Medical Image Analysis 57 (2019) 176–185

Contents lists available at ScienceDirect

Medical Image Analysis

journal homepage: www.elsevier.com/locate/media

TOP-GAN: Stain-free cancer cell classification using deep learning with a small training set

Moran Rubin[a,b], Omer Stein[b], Nir A. Turko[a], Yoav Nygate[a], Darina Roitshtain[a], Lidor Karako[a], Itay Barnea[a], Raja Giryes[b], Natan T. Shaked[a,*]

[a] Department of Biomedical Engineering, Faculty of Engineering, Tel Aviv University, Tel Aviv 69978, Israel
[b] School of Electrical Engineering, Faculty of Engineering, Tel Aviv University, Tel Aviv 69978, Israel

ARTICLE INFO

Article history:
Received 16 December 2018
Revised 18 May 2019
Accepted 25 June 2019
Available online 26 June 2019

Keywords:
Holography
Quantitative phase imaging
Deep learning
Machine learning algorithms
Image classification
Biological cells

ABSTRACT

We propose a new deep learning approach for medical imaging that copes with the problem of a small training set, the main bottleneck of deep learning, and apply it for classification of healthy and cancer cell lines acquired by quantitative phase imaging. The proposed method, called transferring of pre-trained generative adversarial network (TOP-GAN), is hybridization between transfer learning and generative adversarial networks (GANs). Healthy cells and cancer cells of different metastatic potential have been imaged by low-coherence off-axis holography. After the acquisition, the optical path delay maps of the cells are extracted and directly used as inputs to the networks. In order to cope with the small number of classified images, we use GANs to train a large number of unclassified images from another cell type (sperm cells). After this preliminary training, we change the last layers of the network and design automatic classifiers for the correct cell type (healthy/primary cancer/metastatic cancer) with 90–99% accuracies, although small training sets of down to several images are used. These results are better in comparison to other classic methods that aim at coping with the same problem of a small training set. We believe that our approach makes the combination of holographic microscopy and deep learning networks more accessible to the medical field by enabling a rapid, automatic and accurate classification in stain-free imaging flow cytometry. Furthermore, our approach is expected to be applicable to many other medical image classification tasks, suffering from a small training set.

1. Introduction

Cancer is a leading cause of death worldwide. Flow cytometry of body fluids obtained by routine medical tests is expected to identify circulating tumor cells (Paterlini-Brechot and Benali, 2007; Yu et al., 2011). However, isolation of these cancer cells is laborious and typically yields uniformly round cells, which are hard to grade (Crowley et al., 2013). In flow cytometry for cell sorting, one evaluates cellular features through fluorescence markers (Ibrahim and van den Engh, 2007). However, fluorescent markers tend to photobleach, which damages the image contrast and the prognosis decisions (Jensen, 2012). The morphology and texture of cancer cells changes during cancer progression (Wu et al., 2015; Zink et al., al.,2004). Without staining, however, biological cells are nearly transparent, resulting in a low image contrast.

An internal contrast mechanism that can be used when imaging cells without staining is their refractive index. The light beam passing through the imaged cells is delayed, since the cells have a slightly higher refractive index compared to their surroundings. Conventional intensity-based detectors are not fast enough to record this light delay directly. Phase imaging methods, on the other hand, use optical interference to record the delay of light passing through the sample, and thus they yield stain-free contrast in the image. Contrary to qualitative phase contrast methods, interferometric phase microscopy (IPM) yields the full sample wavefront, containing the optical thickness map or optical path delay (OPD) map of the cell, so that on each spatial point of this map, OPD is equal to the integral of the refractive index values across the cell thickness (Girshovitz and Shaked, 2012). In addition to contrast obtained on all cell points without staining, IPM allows calculating quantitative parameters, such as cell volume and dry mass, which have not been available to clinicians so far (Rappaz et al., 2009). In the last years, we made significant efforts to make these wavefront sensors affordable for clinical use (Shaked, 2012; Girshovitz and Shaked, 2013; Roitshtain et al., 2016; Nativ and Shaked, 2017) by attaching a portable interferometric module to the exit port of an existing clinical microscope, making

* Corresponding author.
E-mail address: nshaked@tau.ac.il (N.T. Shaked).

https://doi.org/10.1016/j.media.2019.06.014
1361-8415/

this technology accessible and affordable to the clinicians' direct use.

Combining holographic microscopy with recent developments in the field of machine learning enables automatic stain-free analysis of large numbers of cells, which can assist with the classification of different types of cells (Martinez-Torres et al., 2015; Go et al., 2018; Mirsky et al., 2017; Roitshtain et al., 2017; Hejna et al., 2017; Yoon et al., 2017; Park et al., 2016; Lam et al., 2017; Chen et al., 2016; Moon et al., 2009). In contrast to previous works presenting statistical discrimination of holographic data of cells (Javidi et al., 2005; Moon and Javidi, 2008; Yeom et al., 2006; Watanable et al., 2013; Jaferzadeh and Moon, 2016), the advantage of machine learning classifiers is that they can work on multi-feature space, or even on unknown feature space, for classification between the groups examined. Furthermore, various combinations of machine learning and holographic microscopy have been proposed lately (e.g. (Rivenson et al., 2017, 2018; Jo et al., 2019; Ronneberger et al., 2015)).

The fact that IPM can now be implemented in clinical settings and provide the clinicians with tens of new parameters extracted from the cell OPD map gives rise to the question of how to account for all parameters together in order to classify the cells. Simple machine learning approaches can weigh the various parameters, for example by principle component analysis (PCA), followed by a support vector machine (SVM) classifier (Roitshtain et al., 2017).

Instead of manually designing parameters to be extracted from the OPD maps, an alternative approach is to apply machine learning techniques directly on these maps with the goal of cell classification. This approach is more global, since it creates hidden parameterization that might be missed by the manually designed parameters, and thus, although more computationally heavy, it is expected to yield better classification results.

In the past years, the concept of deep convolutional neural networks (CNNs) has revolutionized the field of image recognition and classification (Jo et al., 2017; LeCun et al., 2015; Oquab et al., 2014). When the number of parameters in the network is large, more samples are required for training in order to avoid overfitting situations (Erhan et al., 2009). One major challenge when building CNNs is that it is frequently hard to acquire very large number of classified training examples. This problem is common to many medical classification tasks solved with deep learning, and researchers thrive to suggest new approaches that allow applying deep learning with limited number of examples (Tajbakhsh et al., 2016; Shi et al., 2016; Litjens, 2017).

A simple possible solution to the problem of a small training set is to expand the training set by using data augmentation (Ronneberger et al., 2015; Krizhevsky et al., 2012). However, the new information gained from this process might be minor and ineffective for many types of data. An alternative solution to the problem of a small training set is transfer learning. It can be used in the case of well-known deep learning networks that have been previously trained on large general training sets (Shin et al., 2016; Bar et al., 2015; Howard et al., 2017). By changing the last layers of these pre-trained networks, researchers have shown the possibility of receiving high accuracy results after training with a small data set, specifically for medical images (Bar et al., 2015). Another approach is based on generative adversarial networks (GANs). This is an unsupervised learning strategy for generating new data points for a given large unclassified dataset (Goodfellow et al., 2014). This model is composed of two individual networks that compete with each other and learn to synthetically create additional reliable images, which increase the available training set (Radford et al., 2015; Frid-Adar et al., 2018). Alternatively, GANs can be used as a semi-supervised learning technique that utilizes a large number of unclassified images in order to achieve better classification results with a small number of classified images of the same type (Odena, 2016; Salimans et al., 2016). These approaches, however, cannot help if neither a sufficient number of classified nor unclassified training images from the same type are available. Another approach is to train a GAN model on a large set of *unclassified* images, and then fine-tune one of its networks on a smaller set of *classified* images (Vondrick et al., 2016; Ahsan et al., 2018). To the best of our knowledge, this approach has not been adapted so far for medical imaging.

In the current paper, we perform classification of quantitative phase images of healthy and cancer cells and of primary cancer and metastatic cancer cells. For these tasks, we successfully apply various machine learning methods, one of which copes with the problem of a small training set by using another set of unclassified images, and is adapted here for medical imaging in the first time. This method, called transferring of pre-trained generative adversarial network (TOP-GAN), exploits the large amount of unclassified data from another set of examples in order to compensate for the lack in classified data. In order to do so, we combine between transfer learning and GANs. TOP-GAN uses a GAN model in order to train a discriminator network on a large number of unclassified images from one type. Then, by changing the last layer of the discriminator with new un-trained layers, we transform the discriminator network into a new classifier network that can be retrained on a small number of classified images from another type. The transformation of the network is done using transfer learning; hence, we train the GAN model on one type of data set, e.g. unclassified sperm cell OPD images, and then implement this information for classification of a smaller classified data set of another type, e.g. classified cancer and non-cancer cell OPD images. To the best of our knowledge, this is the first time that the discriminator network is being used as a transfer learning network for medical imaging in general and for stain-free cell classification in particular.

In this paper, we compare the classification abilities of TOP-GAN and various machine learning methods, including the classic CNN, in the case that only a small and complex dataset of OPD maps of cells is available for training. We show that TOP-GAN yields a more efficient classifier, which can achieve better classification results with a smaller classified training set. Furthermore, this method shown to be much more robust to the selection of the training set than other deep learning approaches.

2. Materials and methods

2.1. Dataset: stain-free quantitative phase maps of cells

To prepare the datasets, we imaged without staining healthy and cancer cell lines and sperm cells from donors using IPM. We chose two pairs of isogenic cell lines: 1) Hs 895.Sk (healthy skin) and Hs 895.T (melanoma), 2) SW 480 (colorectal adenocarcinoma colon cells) and SW 620 (metastatic from lymph node of colorectal adenocarcinoma cells). The cells were purchased from the ATCC. The first comparison (healthy versus cancer) was done between normal human skin fibroblasts (Hs 895.Sk) and fibroblasts from a melanoma tumor (Hs 895.T), where both cell types were isolated from a 48-years-old Caucasian female. The second comparison (primary versus metastatic cancer) was done between cells from colorectal adenocarcinoma in situ (SW 480) and cells from colorectal adenocarcinoma lymph metastasis (SW 620), where both cell types were isolated from the same individual (but not the same individual of the first pair of cell lines). Thus, each pair of cell lines was taken from the same individual and the same organ. The cells were imaged without staining while being unattached and thus were mostly round, as in the case of imaging flow cytometry, which were harder to classify. The first classification task represents a less complex task: discriminating between healthy skin

and cancer cells, while the second classification task represents a more complex task: discriminating between primary cancer and metastatic cancer cells. In addition, we used IPM to image sperm cells from donors. The cell preparation procedures are elaborated in Appendix A.

In order to obtain the OPD topographic maps of the cells, we acquired off-axis image holograms. For imaging the Hs cells, we used the system mentioned in (Roitshtain et al., 2017). For imaging the SW cells, we used the flipping interferometry module (Roitshtain et al., 2016), connected to an inverted microscope. As shown in Fig. 1(a), flipping interferometry is a compact and portable module, so it can be attached into existing clinical microscopes, signifying its advantage for clinical imaging flow cytometry. The microscope was illuminated by a coherent laser (Helium-Neon, 632.8 nm), and a microscope objective (Mitutoyo, 50X, 0.55 NA) was used for imaging. The off-axis image holograms were created on the digital camera (Thorlabs, DCC1545M), positioned right after the interferometric module. Inside the module, a beam splitter splits the beam into a reference beam, which is reflected back and flipped by a retro-reflector, and a sample beam, which is reflected back by a mirror. This configuration requires half of the optical field of view to be empty from sample details. Flipping interferometric geometry is specifically useful for microfluidic channels, since it is easier to make sure that the sample beam half is positioned on the area of the flowing cells, and the other half of the beam, dedicated for the reference beam, is positioned on the bare glass of the channel. Therefore, flipping interferometry can deal with non-sparse samples without creating ghost images, in contrast to modules based on shearing interferometry (e.g., (Moon, 2013)).

We then extracted the OPD maps from the off-axis holograms of the cells by using digital spatial filtering (Girshovitz and Shaked, 2015). As demonstrated in Fig. 1(b–e), the algorithm included a 2D Fourier transform, filtering one of the cross-correlation terms containing the complex wavefront of the sample, and an inverse 2D Fourier transform. The resulting matrix argument was the wrapped phase of the sample. We then subtracted from the wrapped phase map of the sample a phase map, which was extracted from a reference hologram (acquired without the sample present), in order to overcome stationary aberrations and field curvatures. We then created the unwrapped phase map, by using the unweighted least squares phase unwrapping algorithm, and multiplied it by the wavelength of the source divided by 2π, in order to create the quantitative OPD map of the sample, which is defined as follows:

$$OPD_c(x,y) = [\overline{n_c}(x,y) - n_m] \times h_c(x,y) \tag{1}$$

where n_m is the refractive index of the medium, h_c is the thickness profile of the cell, and $\overline{n_c}$ is the cell integral refractive index, which is defined as follows:

$$\overline{n_c}(x,y) = \frac{1}{h_c}\int_0^h n_c(x,y,z)dz \tag{2}$$

No digital propagation was performed since the camera was positioned in the image plane of the sample. All image analysis procedures were carried out using Matlab (R2017b).

The reconstructed OPD maps, each containing multiple cells, were segmented by thresholding, in order to separate the cells from the background, and then cropped into images of single cells. Next, the isolated cells were computationally aligned and centered for further analysis. Using the method described above, we created RGB images with size 128 × 128 × 3 that described the OPD information of the individual cell area only. We chose to use RGB images as the inputs to the deep learning networks rather than grayscale ones (Teare et al., 2017). The transformation from a grayscale image to an RGB image was done by MATLAB using the jet color map. One of the reasons we chose to use RGB images instead of grayscale ones was due to the fact that the comparison process between the TOP-GAN results and other classification methods included also a comparison to MobileNet that was previously trained on the ImageNet database, which includes RGB images. To test if using RGB images degrades the results in comparison to using gray scale images, we inserted to a conventional CNN the original grayscale OPD maps of the cancer/non-cancer cells, and then trained it for classification. The average accuracy results were very similar to the results of the same network that was trained with RGB images.

2.2. TOP-GAN

TOP-GAN is implemented by performing transfer learning with the previously trained discriminator from a GAN model. GAN is an unsupervised learning strategy that utilizes a given large unclassified samples to generate new data points (e.g. images) based on

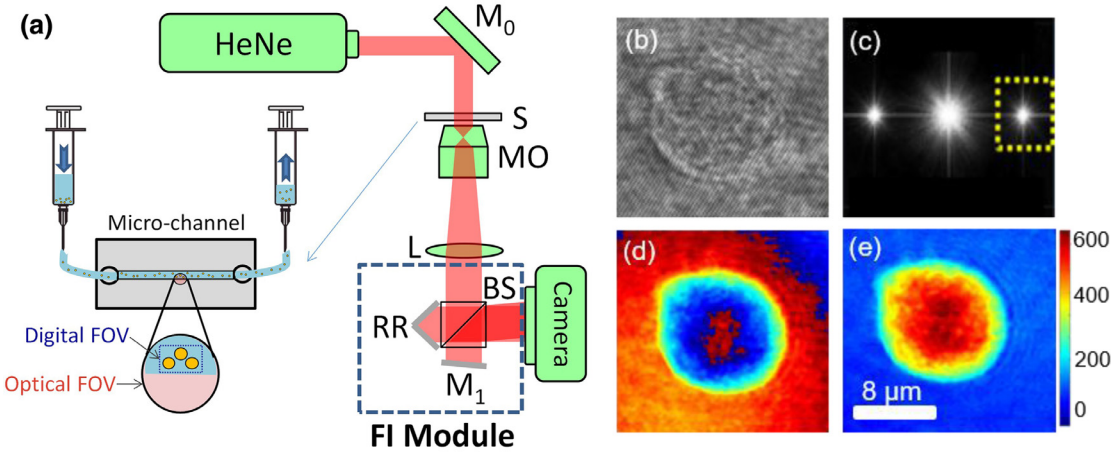

Fig. 1. (a) Flipping interferometry (FI) optical system. The compact interferometer (FI Module) is connected to an inverted microscope. Optical field of view (FOV) is the entire optical beam illuminating the sample, when projected onto the camera plane. Digital FOV is the sample area that the camera acquires (defined by the camera sensor size). Retroreflector (RR) flips the empty half of the optical FOV onto the digital FOV area, so that this half becomes the reference beam of the hologram. HeNe, Helium-Neon laser. M_0, M_1, mirrors. S, cell sample. MO, microscope objective. L, tube lens. BS, beam splitter. Camera, monochrome digital camera. (b) Off-axis hologram of a cancer cell in suspension. (b) Absolute value of the digital Fourier transform of the hologram and the selected cross-correlation term. (d) The resulting wrapped phase map. (e) The final unwrapped OPD map.

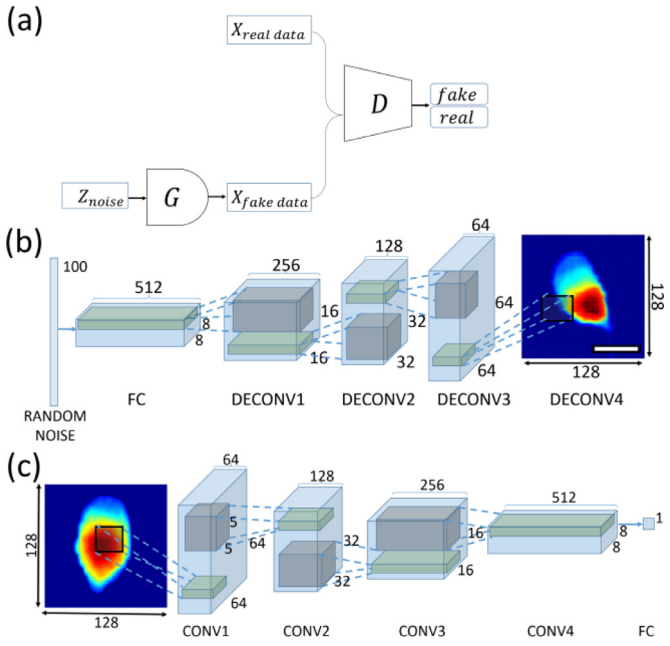

Fig. 2. (a) DCGAN architecture. 'fake' means a generated image, and 'real' means a recorded image. (b) Generator network architecture when generating OPD images of sperm cells. (c) Discriminator network architecture when training on real and generated OPD images of sperm cells.

the same distribution (Goodfellow et al., 2014). This model is based on training two individual CNNs, generator and discriminator. The generator network attempts to synthetically create a realistic images with a similar distribution to the real images, in order to 'fool' the discriminator network, whereas the discriminator have to identify correctly which input image is real (recorded) and which one is fake (generated).

We have based our model on the architecture of deep convolutional GAN (DCGAN), as illustrated in Fig. 2(a) and on the suggested guidelines for stable GAN training proposed in (Radford et al., 2015). The architecture of both networks of the GAN model, generator and discriminator, is illustrated in Fig. 2(b) and Fig. 2(c), respectively.

The generator network takes as an input a vector with a size of 100 elements that are sampled from a random normal distribution. After propagating through the network, it outputs an RGB image that describes an OPD map of a sperm cell with a size of $128 \times 128 \times 3$ pixels. The network architecture consists of a fully connected (FC) layer reshaped to a size of $8 \times 8 \times 512$ and four deconvolution layers with a stride of 2 and a kernel size of 5×5. Batch-normalization layers and ReLU activation functions are applied to all layers except the output layer, which uses a tanh activation function (Radford et al., 2015).

The discriminator network has a classic CNN architecture. The input of the network is an RGB image with a size of $128 \times 128 \times 3$ pixels that describes an OPD map of a cell. The output of the network is a binary decision value that signifies if the image is a real OPD map or a fake (generated) one. The network architecture consists of four convolution layers with a stride of 2 and a kernel size 5×5, and a fully connected layer. A batch normalization layer is applied to each convolution layer except the input and the output layers. Leaky ReLU activation functions, with a slope set to 0.1, are applied to all convolution layers except the output layer, which uses the sigmoid function (Radford et al., 2015). The Sigmoid function gives the probability (0,1) of the image to be 'fake' or 'real' (respectively).

In order to train both the generator and the discriminator networks, we use 2762 different OPD images of human sperm cells captured in our lab by IPM. In order to increase the data set size, we use a data augmentation method, which increases the data set by eight, as will be describe next. During the training process, we use mini-batches of 64 OPD images of human sperm cells, and 64 samples of random noise, Weights are initialized to a zero-centered normal distribution with standard deviation of 0.02. We apply an Adam optimizer with adaptive momentum with the parameters: $\beta_1 = 0.5,\ \beta_2 = 0.99$ and a learning rate of 0.0002 for 75 epochs. We implement the GAN model using the TensorFlow framework (Abadi et al., 2016).

After the GAN is trained on thousands of unlabeled OPD images of human sperm cells, both the generator and discriminator networks have already learned most of the important features that characterize OPD images of biological cells. We thus use the discriminator network from the first section, which has been already trained on sperm cells, and switched its last layer with three untrained fully connected layers. In fact, we have created a classifier that is based on the architecture and knowledge of the discriminator described above. The classifier architecture is the same architecture as the CNN illustrated in Fig. 3, and describes four pre-trained convolution layers followed by three un-trained fully connected layers with sizes of 100, 100, and 2, respectively. In order to reduce the chances of overfitting, each one of the first two un-trained fully-connected layers added to the network is followed by a dropout layer with a probability of 0.5 during training (Srivastava et al., 2014). Like the original discriminator network, the input of the TOP-GAN is an RGB image with a size of $128 \times 128 \times 3$ pixels, which describes an OPD map of a cell. However, the output is a binary decision value that signifies if the image is an image of a healthy skin and cancer cell for the first experiment, and preliminary cancer and metastatic cancer cell for the second experiment, based on a softmax function in the last layer. For the implementation of the TOP-GAN architecture, we use the TensorFlow framework (Abadi et al., 2016). For training, we use an Adam optimizer with a learning rate of 0.00001 and adaptive momentum with the parameters: $\beta_1 = 0.6,\ \beta_2 = 0.99$. The network is trained for 900 epochs or until convergence, whichever comes first.

3. Results

We used IPM to acquire off-axis digital holograms of individual unstained cells for the reconstruction of their OPD maps. Fig. 4(a), (b), (c), and (d) present the OPD maps of unattached cells of types Hs 895.Sk (skin), Hs 895.T (melanoma), SW 480 (colorectal adenocarcinoma colon), and SW 620 (metastatic colorectal adenocarcinoma colon), respectively. These figures demonstrate the difficulty of visually finding significant differences between these groups of cells with a naked eye, even with quantitative phase imaging. This is due to the fact that typically, isolated and unattached cells are mostly round-shaped, and thus are fairly similar to one another (Crowley et al., 2013). This is the case in imaging flow cytometry. Furthermore, we have previously shown (Roitshtain et al., 2017) that none of the classification tasks applied here is trivial, since simple parameters, such as the average OPD value or other single parameters, are not enough to create full separation between the groups.

Our goal is to build a robust, automatic network that can accurately classify between two stages of the cancer cells, even when the number of available classified samples is small. In order to do so, we examine two different tasks with different levels of complexity. In the first experiment, which has a lower complexity level, we classify melanoma cells (Hs 895.T) and healthy skin cells (Hs 895.Sk). In the second experiment, which has a higher complexity level, we classify metastatic cells (SW 620) and primary cells (SW

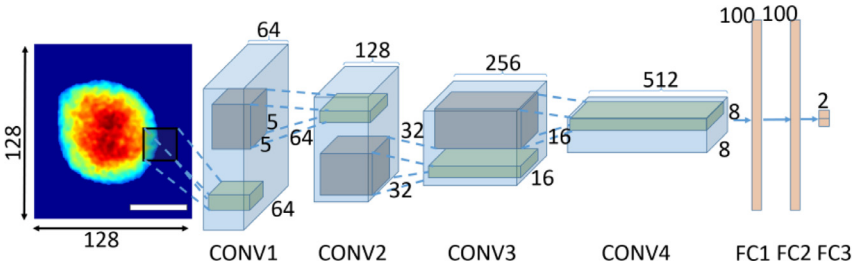

Fig. 3. The architecture of the CNN and of the TOP-GAN for classification of OPD images of cancer cells.

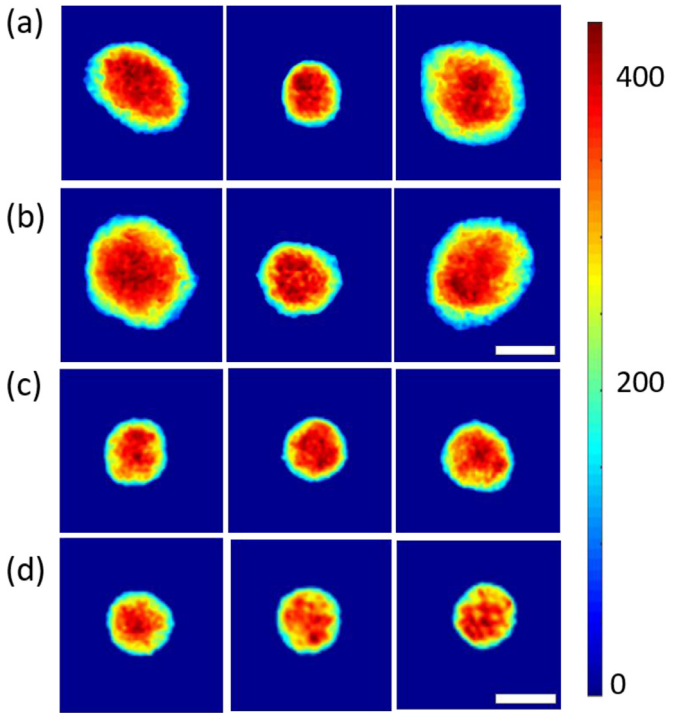

Fig. 4. Quantitative OPD maps of unattached cancer cells, demonstrating the similarity between these groups. (a) Hs 895.Sk (healthy skin cells), (b) Hs 895.T (melanoma cells), (c) SW 480 (colorectal adenocarcinoma colon cells), (d) SW 620 (metastatic colorectal adenocarcinoma colon cells). Color bars represent OPD values in nm. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

480) of colorectal adenocarcinoma colon, imaged during flow. For both experiments, we first use a classic CNN, similar to the one shown in Fig. 3, for classification, and then try to improve its results by different machine learning classification methods, including TOP-GAN. In contrast to TOP-GAN, the classic CNN has not been pre-trained.

To evaluate the classification performance of each method, we use the total classification accuracy. Additionally, we calculate the sensitivity, and specificity for each experiment. These measures are defined as follows:

$$\text{Total Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}, \tag{3}$$

$$\text{Sensitivity} = \frac{TP}{TP+FN}, \tag{4}$$

$$\text{Specificity} = \frac{TN}{TN+FP}, \tag{5}$$

where true (T) and false (F) represent images classified correctly or misclassified, respectively, while positive (P) and negative (N) represent the different categories (e.g. healthy/cancer, primary/metastatic). In order to maximize the sizes of the training and testing sets, and in order to test the effect of the chosen training images on the success of the experiment, we perform five-fold cross-validation. This process includes five iterations, where during each of them we choose different images for training and for testing. Each training procedure is performed separately, and does not influence the results of the other four training steps.

All acquired and synthetic OPD images used for training have the same noise characteristic of OPD images acquired under coherent illumination, rather than incoherent illumination, presenting the worst case scenario for spatial noise.

3.1. Classification using conventional approaches

First, we have implemented a naïve KNN classifier that does not make any assumptions on the data distribution in order to predict the complexity of each experiment. There is a direct relation between the performance of a CNN and the performance of a simple KNN (Cohen et al., 2018). We have chosen to implement KNN in the image space by calculating the Euclidean difference (L1) between the test image and any training images in order to find the K nearest neighbors. This can define the complexity of each class and the similarity between different images in the same class. This classifier provides good classification when the number of classified training images is high relative to the class complexity (Boiman et al., 2008). In our case, $K=9$ has yielded the best accuracy. We first compare healthy and cancer cells. By using a training set of 162 images of melanoma cells and healthy skin cells, the KNN algorithm is able to reach 88% accuracy. This high accuracy describes the low complexity of this dataset, where the images associated with the same class are very similar to each other (Boiman et al., 2008). However, when decreasing the training set size, the value of the mean accuracy decreases as well. For example, by training with 10 images, the KNN algorithm reaches only 69.36% mean accuracy. We then compare primary and metastatic cancer cells. By using a training set of 236 images of primary and metastatic adenocarcinoma cells, the KNN algorithm is able to reach only 82.02% accuracy, lower than the results of the first dataset even when the training set is larger. We thus conclude that images that are associated with each class are less similar to each other than in the first experiment, making it harder to classify them based on a small training set. When decreasing size of the training set, the value of the mean accuracy decreases even further, reaching a mean value of less than 70% with 60 images for training. Furthermore, the KNN results in this experiment have been very sensitive to the training set selection.

Next, we have implemented a conentional CNN, having the same architecture as the TOP-GAN (illustrated in Fig. 3). We first compare healthy and cancer cells. Our CNN is trained as a binary classifier on a training set of 162 classified OPD images of individual melanoma cells and healthy skin cells. The performance of the trained CNN is tested on 40 new OPD maps (20 images of

melanoma cells and 20 images of healthy skin cell). The test images have never been seen previously by the network. The classifier accuracy is surprisingly high, resulting in a mean accuracy of 98.9%, where Hs 895.T is defined as 'positive' and Hs 895.Sk as 'negative'. The high accuracy results for a relatively small training set, with only 162 classified images, can be explained by Cohen et al. (2018) who have demonstrated the connection between the high accuracy results of the naive KNN algorithm and the results of a deep neural network. While reducing the size of the training set, the performance of the conventional CNN drops as well. For example, by training the CNN with 10 images and testing it on 192 images, the network barely reaches 65% accuracy. We then compare primary and metastatic cancer cells. Our CNN is trained as a binary classifier on a training set of 236 classified OPD images of both individual primary adenocarcinoma and metastatic cells. The performance of the trained CNN is tested on 40 new OPD maps of both primary (20) and metastatic (20) cells. The test images have never been seen previously by the network. Also in this experiment, we have created five different training sets in order to evaluate the influence of the training images on the performance of the network, and the accuracy presented is the mean accuracy from all five iterations. In this experiment, the mean accuracy reaches 81.8%, where SW 620 is defined as 'positive' and SW 480 as 'negative', which is much lower than obtained in the first experiment, even when using a larger number of classified images for training. Again, by looking at the result of the KNN classifier for the same number of cells, we find a positive correlation between the success of the KNN and the success of the CNN. Like in the first experiment, as the size of the training set decreases, the accuracy of the classic CNN decreases as well, reaching 72% accuracy with 40 images for training. Furthermore, the selection of training set has significant effect on the results.

Afterwards, we have utilized a MOBILE-NET (Howard et al., 2017) that has been trained on the ImageNet of natural images (Howard et al., 2017), and fine-tuned it on our classified training sets that include OPD images of biological cancer cells from different stages. MOBILE-NET is lightweight in its architecture, and therefore it is useful for transfer learning with a small training set. We first compare healthy and cancer cells. The performance of MOBILE-NET is examined on the same data sets as for the conventional CNN, while reaching higher accuracy for a smaller training size. As shown in Fig. 5(a), MOBILE-NET is able to maintain accuracy of 90% for only 20 images for training and is able to increase the accuracy performance of the conventional CNN by 23% for training of 10 classified images, reaching for 88% accuracy. However, this method is sensitive to the selection of training images, as can be seen in Fig. 5(c). We then compare primary and metastatic cancer cells. As shown in Fig. 5(b), In this case, MOBILE-NET is less effective in the second experiment (primary vs. metastatic), and is able to increase the accuracy of the conventional CNN with only 1% of mean accuracy for the same training set of 236 classified images. As the training set size decreases, the performance of the MOBILE-NET decreases as well, and reaches similar results as the conventional CNN. Also here, the classifier is highly influenced from the selected training set, as illustrated in Fig. 5(d). In fact, the resulting accuracy changes in the range of almost 20%, in the case of 30 training images, based only on the selection of the training set. We have also implemented transfer learning by using a larger network, VGG16, which has been previously trained on ImageNet database as well, and have compared it to the other

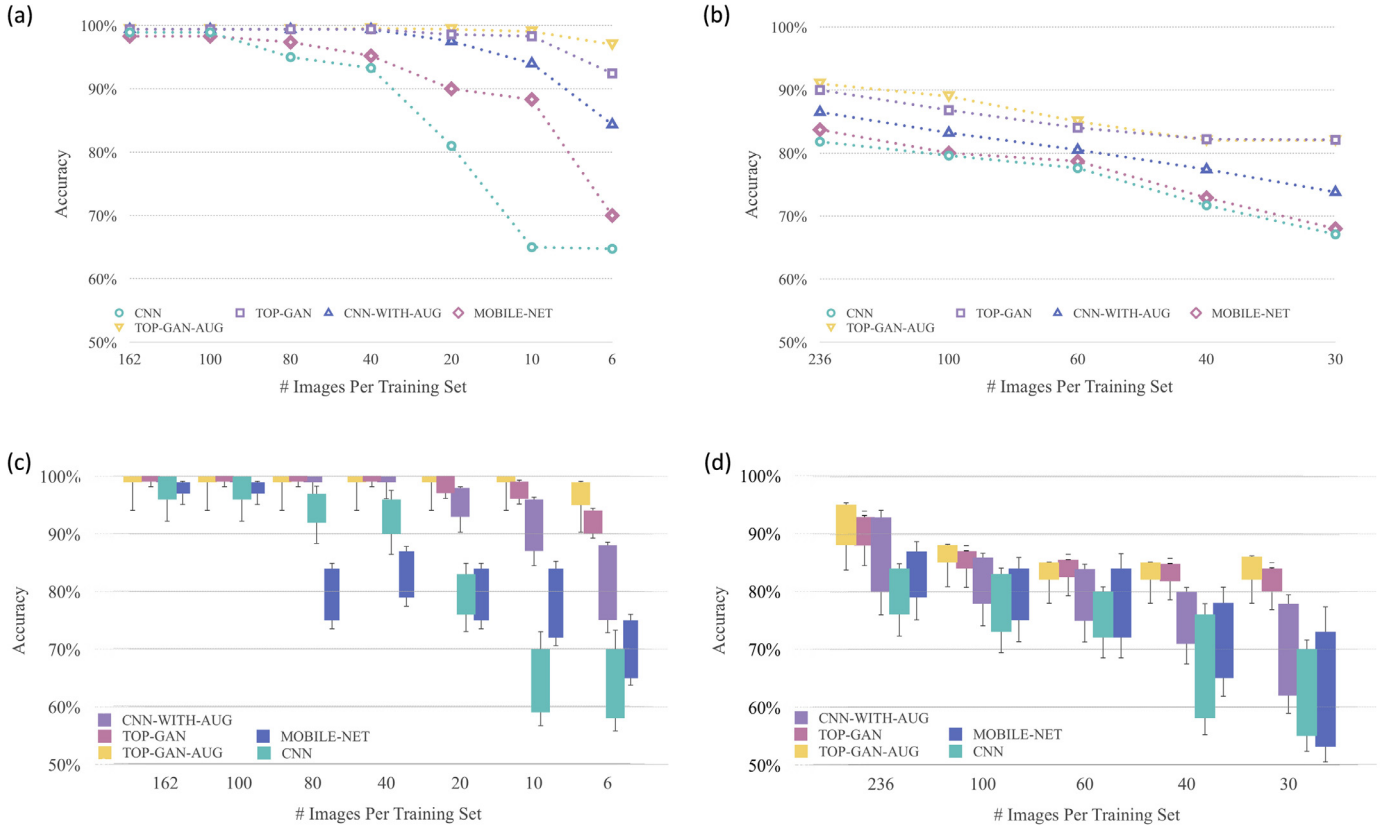

Fig. 5. (a,b) The mean accuracy results calculated on five different training sets with various sizes and different methods (CNN, CNN-WITH-AUG, MOBILE-NET, TOP-GAN, and TOP-GAN-AUG) for classification of melanoma cells and healthy skin cells (a) and for classification of primary and metastatic adenocarcinoma cells (b). (c,d) The range of the accuracy results (percentage) for different sizes of training set. Each color in the graph represents a different method (CNN, CNN-WITH-AUG, MOBILE-NET, TOP-GAN, and TOP-GAN-AUG) for classification of melanoma cells and healthy skin cells (c), and for classification of primary and metastatic adenocarcinoma cells (d).(For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

methods. However, it has yielded worse results than MOBILE-NET presented above.

Finally, we have added data augmentation to the CNN. As described in Krizhevsky et al. (2012), classic data augmentation techniques include rotation, flipping and scaling of the input images. In order to avoid changes in the characteristics and morphology of the cells, we perform OPD image augmentation by combining between flipping and rotation (without scaling). By doing so, we manage to increase our data set by eight times. Following this process, the accuracy of the classification process has increased as well, even while using a smaller training set, as shown in Fig. 5(a) and (b). We first compare healthy and cancer cells. In this experiment, while training on only 10 images and testing the result on the other 192 available images, the augmentation process increases the accuracy from 65% to 94%. Indeed, for the less complex task, data augmentation can improve the results. We then compare primary and metastatic cancer cells. In this experiment, which have a more complex data set (as can be concluded from the KNN results), the improvement in the results when using data augmentation is lower, improving from 81.8% to 86.5%. As shown in Fig. 5(d), this method is also very sensitive to the selection of training image, with accuracy range of 80%–93% for the same training set size.

3.2. Classification using TOP-GAN

TOP-GAN is meant to cope with a situation where only limited data sets of classified (known-class) images are available (classified OPD maps of cancer cells in our case) by creating a transfer learning network that fits *another* types of unclassified (unknown-class) images (unclassified OPD maps of sperm cells in our case). Our classification model is constructed by first building a GANs, based on the architecture of the deep convolutional GAN (DCGAN) model described in Radford et al. (2015) and illustrated in Fig. 2(a). In our case, since thousands of unclassified OPD of sperm cells are available to us, and only a few tens or hundreds of classified OPD maps of cancer cells, the GAN training is performed by using the large number of OPD maps of unclassified sperm cells, and not the small number of classified cancer cells.

The discriminator network that is pre-trained on OPD images of sperm cells becomes a transfer learning network for the classification of other OPD images of biological cells, such as of cancer cells. This technique is based on the idea that most OPD images of isolated and unattached biological cells, especially in flow cytometry, appear highly morphologically similar. Therefore, the general features that characterize OPD maps of cancer cells are similar to those that characterize OPD maps of other biological cells, like sperm cells. Thus, by using GANs to pre-train the discriminator network on a large number of unclassified OPD maps of sperm cells, we allow the network to learn most of the generic features that characterize OPD maps of biological cells before using the small classified dataset. Note that in standard transfer learning, where the weights of a pre-trained network on another class with many examples are used, there is a need to have the data of the other class also classified. In TOP-GAN, on the other hand, we just need unclassified examples from the other class.

During the GAN training, the generator learns the features that characterize the OPD maps of sperm cells and is able to generate new OPD images with the same data distribution as the real OPD images. In Fig. 6(a), the images generated by the generator are illustrated for different stages of the GAN training process (from left to right). As the training process advances, the generated images become more and more similar to the OPD images of actual human sperm cells (see Fig. 6(b)). It is important to note that the GAN model is based on an adversarial process. Therefore, while the generator abilities in creating new images improve as the training progresses, the abilities of the discriminator network in classifying

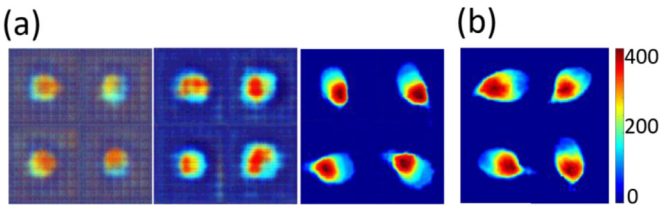

Fig. 6. The generated images from the generator output obtained for training on OPD images of sperm cells in comparison to the OPD images of actual human sperm cells. (a) The generated images from left to right after 10, 20, and 75 epochs, respectively. (b) The experimentally acquired OPD images of human sperm cells. Color bars represent OPD values in nm. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

between these generated images and the real images improves as well. In fact, at the end of the training process, both networks, the generator and the discriminator, become familiar with the different features that characterize the OPD images of the sperm cells. After this preliminary training of the discriminator, we perform transfer learning, replacing the last layer of the discriminator network with three new untrained fully connected layers, and change the classification labels from 'fake' and 'real' images to 'positive' and 'negative' images, according to our classification task. By using a pre-trained network, we manage to create a new classifier that is more efficient in classifying a small number of unseen biological cells, because most of its layers have already been trained on other data sets.

The final accuracy, specificity and sensitivity obtained from this training process are the mean value from all five iterations for each task. The comparison between the different methods (except for the KNN) is illustrated in Fig. 5(a) and (b), for the first and second experiments, respectively. We also checked the robustness of each method to the selected training set by inspecting the range of resulting accuracies, as shown in Fig. 5(c) and (d). Tables 1 and 2 demonstrate the best performance of the TOP-GAN in terms of sensitivity, specificity and area under curve (AUC), for the smallest training set as possible and compare it to other methods (CNN, CNN-WITH-AUG, and MOBILE-NET). Table 1 summarizes the mean results from 5 different folds for specificity, sensitivity and AUC for the first experiment (healthy vs. cancer) by using 10 images for training and 192 images for testing. Table 2 summarizes the same

Table 1
Performance comparison between the different methods (CNN, CNN-WITH-AUG, MOBILE-NET, TOP-GAN, and TOP-GAN-AUG) for Hs 895.Sk (healthy) vs. Hs 895.T (cancer) with 10 images per training and 192 images per testing.

Method	Sensitivity (%)	Specificity (%)	AUC
TOP-GAN-AUG	**98.96**	**99.61**	**0.995**
TOP-GAN	98.08	96.76	0.991
MOBILE-NET	89.21	93.6	0.949
CNN-WITH-AUG	88.28	93.2	0.952
CNN	65.98	66.44	0.717

Table 2
Performance comparison between the different methods (CNN, CNN-WITH-AUG, MOBILE-NET, TOP-GAN, and TOP-GAN-AUG) for SW 620 (primary cancer) vs. SW 480 (metastatic) with 236 images per training and 40 images per testing.

Method	Sensitivity (%)	Specificity (%)	AUC
TOP-GAN-AUG	**93.01**	**93.31**	**0.947**
TOP-GAN	90.61	92.94	0.942
MOBILE-NET	85.01	79.21	0.881
CNN-WITH-AUG	86.01	83.21	0.892
CNN	82.32	78.81	0.878

for the second experiment (primary cancer vs. metastatic) by using 236 images for training and 40 images for testing. All results for all methods are calculated on the same training and testing sets. All training processes are performed using the Google Cloud Platform on the NVIDIA's Tesla P100 GPU.

We first compare healthy and cancer cells. In this case of low complexity data set, there is no need to use TOP-GAN for a training set containing tens of images (e.g., conventional CNN has reached accuracy of 98.9% for 40 images in the training set). However, if a smaller training set is available, TOP-GAN should be used. For example, by training on just 10 classified images (5 melanoma cells and 5 healthy skin cells) TOP-GAN is still able to maintain a mean accuracy of more than 98%, higher than any of the classification method presented before, as shown in Fig. 5(a). Additionally, as shown in Table 1, TOP-GAN yields the highest sensitivity, specificity and area under curve (AUC) in comparison to all other classification methods tested.

We then compare primary and metastatic cancer cells. In this experiment, the accuracy results of the conventional CNN are low. In this case, TOP-GAN reaches accuracy of 89.12% by training on the same 236 images as the CNN, presenting a significant improvement, as shown in Fig. 5(b). Even after reducing the training sets massively, down to only 60 images for training, the TOP-GAN is still able to yield more than 83% accuracy. Additionally, as shown in Table 2, TOP-GAN yields the best sensitivity, specificity and AUC results.

Furthermore, it was noticeable that while other methods yield a wide range of possible accuracy values for the same training size, based only on the selection of the training set, as can be seen in Fig. 5(c) and (d), TOP-GAN yields a much more stable accuracy value for each training size. This signifies another advantage of TOP-GAN, namely being more robust to the selection of the training set.

Note that though our model is over-parameterized and reaches almost zero error on the training set, we believe that it does not suffer from overfitting of the small datasets used due to the following reasons: (i) All state-of-the-art networks achieve almost zero error on the same training data. Yet, this memorization does not contradict their ability to generalize well; (ii) All the methods used attain zero training error. Still, our performance on the test set is better; (iii) We use the same architecture for both the Hs 895.T vs. Hs 895.Sk databases, and the SW 620 vs. SW 480 datasets. For the second, the test error is far from 100%, so it is clear that our model is not overfitting of the testing data. We thus conclude that the very high accuracy in the first case is just to the the low complexity of the data. Indeed, the KNN performance reported in the paper for each of the two cases explains the gap between the two experiments, and coicides with previous studies (Boiman et al., 2008; Cohen et al., 2018).

In addition to pre-training of the GAN model on OPD images of sperm cells, we have also tried to pre-trained the discriminator network on natural images of unclassified human faces and of horses (from CelebA and the ImageNet datasets), as well as on unclassified OPD images of red blood cells. These cells do not have a nucleus, thus have concave morphology, and are less similar in structure to the target dataset (e.g. cancer cells). The accuracy of the classification of the discriminator, which has been pre-trained on each one of these datasets and tested on the second task (SW 480 vs. SW 620), has been higher than the accuracy obtained from the conventional CNN, reaching 82.2% mean accuracy for the natural images and 85.2% mean accuracy for the OPD images. Pre-training on OPD images of red blood cells has yielded higher accuracy than the accuracy obtained when the pre-training on natural images. However, the accuracies resulted when pre-training on both natural images and red blood cell OPD images have been lower than these obtained when the discriminator has been pre-trained on nucleated cells with a more similar morphology by 7% and 4%, respectively. Therefore, we conclude that the features that are extracted from OPD images of biological cells are more relevant than natural images to the performance of the network when the actual classification is done on OPD images of other types of biological cells. Furthermore, there is a positive correlation between the similarity of the pre-trained images to the target classification images and the success of the classification.

For comparison, SVM applied on a similar data have obtained 81% sensitivity and 83% specificity for discriminating between melanoma (Hs 895.T) and healthy skin (Hs 895.Sk) cells, and 82% sensitivity and 81% specificity for discriminating between primary (SW 480) and metastatic (SW 620) colorectal adenocarcinoma cells (Roitshtain et al., 2017).

Finally, by combining data augmentation with the proposed TOP-GAN model, we are able to improve the results even further. We first compare healthy and cancer cells. Here, by training with only 10 classified images and testing them on all other 192 images, we reach 99.054% accuracy. For this best performance method, TOP-GAN-AUG, Table 1 presents sensitivity of 98.96%, specificity of 99.61%, and AUC of 99.51%. We then compare primary and metastatic cancer cells. Here, by training with 236 classified images and testing them on 40 images, we reach 90.6% accuracy. For the best performance method, TOP-GAN-AUG, Table 2 presents sensitivity of 93.01%, specificity of 93.31%, and AUC of 94.73%.

4. Discussion

The method for automatic classification of stain-free individual cells suggested here, TOP-GAN, is based on the combination of deep learning and holographic microscopy, and can be used even with a small classified training set, given that a large unclassified set of other biological cells is available. In our case, a large amount of data from one type (sperm cells) and a small amount of data from other type (cancer cells) are available. We have provided a comparison between many machine learning approaches. First, we have compared TOP-GAN results to those of CNN. Furthermore, we have performed additional experiments with CNN and transfer learning. MobileNet is a CNN that has been previously trained on general images from the ImageNet database. As a prior stage, we have tested other CNNs, such as VGG-16, also pre-trained on ImageNet, but selected MobileNet, as it has achieved better transfer learning performance compared to the pre-trained networks we have checked. After preforming transfer learning on this network, we have fine-tuned it on the target biological images. As shown, the proposed method, TOP-GAN, has still been able to reach higher classification results than those of the MobileNet with transfer learning. We have not provided comparison to methods that require a large amount of unlabeled data, and a small amount of labeled data, both from the same type, since this has not been the case in our paper. We demonstrate that by using a sufficient number of classified images in the training set, the accuracy of the conventional CNN can reach 98.9% when classifying between melanoma cells and healthy skin cells, based only on their OPD maps. In contrast to classic computer vision algorithms, like support vector machine (SVM) or KNN, deep learning, applied directly on the OPD maps, automatically recognizes the important features that characterize cancer cells and healthy skin cells, and uses them in order to distinguish between the different disease stages. Hence, we eliminate the need in subjective and manual selection of features for optimization, and enable rapid and efficient identification of individual cells without the need for staining.

We also demonstrate the weakness of a conventional CNN when the complexity of the data setof the classes increases or the available training set size decreases. We have shown that TOP-GAN, which is based on an integration of two learning approaches,

transfer learning and GAN, is more robust for these cases. By combining TOP-GAN with simple data augmentation we are able to increase the classification results even more. Note that since our network is designed to identify the full structure of each cell, we could not use more advanced augmentation methods, such as elastic augmentation, as it changes the meaning of the size and the shape of each cell. There are many additional augmentation techniques. However, note that all comparisons – CNN versus TOP-GAN, etc. are performed under the same conditions (same types of augmentations). Indeed, the combination of TOP-GAN and classic augmentation (TOP-GAN-AUG) improves the TOP-GAN performance as well. Specifically, in the case of normal human skin cells (Hs 895.Sk) and melanoma tumor cells (Hs 895.T). taken from the same individual, based on only 10 images for training, the classification results reach 98.96% sensitivity, 99.61% specificity and 99.054% accuracy. In the case of comparing between primary colorectal adenocarcinoma colon cells (SW 480) and metastatic from lymph node of colorectal adenocarcinoma cells (SW 620), taken from the same individual, based on 236 images for training, the classification results reach 93.01% sensitivity, 93.31% specificity and 90.6% accuracy. Thus, we eliminate the need to collect a large amount of classified data, which might be a great challenge in various medical and non-medical classification tasks.

Regarding computation time, the GAN pre-training can last up to several hours, and requires a large number of unclassified images, in contrast to simpler methods like data augmentation. However, after pre-training the GAN on a unclassified biological cell images of other types (which are typically more accessible), the TOP-GAN training only depends on the size of the classified training set, which is relatively small, which makes the training process very quick. For example, for 160 images, the training process lasts 3.38 s for TOP-GAN and 13 s for MOBILE-NET, and for testing 40 images, it lasts 0.0226 s for TOP-GAN and 0.62 s for MOBILE-NET.

In contrast to flow cytometry that obtains one accumulative measurement per cell, imaging flow cytometry allows measuring the cell morphology. Current imaging flow cytometers use cell staining, and allow imaging of up to several thousands of cells per second. In this paper, we presented proof-of-concept stain-free quantitative imaging for flow cytometry with much lower throughput of several cells per second. However, in principle, the techniques presented here can reach, with the suitable flow setups and fast cameras, to throughputs similar to the current imaging flow cytometers, but without cell staining and with the possibility of extracting more quantitative data from the cells measured, which create a better basis for machine learning classifiers.

5. Conclusion

We have suggested a new deep learning approach for medical imaging that addresses the problem of a small training set, and applied it for stain-free classification of healthy and cancer cells under quantitative phase microscopy. We believe that our method will provide a valuable tool for an automatic classification process of individual cells for medical diagnosis in flow cytometry, as well as a new classification approach for other medical imaging modalities, where the collection of many training images that are typically needed for deep learning is too expensive or not possible. We thus believe that TOP-GAN will expand the use of deep learning in medical data.

Conflicts of interest

We wish to confirm that there are no known conflicts of interest associated with this publication.

Acknowledgments

This work was supported in part by the H2020 European Research Council (ERC) StG Grant 678316.

Appendix A

A.1. Preparation of human healthy and cancer cell lines

We performed stain-free IPM measurements on Hs 895.Sk (healthy skin cells) and Hs 895.T (melanoma cells), as well as SW 480 (colorectal adenocarcinoma colon cells) and SW 620 (metastatic from lymph node of colorectal adenocarcinoma cells). The complete growth medium used for the Hs cells was Dulbecco's Modified Eagle's Medium (DMEM), supplemented with 10% Fetal Bovine Serum (FBS) (BI) and 2 mM L-glutamine (all purchased from biological industries, Beit Haemek, Israel). The complete growth medium used for the SW cells was BI Roswell Park Memorial Institute (RPMI) 1640 Medium, without L-glutamine supplemented, with 10% FBS, and 2 mM L-glutamine (all purchased from biological industries, Beit Haemek, Israel).

All cell lines were incubated under standard cell culture conditions at 37°C and 5% CO_2 in a humidified incubator until 80% confluence was achieved. Prior to the imaging experiment, the cells were trypsinized for suspension, supplemented adhesive chamber, volume 18 μL, 13 mm diameter $\times$ 0.15 mm thickness, 1.5 mm ports diameter, Sigma Aldrich SN. GBL611101) attached to a cover slip. This chamber induced a contrast thickness value on the entire imaged sample, which is important for the flatness of the final phase map. All cell lines then were imaged by IPM without staining.

A.2. Preparation of human sperm cells

We obtained sperm cells from six human donors at their 20 s. The study was approved by the institutional ethics committee of Tel Aviv University. All sperm donors signed a written informed consent form. After ejaculation, the semen was liquefied at room temperature for 30 min and then the sperm cells were isolated using the PureCeption Bi-layer kit (Origio, Målov, Denmark) in accordance with manufacturer's instructions. The semen and the non-spermatozoa cells were discarded, and the pellet that included sperm cells was resuspended in 5 mL of modified human tubal fluid (HTF) medium (Irvine Scientific, California) and centrifuged at 500 g for 5 min. For fixation, the HTF medium was discarded, the pellet was resuspended with 0.1 ml HTF medium, and then 10 ml fixative solution (3:1 methanol to acetic acid) was added drop by drop. After 5 min at room temperature, the cells were centrifuged at 800 g for 5 min, the supernatant was discarded, and the pellet was resuspended in 0.2 mL of fixative solution. 10 µl of the fixed cells were smeared on 60 $\times$ 20 mm #1 cover slips and put to dry overnight to ensure the evaporation of the fixative solution. Sperm cells were then imaged by IPM without staining.

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, GS., Davis, A., Dean, J., Devin, M., 2016. Tensorflow: lLarge-scale machine learning on heterogeneous distributed systems. arXiv:1603.04467.

Ahsan, U., Sun, C., Essa, I., 2018. Discrimnet: sSemi-supervised action recognition from videos using generative adversarial networks. arXiv:1801.07230.

Bar, Y., Diamant, I., Wolf, L., Greenspan, H., 2015. Deep learning with non-medical training used for chest pathology identification. SPIE Med. Imag. pp. 94140V–94140V https://doi.org/10.1117/12.2083124 .

Boiman, O., Shechtman, E., Irani, M., 2008. In defense of nearest-neighbor based image classification. In: CVPR, pp. 1–8. https://doi.org/10.1109/CVPR.2008.4587598.

Chen, C.L., Mahjoubfar, A., Tai, L.C., Blaby, I.K., Huang, A., Niazi, K.R., Jalali, B., 2016. Deep learning in label-free cell classification. Sci. Rep. 6, 21471. https://doi.org/10.1038/srep21471.

Cohen, G., Sapiro, G., Giryes, R., 2018. DNN or K-NN that is the generalize vs. memorize question. arXiv:1805.06822.

Crowley, E., Di Nicolantonio, F., Loupakis, F., Bardelli, A., 2013. Liquid biopsy: monitoring cancer-genetics in the blood. Nat. Rev. Clin. Oncol. 10, 472–484. https://doi.org/10.1038/nrclinonc.2013.110.

Erhan, D., Manzagol, P.A., Bengio, Y., Bengio, S., Vincent, P., 2009. The difficulty of training deep architectures and the effect of unsupervised pre-training. In: ICAIS, pp. 153–160.

Frid-Adar, M., Diamant, I., Klang, E., Amitai, M., Goldberger, J., Greenspan, H., 2018. GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification. arXiv: 1803.01229.

Girshovitz, P., Shaked, N.T., 2012. Generalized cell morphological parameters based on interferometric phase microscopy and their application to cell life cycle characterization. Biomed. Opt. Express 3, 1757–1773. https://doi.org/10.1364/BOE.3.001757.

Girshovitz, P., Shaked, N.T., 2013. Compact and portable low-coherence interferometer with off-axis geometry for quantitative phase microscopy and nanoscopy. Opt. Express 21, 5701–5714. https://doi.org/10.1364/OE.21.005701.

Girshovitz, P., Shaked, N.T., 2015. Fast phase processing in off-axis holography using multiplexing with complex encoding and live-cell fluctuation map calculation in real-time. Opt. Express 23, 8773–8787.

Go, T., Kim, J.H., Byeon, H., Lee, S.J., 2018. Machine learning-based in-line holographic sensing of unstained malaria-infected red blood cells. J. Biophotonics. 11, e201800101. https://doi.org/10.1002/jbio.201800101.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y., 2014. Generative adversarial nets. In: NIPS, pp. 2672–2680.

Hejna, M., Jorapur, A., Song, J.S., Judson, R.L., 2017. High accuracy label-free classification of single-cell kinetic states from holographic cytometry of human melanoma cells. Sci. Rep. 7, 11943. https://doi.org/10.1038/s41598-017-12165-1.

Howard, A., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., Adam, H., 2017. Mobilenets: efficient convolutional neural networks for mobile vision applications. arXiv: 1704.04861.

Ibrahim, S.F., van den Engh, G., 2007. Flow cytometry and cell sorting. Adv. Biochem. Eng. Biotechnol. 106, 19–39. https://doi.org/10.1007/10_2007_073.

Jaferzadeh, K., Moon, I., 2016. Human red blood cell recognition enhancement with three-dimensional morphological features obtained by digital holographic imaging. J. Biomed. Opt. 21, 126015.

Javidi, B., Moon, I., Yeom, SK., Carapezza, E., 2005. Three-dimensional imaging and recognition of microorganism using single-exposure on-line (SEOL) digital holography. Opt. Express 13, 4492–4506. https://doi.org/10.1007/0-387-28001-4_7.

Jensen, E.C., 2012. Use of fluorescent probes: their effect on cell biology and limitations. Anat Rec. 295, 2031–2036. https://doi.org/10.1002/ar.22602.

Jo, Y., Cho, H., Lee, S.Y., Choi, G., Kim, G., Min, H.S., Park, Y., 2019. Quantitative phase imaging and artificial intelligence: a review. IEEE J. Sel. Top. Quant. Electron. 25, 1–14. https://doi.org/10.1109/JSTQE.2018.2859234.

Jo, Y., Park, S., Jung, J., Yoon, J., Joo, H., Kim, M.H., Kang, S.J., Choi, M.C., Lee, S.Y., Park, Y., 2017. Holographic deep learning for rapid optical screening of anthrax spores. Sci. Adv. 3, e1700606. https://doi.org/10.1126/sciadv.1700606.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. In: NIPS, pp. 1097–1105. https://doi.org/10.1145/3065386.

Lam, V.K., Nguyen, T.C., Chung, B.M., Nehmetallah, G., Raub, C.B., 2017. Quantitative assessment of cancer cell morphology and motility using telecentric digital holographic microscopy and machine learning. Cytometry Part A 93, 334–345. https://doi.org/10.1002/cyto.a.23316.

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436–444. https://doi.org/10.1038/nature14539.

Litjens, G., 2017. A survey on deep learning in medical image analysis. Med. Image Anal. 42, 60–88. https://doi.org/10.1016/j.media.2017.07.005.

Martinez-Torres, C., Laperrousaz, B., Berguiga, L., Boyer-Provera, E., Elezgaray, J., Nicolini, F.E., Maguer-Satta, V., Arneodo, A., Argoul, F., 2015. Deciphering the internal complexity of living cells with quantitative phase microscopy: a multiscale approach. J. Biomed. 20, 96005. https://doi.org/10.1117/1.JBO.20.9.096005.

Mirsky, S.K., Barnea, I., Levi, M., Greenspan, H., Shaked, N.T., 2017. Automated analysis of individual sperm cells using stain-free interferometric phase microscopy and machine learning. Cytom. Part A 91, 893–900. https://doi.org/10.1002/cyto.a.23189.

Moon, I., 2013. Identification of malaria infected red blood cells via digital shearing interferometry and statistical inference. IEEE Photonics J. 5, 6900207. https://doi.org/10.1109/JPHOT.2013.2278522.

Moon, I., Daneshpanah, M., Javidi, B., Stern, A., 2009. Automated three-dimensional identification and tracking of micro/nanobiological organisms by computational holographic microscopy. Proc. IEEE 97, 990–1010. https://doi.org/10.1109/JPROC.2009.2017563.

Moon, I., Javidi, B., 2008. 3-D visualization and identification of biological microorganisms using partially temporal incoherent light in-line computational holographic imaging. IEEE Trans. Med. Imaging 27, 1782–1790. https://doi.org/10.1109/TMI.2008.927339.

Nativ, A., Shaked, N.T., 2017. Compact interferometric module for full-field interferometric phase microscopy with low spatial coherence illumination. Opt. Lett. 42, 1492–1495. https://doi.org/10.1364/OL.42.001492.

Odena, A., 2016. Semi-supervised learning with generative adversarial networks. arXiv:1606.01583.

Oquab, M., Bottou, L., Laptev, I., Sivic, J., 2014. Learning and transferring mid-level image representations using convolutional neural networks. In: IEEE Conf. CVPR, pp. 1717–1724. https://doi.org/10.1109/CVPR.2014.222.

Park, H.S., Rinehart, M.T., Walzer, K.A., Chi, J.T.A., Wax, A., 2016. Automated detection of P. falciparum using machine learning algorithms with quantitative phase images of unstained cells. PLoS One 11, e0163045. https://doi.org/10.1371/journal.pone.0163045.

Paterlini-Brechot, P., Benali, N.L., 2007. Circulating tumor cells (CTC) detection: clinical impact and future directions. Cancer Lett. 253, 180–204. https://doi.org/10.1016/j.canlet.2006.12.014.

Radford, A., Metz, L., Chintala, S., 2015. Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv:1511.06434.

Rappaz, B., Cano, E., Colomb, T., Kühn, J., Depeursinge, C., Simanis, V., Magistretti, P.J., Marquet, P., 2009. Noninvasive characterization of the fission yeast cell cycle by monitoring dry mass with digital holographic microscopy. J. Biomed. Opt. 14, 34049. https://doi.org/10.1117/1.3147385.

Rivenson, Y., Liu, T., Wei, Z., Zhang, Y., Ozcan, A., 2018. PhaseStain: dDigital staining of label-free quantitative phase microscopy images using deep learning. arXiv:1807.07701.

Rivenson, Y., Zhang, Y., Gunaydin, H., Teng, D., Ozcan, A., 2017. Phase recovery and holographic image reconstruction using deep learning in neural networks. arXiv:1705.04286.

Roitshtain, D., Turko, N.A., Javidi, B., Shaked, N.T., 2016. Flipping interferometry and its application for quantitative phase microscopy in a micro-channel. Opt Lett. 41, 2354–2357. https://doi.org/10.1364/OL.41.002354.

Roitshtain, D., Wolbromsky, L., Bal, E., Greenspan, H., Satterwhite, L.L., Shaked, N.T., 2017. Quantitative phase microscopy spatial signatures of cancer cells. Cytometry Part A 91, 482–493. https://doi.org/10.1002/cyto.a.23100.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: convolutional networks for biomedical image segmentation. In: Proc. Int. Conf. Medical Image Comput. Comput.-Assisted Intervention, pp. 234–241. https://doi.org/10.1007/978-3-319-24574-4_28.

Salimans, T., Goodfellow, I., Zaremba, W., Cheung, V., Radford, A., Chen, X., 2016. Improved techniques for training gans. arXiv:1606.03498.

Shaked, N.T., 2012. Quantitative phase microscopy of biological samples using a portable interferometer. Opt. Lett. 37, 2016–2019. https://doi.org/10.1364/OL.37.002016.

Shi, J., Zhou, S., Liu, X., Zhang, Q., Lu, M., Wang, T., 2016. Stacked deep polynomial network based representation learning for tumor classification with small ultrasound image dataset. Neurocomputing 194, 87–94. https://doi.org/10.1016/j.neucom.2016.01.074.

Shin, H.C., Roth, H.R., Gao, M., Lu, L., Xu, Z., Nogues, I., Yao, J., Mollura, D., Summers, R.M., 2016. Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. IEEE Tran. Med. Imaging 35, 1285–1298. https://doi.org/10.1109/TMI.2016.2528162.

Srivastava, N., Hinton, G.E., Krizhevsky, A., Sutskever, I., Salakhutdinov, R., 2014. Dropout: a simple way to prevent neural networks from overfitting. J. Mach. Learn. Res. 15, 1929–1958.

Tajbakhsh, N., Shin, J.Y., Gurudu, S.R., Hurst, R.T., Kendall, C.B., Gotway, M.B., Liang, J., 2016. Convolutional neural networks for medical image analysis: full training or fine tuning. IEEE Tran. Med. Imaging. 35, 1299–1312. https://doi.org/10.1109/TMI.2016.2535302.

Teare, P., Fishman, M., Benzaquen, O., Toledano, E., Elnekave, E., 2017. Malignancy detection on mammography using dual deep convolutional neural networks and genetically discovered false color input enhancement. J. Digit. Imaging 30, 499–505. https://doi.org/10.1007/s10278-017-9993-2.

Vondrick, C., Pirsiavash, H., Torralba, A., 2016. Generating videos with scene dynamics. In: NIPS, pp. 613–621.

Watanabe, E., Hoshiba, T., Javidi, B., 2013. High-precision microscopic phase imaging without phase unwrapping for cancer cell identification. J. Opt. Lett. 38, 1319–1321. https://doi.org/10.1364/OL.38.001319.

Wu, P.H., Phillip, J.M., Khatau, S.B., Chen, W.C., Stirman, J., Rosseel, S., Tschudi, K., Van Patten, J., Wong, M., Gupta, S., 2015. Evolution of cellular morphophenotypes in cancer metastasis. Sci. Rep. 5, 18437. https://doi.org/10.1038/srep18437.

Yeom, S., Moon, I., Javidi, B., 2006. Real-time 3D sensing, visualization and recognition of biological microorganisms. Proc. IEEE J. 94, 550–567. https://doi.org/10.1109/JPROC.2006.870691.

Yoon, J., Min-hyeok, K., Kyoohyun, K., SangYun, L., Suk-Jo, K., YongKeun, P., 2017. Identification of non-activated lymphocytes using three-dimensional refractive index tomography and machine learning. Sci. Rep. 7, 6654. https://doi.org/10.1038/s41598-017-06311-y.

Yu, M., Stott, S., Toner, M., Maheswaran, S., Haber, D.A., 2011. Circulating tumor cells: approaches to isolation and characterization. J. Cell Biol. 192, 373–382. https://doi.org/10.1083/jcb.201010021.

Zink, D., Fischer, A.H., Nickerson, J.A., 2004. Nuclear structure in cancer cell. Nat. Rev. Cancer 4, 677–687. https://doi.org/10.1038/nrc1430.