and x are the ground truth and the acquired transmitted-light image, respectively. λ_1 and λ_2 denote the weight coefficients for different terms. According to Ref. 36, here, $\lambda_1 = 16$ and $\lambda_2 = 84$. $\ell_{c\text{GAN}}$ denotes the original adversarial loss function used in cGAN.[27] It should be noted that for unsupervised training, the last two terms should be replaced by cycle consistency[37] loss incorporated with MS_SSIM loss, as follows:

$$L_{\text{stage1}} = \arg \min_G \max_D \ell_{c\text{GAN}}(G, D) + \ell_{cyc\_\text{ssim}}(G). \quad (2)$$

In this work, the second stage network is used to detect and count the cell from the transformed fluorescent images obtained by cGAN network. According to Ref. 32, the formula of the loss function can be described as follows:

$$L_{\text{stage2}} = \ell_{\text{box}}(S_{\text{out}}, S_{\text{tar}}) + \ell_{\text{obj}}(C_{\text{out}}, C_{\text{tar}}), \quad (3)$$

where ℓ_{box} denotes the center point error, width, and height error of the bounding box between the output and label. S_{out} and S_{tar} are the bounding box of output and label, respectively. ℓ_{obj} denotes the confidence loss of the bounding box containing the cell. C_{out} and C_{tar} are the probabilities of the predicted bounding box containing the cell and the label, respectively.

2.3. *Comparison methods*

To assess the performance of the cGAN-YOLO method, two cell counting methods are used as comparison. One method is the previously reported YOLO-based one-stage detection method[10,13] and the other method is the previously reported image-translation-based detection method[24] (termed as cGAN-Watershed in this work).

It is noteworthy that when using the first comparison method, YOLO is directly used to analyze the transmitted-light microscopy image. That is, the cell counting task is accomplished by using YOLO to learn the relevance between the transmitted-light images and labels. When using cGAN-Watershed, the transmitted-light image is first transformed into fluorescent images. Here, for a fair comparison, similar to the cGAN-YOLO method, cGAN[27] is used to implement the virtual fluorescent image translation. Similarly, cGAN model can be trained by using the paired or unpaired training data. After that, based on the generated fluorescent images, the cell counting task is completed by using watershed[7] and region detection methods. It should also be pointed out that for improving the detection accuracy based on the watershed method, a gaussian filter is used to highlight the position of the nucleus after translation. The sharpening method is used to further enhance the contrast of the generated fluorescent image.

To quantitatively evaluate the transformed performance of stage 1, here, peak signal-to-noise ratio (PSNR) and learned perceptual image patch similarity[38] (LPIPS) are utilized. Referring to Ref. 38, LPIPS can effectively measure the differences between images by calculating L2 distance of deep features in images, which is consistent with human perception.

In addition, to quantitatively evaluate the final count performance, the precision rate (PR), recall rate (RC), mean average precision (mAP), and RA[39] indicators are also calculated, respectively, as follows:

$$\text{PR} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \quad (4)$$

$$\text{RC} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \quad (5)$$

$$\text{mAP} = \int_0^1 \text{PR}(\text{RC}) d\text{RC}, \quad (6)$$

$$\text{RA} = 1 - \frac{|N_{\text{GT}} - N_{\text{out}}|}{N_{\text{GT}}}, \quad (7)$$
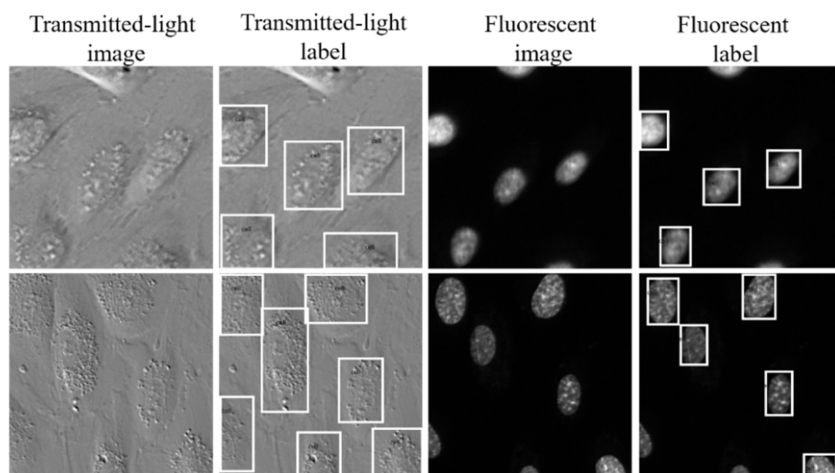
where TP, FP, and FN represent true positive, false positive, and false negative respectively. mAP

denotes mean average precision. N_{GT} denotes the cell number in the ground truth, and N_{out} denotes the cell number after detection.
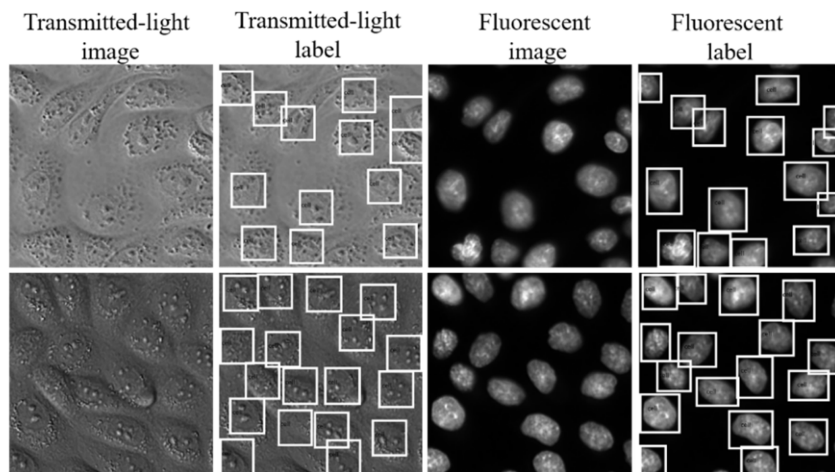
2.4. Cell dataset

In this work, three sets of microscopy images[40] including the transmitted-light images and the fluorescent images are used. In detail, the HUVEC under low-density conditions, the MDCK under high-density conditions, and the mouse keratinocytes are used to assess the detection ability and generalization ability of the cGAN-YOLO method. Here, the size of these images is 512×512 pixels.

In detail, the training set contains 4000 paired images for HUVEC and MDCK, which are selected from the above datasets. Here, all labeled data are manually annotated under the guidance of professionals, which is realized by using LabelImg software (https://github.com/tzutalin/labelImg). It is noteworthy that in fluorescent images, the cell nuclei have clear boundaries and can be easily marked. However, the whole cell needs to be marked in transmitted-light images when labeling due to the low contrast. Figure 4 shows some labeled images used in this work. Concretely, the transmitted-light images and corresponding transmitted-light labels are used to train the YOLO-based detection

(a) HUVEC

(b) MDCK

Fig. 4. Some cell images and labels used in this work. (a) The images of HUVEC. (b) The MDCK images. The first column is the transmitted-light images, the second column is the labels of the transmitted-light images, the third is the fluorescent images, and the fourth is the labels of the fluorescent images.

Table 1. The data used in training different detection methods.

Methods	Input	Target
Stage 1 of cGAN-YOLO	Transmitted-light images	Fluorescent images
Stage 2 of cGAN-YOLO	Fluorescent images	Fluorescent labels
YOLO	Transmitted-light images	Transmitted-light labels
cGAN-Watershed	Transmitted-light images	Fluorescent images

method. The transmitted-light images and corresponding fluorescent images are adopted for training in stage 1 of the proposed cGAN-YOLO method, and the fluorescent images and corresponding fluorescent labels are adopted for training in stage 2 of the proposed cGAN-YOLO method. Also, the training data of cGAN-Watershed is the same as the dataset used in stage 1 of the cGAN-YOLO method. Table 1 reveals the details of the training dataset.

2.5. *Training*

In this work, cGAN model and YOLOv3 model are trained, respectively. In detail, the model of fluorescent image translation based on cGAN is trained for 200 epochs. During the training processes, Adam optimization is utilized. The batch size is 4. In order to make the training converge effectively and stably, a learning rate attenuation strategy is used. Concretely, for the first 100 epochs, the learning rate is 0.0002, for the last 100 epochs, the learning rate decreases linearly to 0. In addition, it is noteworthy that cGAN can be trained with paired or unpaired data. The use of the paired data can achieve high accuracy translation, and the use of the unpaired data can flexibly implement translation. Based on the above considerations, to better clarify the effectiveness of the method, here, cGAN is trained with paired data and unpaired data, respectively.

The cell detection model based on YOLOv3 is trained for 200 epochs, and the stochastic gradient descent optimization is utilized. The learning rate is set to 0.001. Here, the training process is performed with an NVidia Tesla V100 GPU (16 GB RAM), 2 Intel Xeon Gold 6130 (2.1 GHZ), and 192 G DDR4 REG ECC.

3. Results and Discussion

3.1. *Detection capability under low cell density conditions*

Figure 5 shows the detection and comparison results of HUVECs under low-cell density obtained by different methods. Here, the first column displays the transmitted-light images as input. The second column describes the corresponding fluorescent images dyed by DAPI for labeling nuclei. Detection results obtained by the YOLO method from transmitted-light images are shown in the third column. The fourth and fifth columns display the counting results of the cGAN-Watershed method in an unpaired and paired way, respectively. The last two columns represent the results of the cGAN-YOLO method in an unpaired and paired way, respectively.

These results demonstrate that the YOLO counting method based on the transmitted-light images cannot accurately detect the cell with different shapes, as shown in the third column of Fig. 5. Although the detection performance of the cGAN-Watershed method is improved, the problem of missed detection still exists, see the fourth and fifth columns in Fig. 5. Compared with the above method, an obvious improvement can be observed in the detection results acquired by the proposed method cGAN-YOLO, which means cGAN-YOLO method provides an efficient tool to implement accurate cell counting.

Furthermore, in order to evaluate the quality of the generated images, 200 images are randomly chosen from test datasets to calculate the quantitative indexes. Table 2 shows the corresponding quantitative results of the virtual fluorescent images obtained from stage 1 in the proposed method. From these results, it can be demonstrated that the synthesis of fluorescence images is effective, especially for paired training strategy. Also, 20 images are randomly selected to assess the counting performance. The comparison of RA, PR, RC, and mAP obtained by YOLO, cGAN-Watershed method trained with unpaired data, cGAN-Watershed method trained with paired data, cGAN-YOLO method trained with unpaired data, and cGAN-YOLO method trained with paired data are shown in Table 3. The results demonstrate that as opposed to the YOLO counting method, the accuracy of the cGAN-YOLO method trained with paired data is improved by 10.42%. Further,
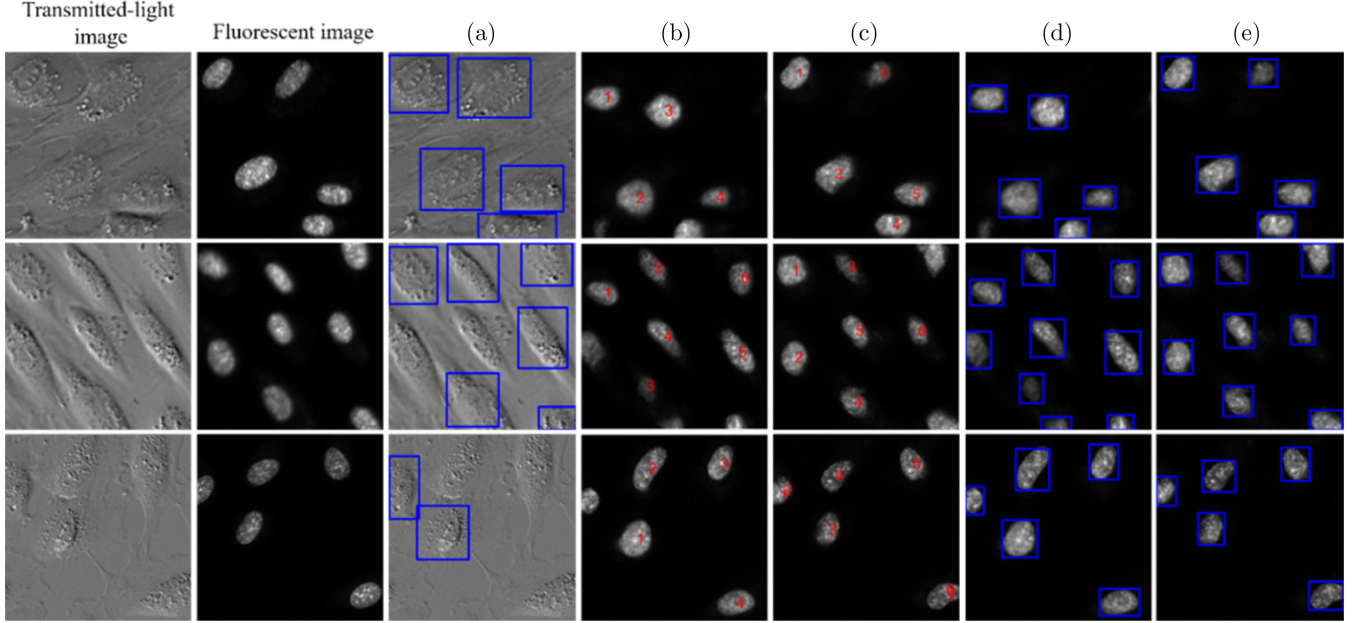
Fig. 5. The detection results of different methods from HUVECs. Each row represents the different HUVEC samples. The first and the second columns show the transmitted-light images and corresponding fluorescent images, respectively. (a) The detection results are obtained by YOLO from the experimental acquired transmitted-light images. (b) The counting results of cGAN-Watershed are trained with the unpaired data. (c) The counting results of cGAN-Watershed are trained with the paired data. (d) The counting results of the cGAN-YOLO method are trained with the unpaired data. (e) The counting results of cGAN-YOLO are trained with the paired data.

Table 2. The LPIPS and PSNR of HUVECs in stage 1 calculated from 200 images.

	LPIPS	PSNR (dB)
Paired cGAN	0.10 ± 0.01	25.44 ± 0.17
Unpaired cGAN	0.17 ± 0.03	20.52 ± 0.45

compared with the paired cGAN-Watershed method, the accuracy of the paired cGAN-YOLO is improved by 18.75%. Similar results can also be observed in other indexes, e.g., mAP. Here, it can be seen that the methods based on the unpaired data have relatively low results, and this may be caused by the poor virtual fluorescent images generated from stage 1 (see the first row (b) and (d) of Fig. 5).

Table 3. Comparison of quantitative index in HUVECs by different methods.

Methods	RA (%)	PR	RC	mAP
YOLO	86.25	0.98	0.88	0.85
Unpaired cGAN-Watershed	68.67	0.76	0.76	0.64
Paired cGAN-Watershed	77.92	0.96	0.79	0.80
Unpaired cGAN-YOLO	73.97	0.75	0.93	0.81
Paired cGAN-YOLO	96.67	0.95	0.98	0.97

3.2. Detection capability under high cell density conditions

Figure 6 shows the detection results of MDCKs under high-cell density obtained by YOLO method, cGAN-Watershed method, and cGAN-YOLO method. Here, the first two columns show the transmitted-light images and corresponding fluorescent images, respectively. Detection results obtained by the YOLO method from transmitted-light images are displayed in the third column. The fourth and fifth columns show the counting results obtained by the cGAN-Watershed in an unpaired and paired way, respectively. The last two columns show the results of the cGAN-YOLO method in an unpaired and paired way, respectively. The results demonstrate that with the increase in cell density, the missed detection of the YOLO counting method is aggravated. Comparatively, the cGAN-YOLO method can achieve accurate detection even in the case of cell adhesion and overlap, as shown in Figs. 6(d) and 6(e).

Table 4 shows the quantitative indexes of the MDCK virtual fluorescent images obtained by stage 1 of the proposed method. Table 5 compares the quantitative evaluate results of the YOLO method,
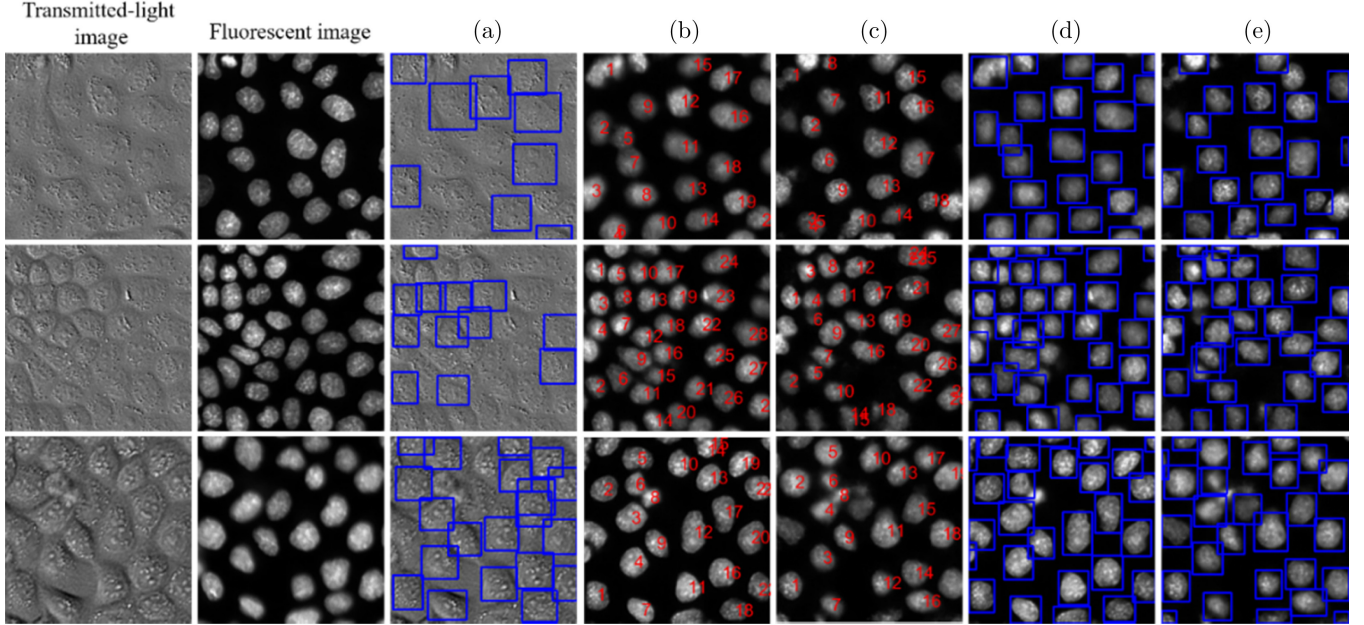
Fig. 6. The detection results of MDCK cells under high density conditions. Each row represents the different MDCK samples. The first and the second columns show the experimental acquired transmitted-light images and corresponding fluorescent images, respectively. (a) The detection results obtained by YOLO method from the acquired transmitted-light images. (b) The counting results of cGAN-Watershed trained with unpaired data. (c) The counting results of cGAN-Watershed trained with paired data. (d) The counting results of cGAN-YOLO trained with unpaired data. (e) The counting results of cGAN-YOLO trained with paired data.

Table 4. The LPIPS and PSNR of HUVECs in stage 1 calculated from 200 images.

	LPIPS	PSNR (dB)
Paired cGAN	0.18 ± 0.01	19.91 ± 0.78
Unpaired cGAN	0.20 ± 0.04	17.63 ± 2.00

cGAN-Watershed method, and cGAN-YOLO method, which are calculated from 20 random MDCK images. The quantitative results demonstrate that the accuracy of the YOLO counting method based on the transmitted-light image is low, which is 62.78%. The accuracy of the cGAN-Watershed method is also unsatisfactory, with an RA of 79.06% in an unpaired way and 80.47% in a

paired way. Comparatively, the cGAN-YOLO method has a better performance. Specifically, the accuracy of the cGAN-YOLO method in an unpaired way is improved by 25.18%, and the cGAN-YOLO method in a paired way is enhanced by 29.80% compared to the YOLO counting method. Considering the other indexes e.g., mAP, the cGAN-YOLO method with paired data shows better performance in high cell density conditions.

The quantitative results in Tables 3 and 5 demonstrate that the proposed cGAN-YOLO counting method is not only applicable to images with low-cell density and simple cell morphology, but also has a good performance for images with high-cell density and various cell morphologies. As a result, this method demonstrates good potential for clinical application.

3.3. *Generalization capability*

To assess the generalization capacity of the cGAN-YOLO method, mouse keratinocytes cells are used for testing. It should be noted that the transmitted-light images of mouse keratinocytes are obtained by phase contrast microscope, which is different from the training dataset transmitted-light images of

Table 5. Comparison of quantitative index in MDCKs by different methods.

Methods	RA (%)	PR	RC	mAP
YOLO	62.78	0.91	0.58	0.65
Unpaired cGAN-Watershed	79.06	0.95	0.76	0.81
Paired cGAN-Watershed	80.47	0.96	0.78	0.84
Unpaired cGAN-YOLO	87.96	0.87	0.92	0.89
Paired cGAN-YOLO	92.58	0.93	0.95	0.94

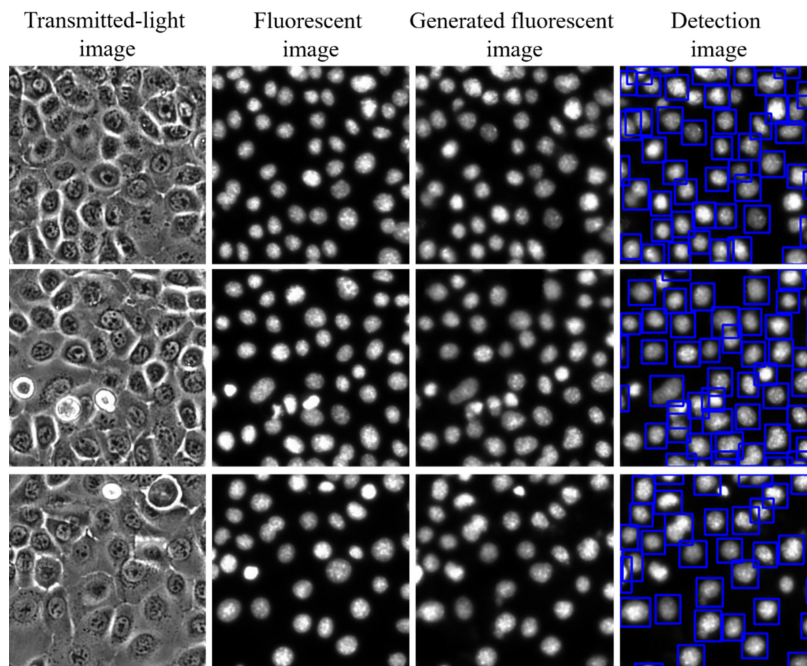| Transmitted-light image	Fluorescent image	Generated fluorescent image	Detection image

Fig. 7. The detection results obtained by cGAN-YOLO with the mouse keratinocytes. The first column is the transmitted-light image of different samples, the second column is the real fluorescent image, the third and the fourth columns are the generated fluorescent images and detection results acquired by the proposed cGAN-YOLO model in a paired way, respectively.

MDCK cells obtained by differential interference contrast (DIC) microscope. Figure 7 shows the count results of mouse keratinocytes cells. Here, the first column is the transmitted-light images of different samples. The second column shows the real fluorescent images. The third and the fourth columns are the generated fluorescent images and detection results of the cGAN-YOLO model in a paired way, respectively. Also, 30 images are randomly chosen to quantitatively assess the effectiveness of the cGAN-YOLO method, and the RA is 92.33%. These results show that even if the cGAN-YOLO model is applied to different data sets, good RA can be obtained.

3.4. Comparisons of counting scheme of directly from real fluorescent images

Finally, we also analyze and compare the performance of the proposed two-stage cGAN-YOLO method with a conventional cell counting scheme, i.e., applying the detection method to the real fluorescent images. Figure 8 shows the corresponding comparison results from the HUVEC and MDCK cells, where the first row shows the detection results obtained by YOLO method and the second row shows the detection results obtained by the cGAN-YOLO. Here, it is worth noting that in
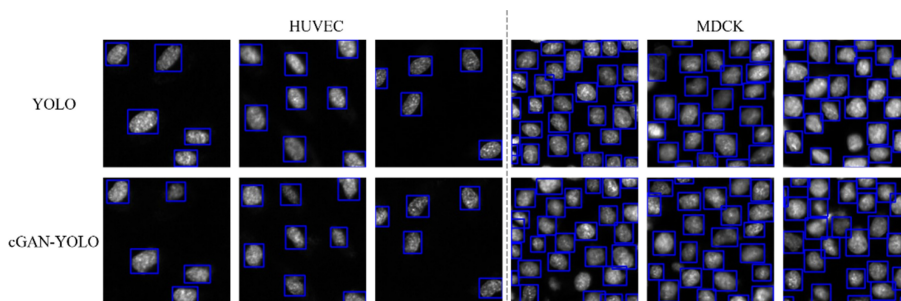

Fig. 8. The comparison results of HUVEC and MDCK cells obtained by YOLO and cGAN-YOLO. The first row shows the detection results obtained by applying YOLO method to the real fluorescence images. The second row shows the detection results obtained by applying the proposed two-stage cGAN-YOLO method to the transmitted-light images.

this case, YOLO is directly applied to the real fluorescent images. Comparably, the proposed two-stage cGAN-YOLO method is applied to the transmitted-light images. Based on the above results, we can find that when applied to the real fluorescent images, YOLO method can effectively implement cell counting with a high accuracy. The proposed cGAN-YOLO method also demonstrates a similar counting performance, even if it is applied to the transmitted-light images. It is further confirmed by the quantitative results. For example, when counting the HUVEC cells in Fig. 8, the RA, PR, and RC values from both methods are the same, i.e., 100%, 1, and 1, respectively. For the MDCK cells in Fig. 8, the RA, PR, and RC values from YOLO are 95.99%, 0.97, and 0.97. Comparably, the RA, PR, and RC values from cGAN-YOLO are 89.76%, 0.93, and 0.94, respectively.

Nevertheless, there are still some limitations when using the proposed two-stage cGAN-YOLO method in cell counting. First, there is a common obstacle in the DL-based methods where physical mechanism in imaging processes, e.g., light propagation in tissues, is not revealed totally. To address this problem, the integrating physical relationships[41] or mathematical methods[42,43] with current DL-based method could be helpful to break through the limitations of the physical principles. In addition, it should be pointed out that the performance of DL-based method may be greatly affected by the amount of training dataset and network structure. Also, in this work, only several types of cells are used for experiments. Clinically, the cell types involved are more diverse and complex, so it is necessary to collect and evaluate different types of cell data. Moreover, there is a problem of adherent cell misdetection, which could be improved by introducing a pixel-level detection method or segmentation method.[44] In the future, these related works will be further investigated.

4. Conclusions

In this paper, we design a two-stage cell counting method based on DL (cGAN-YOLO) to enhance the performance of cell detection in transmitted-light images. The experimental results from various cell types demonstrate that compared to the cGAN-Watershed method and the one-stage YOLO method, cGAN-YOLO can efficiently realize cell counting at various densities and morphologies.

Correspondingly, cGAN-YOLO can be used as a tool to implement cell counting directly from transmitted-light images with high flexibility and good accuracy, which is beneficial in assisting researchers in cell experiments and studying cell growth.

Acknowledgments

Conflicts of Interest

The authors declare that there are no conflicts of interest relevant to this article.

References

1. M. Madjid, O. Fatemi, "Components of the complete blood count as risk predictors for coronary heart disease: In-depth review and update," *Tex. Heart Inst. J.* **40**(1), 17–29 (2013).

2. S. Acharjee, S. Chakrabartty, M. L. Alam, N. Dey, V. Santhi, A. S. Ashour, A semiautomated approach using GUI for the detection of red blood cells, *Proc. Int. Conf. Electr. Electron. Optim. Techn., ICEEOT*, pp. 525–529, IEEE (2016).

3. G. Johnston, "Automated handheld instrument improves counting precision across multiple cell lines," *BioTechniques* **48**(4), 325–327 (2010).

4. Z. Ejaz, C. A. Ahmad, H. Justin, "Automatic red blood cell detection and counting system using Hough transform," *Am. J. Pharm. Sci.* **5**(8), 7913–7920 (2018).

5. V. Acharya, P. Kumar, "Identification and red blood cell automated counting from blood smear images using computer-aided system," *Med. Biol. Eng. Comput.* **56**(3), 483–489 (2018).

6. S. M. Mazalan, N. H. Mahmood, M. A. A. Razak, Automated red blood cells counting in peripheral blood smear image using circular hough transform, *Int. Conf. Artif. Intell. Modelling Simul. (AIMS)*, pp. 320–324, IEEE (2013).

7. S. Biswas, D. Ghoshal, "Blood cell detection using thresholding estimation based watershed transformation with sobel filter in frequency domain," *Procedia Comput. Sci.* **89**, 651–657 (2016).

8. H. Liu, H. Cao, E. Song, "Bone marrow cells detection: A technique for the microscopic image analysis," *J. Med. Syst.* **43**, 82 (2019).

9. A. S. Ramos, C. H. Fontes, A. M. Ferreira, C. C. Baccili, K. N. Silva, V. Gomes, G. J. A. Melo, "Somatic cell count in buffalo milk using fuzzy clustering and image processing techniques," *J. Dairy Res.* **88**(1), 69–72 (2021).

10. Z. Jiang, Z. Yan, W. Gu, J. Jiang, "Improved detection performance in blood cell count by an attention-guided deep learning method," *OSA Continuum* **4**(2), 323–333 (2021).

11. S. Baek, M. Song, J. Jang, G. Kim, S. Paik, "Face detection in untrained deep neural networks," *Nat. Commun.* **12**, 7328 (2021).

12. X. Cao, Z. Wang, P. Yan, X. Li, "Transfer learning for pedestrian detection," *Neurocomput.* **100**, 51–57 (2013).

13. J. Redmon, S. Divvala, R. Girshick, A. Farhadi, You only look once: Unified, real-time object detection, *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp. 779–788, IEEE (2016).

14. R. Padilla, S. L. Netto, D. Silva, A survey on performance metrics for object-detection algorithms, *Int. Conf. Syst. Signals Image Process. (IWSSIP)*, pp. 237–242, IEEE (2016) .

15. V. Nguyen, T. D. Ngo, "Single-image crowd counting: A comparative survey on deep learning-based approaches," *Int. J. Multimed. Info. Retr.* **9**(2), 63–80 (2020).

16. T. Falk, D. Mai, R. Bensch, Ö. Çiçek, A. Abdulkadir, Y. Marrakchi, A. Böhm, J. Deubner, Z. Jäckel, K. Seiwald, A. Dovzhenko, O. Tietz, C. D. Bosco, S. Walsh, D. Saltukoglu, T. L. Tay, M. Prinz, K. Palme, M. Simons, I. Diester, T. Brox, O. Ronneberger, "U-Net: Deep learning for cell counting, detection, and morphometry," *Nat. Meth.* **16**(1), 67–70 (2019).

17. Z. Wang, Z. Yin, Annotation-efficient cell counting, *Proc. Int. Conf. Med. Image Comput. Comput.-Assisted Intervention (MICCA)*, pp. 405–414 (Springer, Cham, 2021).

18. T. H. Song, V. Sanchez, H. Eldaly, N. Rajpoot, Simultaneous cell detection and classification with an asymmetric deep autoencoder in bone marrow histology images, *Annu. Conf. Med. Image Underst. Anal. (MIUA)*, pp. 829–838 (Springer, Cham, 2017).

19. G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. Laak, B. V. Ginneken, C. I. Samchez, "A survey on deep learning in medical image analysis," *Med. Image. Anal.* **42**, 60–88 (2017).

20. Y. Guo, O. Krupa, J. Stein, G. Wu, A. Krishnamurthy, "SAU-net: A unified network for cell counting in 2d and 3d microscopy images," *IEEE/ACM Trans. Comput. Biol. Bioinform.* **19**(4), 1920–1932 (2021).

21. S. He, K. T. Minn, L. Solnica-Krezel, M. Anastasio, H. Li, "Deeply-supervised density regression for automatic cell counting in microscopy images," *Med. Image Anal.* **68**, 101892 (2021).

22. Y. Xie, F. Xing, X. Shi, X. Kong, H. Su, L. Yang, "Efficient and robust cell detection: A structured regression approach," *Med. Image Anal.* **44**, 245–254 (2018).

23. J. W. Choi, Y. Ku, B. W. Yoo, J. Kim, D. S. Lee, Y. J. Chai, H. Kong, H. C. Kim, "White blood cell differential count of maturation stages in bone marrow smear using dual-stage convolutional neural networks," *Plos One* **12**(12), e0189259 (2017).

24. Z. Zhang, K. Leong, K. Vliet, G. Barbastathis, A. Ravasio, "Deep learning for label-free nuclei detection from implicit phase information of mesenchymal stem cells," *Biomed. Opt. Exp.* **12**(3), 1683–1706 (2021).

25. C. Ounkomol, S. Seshamani, M. Maleckar, F. Collman, G. Johnson, "Label-free prediction of three-dimensional fluorescence images from transmitted-light microscopy," *Nat. Meth.* **15**(11), 917–925 (2018).

26. D. Kim, Y. Min, J. Oh, Y. Cho, "AI-powered transmitted light microscopy for functional analysis of live cells," *Sci. Rep.* **9**(1), 1–9 (2019).

27. P. Isola, J. Zhu, T. Zhou, A. Efros, Image-to-image translation with conditional adversarial networks, *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp. 5967–5976, IEEE (2017).

28. Z. Jiang, B. Li, T. Tran, J. Jiang, X. Liu, D. Ta, "Fluo-Fluo translation based on deep learning," *Chin. Opt. Lett.* **20**, 031701 (2022).

29. Z. Yu, Z. Ju, X. Zhang, Z. Meng, F. Yin, K. Xu, "High-speed multimode fiber imaging system based on conditional generative adversarial network," *Chin. Opt. Lett.* **19**, 081101 (2021).

30. T. Tong, G. Li, X. Liu, Q. Gao, Image super-resolution using dense skip connections, *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, pp. 4799–4807, IEEE (2017).

31. J. Redmon, A. Farhadi, YOLO9000: better, faster, stronger, *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp. 21–26, IEEE (2017).

32. J. Redmon, A. Farhadi, "YOLOv3: An incremental improvement," arXiv:1804. 02767 (2018).

33. C. J. Xu, X. F. Wang, Y. D. Yang, "Attention-YOLO: YOLO detection algorithm that introduces attention mechanism," *Comput. Eng. Appl.* **55**(6), 13–23 (2019).

34. K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, *Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp. 770–778, IEEE (2016).

35. Z. Wang, E. P. Simoncelli, A. C. Bovik, Multi-scale structural similarity for image quality assessment, *Asilom. Conf. Signals, Syst. Comput. (ACSSC)*, pp. 1398–1402, IEEE (2003).

36. H. Zhao, O. Gallo, I. Frosiom, J. Kautz, "Loss functions for image restoration with neural networks," *IEEE Trans. Comput. Imaging* **3**(1), 47–57 (2017).

37. J. Zhu, T. Park, P. Isola, A. A. Efros, Unpaired image-to-image translation using cycle-consistent adversarial networks, *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, pp. 2242–2251, IEEE (2017).

38. R. Zhang, P. Isola, A. A. Efros, E. Shechtman, O. Wang, The unreasonable effectiveness of deep features as a perceptual metric, *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp. 586–595, IEEE (2018).

39. M. M. Alam, M. T. Islam, "Machine learning approach of automatic identification and counting of blood cells," *Healthc. Technol. Lett.* **6**(4), 103–108 (2019).

40. J. LaChance, D. J. Cohen, "Practical fluorescence reconstruction microscopy for large samples and low-magnification imaging," *PLoS Comput. Biol.* **16**(12), e1008443 (2020).

41. F. Wang, Y. Bian, H. Wang, M. Lyu, G. Pedrini, W. Osten, G. Barbastathis, G. Situ, "Phase imaging with an untrained neural network," *Light Sci. Appl.* **9**, 77 (2020).

42. R. Anirudh, J. J. Thiagarajan, B. Kailkhura, T. Bremer, "An unsupervised approach to solving inverse problems using generative adversarial networks," preprint, arXiv:1805.07281 [cs.CV] (2018).

43. J. Liu, Y. Sun, X. Xu, U. S. Kamilov, Image restoration using total variation regularized deep image prior, *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, pp. 7715–7719, IEEE (2019).

44. R. Hollandi, A. Szkalisity, T. Toth, E. Tasnadi, C. Molnar, B. Mathe, I. Grexa, J. Molnar, A. Balind, M. Gorbe, M. Kovacs, E. Migh, A. Goodman, T. Balassa, K. Koos, W. Wang, J. Caicedo, N. Bara, F. Kovacs, L. Paavolainen, T. Danka, A. Kriston, A. Elizabeth, K. Smith, P. Horvath, "NucleAIzer: A parameter-free deep learning framework for nucleus segmentation using image style transfer," *Cell Syst.* **10**(5), 453–458 (2020).