Transferring Microscopy Image Modalities with Conditional Generative Adversarial Networks

Liang Han
Department of Computer Science
Missouri University of Science and Technology
lh248@mst.edu

Zhaozheng Yin
Department of Computer Science
Missouri University of Science and Technology
yinz@mst.edu

Abstract

Phase Contrast (PC) and Differential Interference Contrast (DIC) microscopy are two popular non-invasive techniques for monitoring live cells. Each of these two image modalities has its own advantages and disadvantages to visualize specimens, so biologists need these two complementary modalities together to analyze specimens. In this paper, we investigate a conditional Generative Adversarial Network (conditional GAN), which contains one generator and two discriminators, to transfer microscopy image modalities. Given a training dataset consisting of pairs of images (source and destination) captured on the same set of specimens by DIC and Phase Contrast microscopes, we can train a conditional GAN, and with this well-trained GAN, we can generate the corresponding Phase Contrast image given a new DIC image, vice versa. The preliminary experiments demonstrate that our approach outperforms one state-of-the-arts method, and can provide biologists a computational way to switch between microscopy image modalities, so biologists can combine the advantages of different image modalities to better visualize and analyze specimens over time, without purchasing all types of microscopy image modalities or switching between imaging systems back-and-forth during time-lapse experiments.

1. Introduction

Live cells are normally transparent and invisible to human eyes under bright field microscopes. Though the fluorescence microscopy can stain them with chemical dyes and radiate them with the specific wavelength light, the invasive staining will do harm to the cell viability. Differential Interference Contrast (DIC) and Phase Contrast (PC) microscopy, two noninvasive techniques, were invented to visualize live cells without altering them in the last century [1][2].

Figure 1. Different microscopy image modalities. (a): Phase Contrast microscopy image. (b): Differential interference contrast microscopy image.

1.1. Motivation

DIC and Phase Contrast microscopy are usually simultaneously used to minotor living cells. Some cells imaged by the Phase Contrast microscope are shown in Fig.1(a). The Phase Contrast can image many internal cellular details at a very high resolution. Unfortunately, some shade-off and halo artifacts degrade the image, especially the halos surrounding the periphery of the cell membrances, which obscure the detailed information about intracellular contacts within halo regions dramatically. Fig.1(b) presents the same set of specimens with DIC optics, it does not suffer from the halo artifacts, and presents the periphery of the cellular membranes much more clearly than that with the Phase Contrast microscope, though the internal cellular details are less obvious. In short, the DIC image can present the close proximity of neighboring cells evidently, and the Phase Contrast image can clearly show the internal details of the cells, which renders that the DIC images are more suitable for intercellular studies, such as contact inhibition, while the Phase Contrast images are more suitable for the internal cellular details analysis, such as autophagy.

Accordingly, both Phase Contrast and DIC have its own strength and weaknesses to analyze cells (further thorough

2160-7516/17 $31.00 © 2017 IEEE
DOI 10.1109/CVPRW.2017.118

comparisons can be found from [3] and [4]), and these two image modalities are complementary to each other. Usually, both of these two image modalities are used to monitor the same specimens, which helps biologists to study the specimens from different perspectives. Purchasing both of these two microscopes in the laboratory is a feasible way to get these two image modalities, but it is money consuming. Moreover, when we need to perform high-throughput time-lapse experiments on live cells over days or even weeks, it is infeasible to switch the specimen culturing dish back-and-forth between microscopes since it is too tedious and will involve some technical challenges such as image registration between two image modalities.

These problems motivate us to think whether we can transfer one microscopy image modality to the other by developing some computational algorithms, when only one microscopy imaging system is available. In another word, when only one microscope hardware is available, we aim to create multi-model imaging capabilities in a software way by transferring the captured microscopy image to other modalities. Furthermore, by implementing the image transfer, it is feasible to make a long-term multi-modal observation on the same specimens by monitoring the specimens using one microscope system and then transfer the captured time-lapse image sequence to other image modalities using an efficient computational algorithm, which means the image transfer can help us avoid swapping microscope imaging systems back-and-forth when perform time-lapse multi-modal observations on specimens.

1.2. Related Work

Generative Adversarial Nets (GANs) was first proposed by Ian J. Goodfellow in 2014 [5], and has been attracting increasing interests recently. Wang *et al*. proposed a style and structure generative adversarial network, which contains a style-GAN and a structure-GAN to generate new images [6]. Dong *et al*. presented a general approach which based on deep convolutional and conditional GANs to transform an image from its original form to some synthetic forms [7]. Yoo *et al*. introduced an image-conditional image generation model based on the GANs to transfer an input domain to a target domain in the semantic level, and generate the target image in the pixel level [8]. Isola *et al*. investigated conditional generative adversarial networks as a general-purpose solution to image-to-image translation problems [9].

The success of GANs motivates us to try to transfer the microscopy image modalities with GANs, however, directly adopting the previous methods cannot solve our problem perfect because of some limitations of the DIC images. Accordingly, we propose a new conditional GANs which contains two discriminators to solve our microscopy image modalities transfer problem.

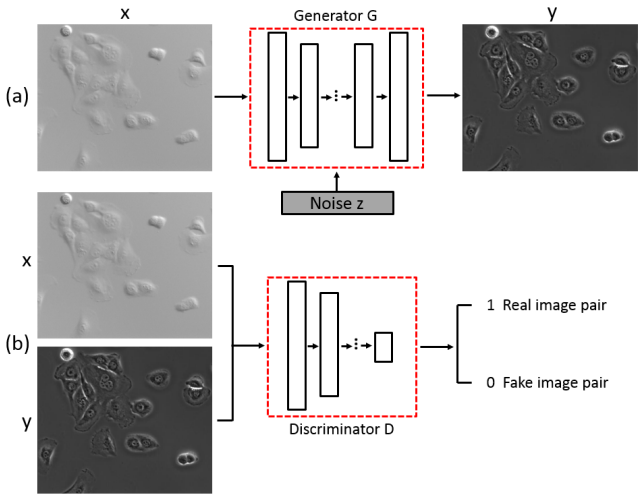

Figure 2. Generator G and discriminator D in the conditional GANs.

1.3. Research Contributions

We investigate a new conditional Generative Adversarial Network (conditional GAN) to transfer microscopy image modalities.

- We improve the structure of the generator in the conditional GAN, which can take more additional information into consideration to generate better destination images.
- We introduce one more discriminator in the conditional GAN, which makes the generated images be distinguished more easily by these two discriminators if the generated images are not good enough.
- We apply the conditional GAN on the microscopy image area, and prove that this technique can do a very good job on transferring microscopy image modality.

2. Background

Generative Adversarial Nets (GANs) were introduced as a novel way to train generative models which can generate realistic images, while the conditional Generative Adversarial Nets were proposed to extend the GANs to a conditional version in which the generative models are conditioned on some aditional information. The biggest difference between GANs and conditional GANs is the input. GANs aim to learn a mapping from a latent random noise vector z to output image y, while conditional GANs try to learn a mapping from a latent random noise vector z and an observed image x to output image y. Both of these two generative models contain two submodels: generator G and discriminator D.

In conditional GANs, the generator G is trained to take latent random noise vector z and observed image x as the input and try to generate realistic images y that cannot be distinguished from the real images by discriminator D (Fig.2 (a)). The discriminator D is trained to perform binary clas-

sification to try to differentiate the images generated by G from the real images (Fig.2 (b)). In another word, generator G and discriminator D act as two adversaries: generator G is trained to generate images which will fool discriminator D to classify the real images and generated images, while D is trained to avoid being fooled by G.

The objective function of a conditional GAN can be formulated as

$$\mathcal{L}_{cGAN}(G,D) = \mathbb{E}_{x,y\sim p_{data}(x,y)}[logD(x,y)]+ \mathbb{E}_{x\sim p_{data}(x),z\sim p_z(z)}[log(1-D(x,G(z|x))], \quad (1)$$

where x is the real source image (e.g., DIC image), y is the real objective image (e.g., Phase Contrast image), $G(z|x)$ is the generated objective image, $D(x,y)$ is the output of the discriminator when taking image pair (x,y) as input. G is trained to minimize this objective function and D is trained to maximize this objective function, i.e., $(G^*,D^*) = arg\min_G \max_D \mathcal{L}_{cGAN}(G,D)$.

3. Methodology

Conditional GANs can be used to do image to image translation, so an intrinsic solution of our microscopy image translation problem is using the conditional GANs to transfer DIC image to Phase Contrast image, or from Phase Contrast image to DIC image. Some previous studies on conditional GANs show that it benefits from combining the GAN objective with a traditional loss function [10], e.g., a ℓ_1 constraint defined as

$$\mathcal{L}_{\ell_1}(G) = \mathbb{E}_{x_1,x_2,y\sim p_{data}(x_1,x_2,y)}[\|y - G(z|x_1,x_2)\|_{\ell_1}]. \quad (2)$$

and the final objective is

$$(G^*,D^*) = arg\min_G \max_D \Big(\mathcal{L}_{cGAN}(G,D) + \mu\mathcal{L}_{\ell_1}(G)\Big), \quad (3)$$

where μ is a weight parameter. The reason why we prefer the ℓ_1 distance to the ℓ_2 distance is that ℓ_1 encourages less blurring than ℓ_2.

However, when the surrounding medium and the specimens have very similar optical path lengths, and the cells' optical path length has very small gradient, neither the DIC image nor the Phase Contrast image can show detailed cell structure and clear cell edges, which makes it very challenging to transfer the microscopy image perfectly, as shown in Fig.7 and Fig.8.

3.1. Cell segmentation

Cell region segmentation can be regarded as the additional information for the conditional GANs to do image translation. Jiang and Yin propose a motion-based DIC image segmentation and restoration algorithm [11]. The tiny motion of each cell pixel is magnified by filtering a time-series of gradient signals on the pixel location using an ideal bandpass filter, while the intensity variation on the background pixels is attenuated. The motion information of a target image is further magnified by a weighted sum of a series of motion images from time-lapse image sequences. With the motion information of cell pixels, we can estimate the cells mask of a DIC image, as shown in Fig.3.

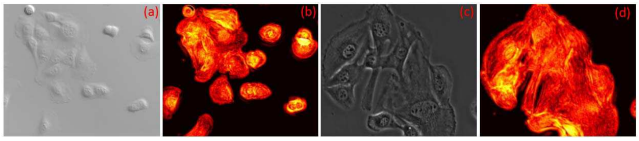

Figure 3. Microscopy image cell segmentation. (a): DIC image. (b): cell mask of the DIC image. (c): Phase Contrast image. (d): cell mask of the Phase Contrast image

3.2. Formulation

With the additional information of cell mask, we can formulate our microscopy image translation problem as a new conditional GAN

$$\begin{aligned}&\mathcal{L}_{cGAN}(G,D_1,D_2) = \\ &\mathbb{E}_{x_1,y\sim p_{data}(x_1,y)}[logD_1(x_1,y)]+ \\ &\mathbb{E}_{x_2,y\sim p_{data}(x_2,y)}[logD_2(x_2,y)]+ \\ &\mathbb{E}_{x_1\sim p_{data}(x_1),z\sim p_z(z)}[log(1-D_1(x_1,G(z|x_1,x_2))]+ \\ &\mathbb{E}_{x_2\sim p_{data}(x_2),z\sim p_z(z)}[log(1-D_2(x_2,G(z|x_1,x_2))],\end{aligned} \quad (4)$$

where x_1 is the real source image (e.g, DIC microscopy image), x_2 is the beforehand obtained cell mask image with the method in [11], and y is the real objective image (e.g., Phase Contrast microscopy image), G is a generator which takes the real source image x_1, the cell mask image x_2, and a latent random noise vector z as the input, and generate some objective images which are very similar to the real ones and cannot be distinguished by the discriminators. D_1 and D_2 are two discriminators which try to differentiate the generated objective images from the real ones by performing binary classification. The input of D_1 is an image pair which is either a pair of real source image and real objective image or a pair of real source image and generated objective image, D_1 is trained to output 1 when the input is a pair of real source image and real objective image, and 0 otherwise. The input of D_2 is an image pair which is either a pair of cell mask image and real objective image or a pair of cell mask image and generated objective image. D_2 is trained to output 1 when the input is a pair of cell mask image and real objective image, and 0 otherwise.

Adding the ℓ_1 constraint to our formulation and we can

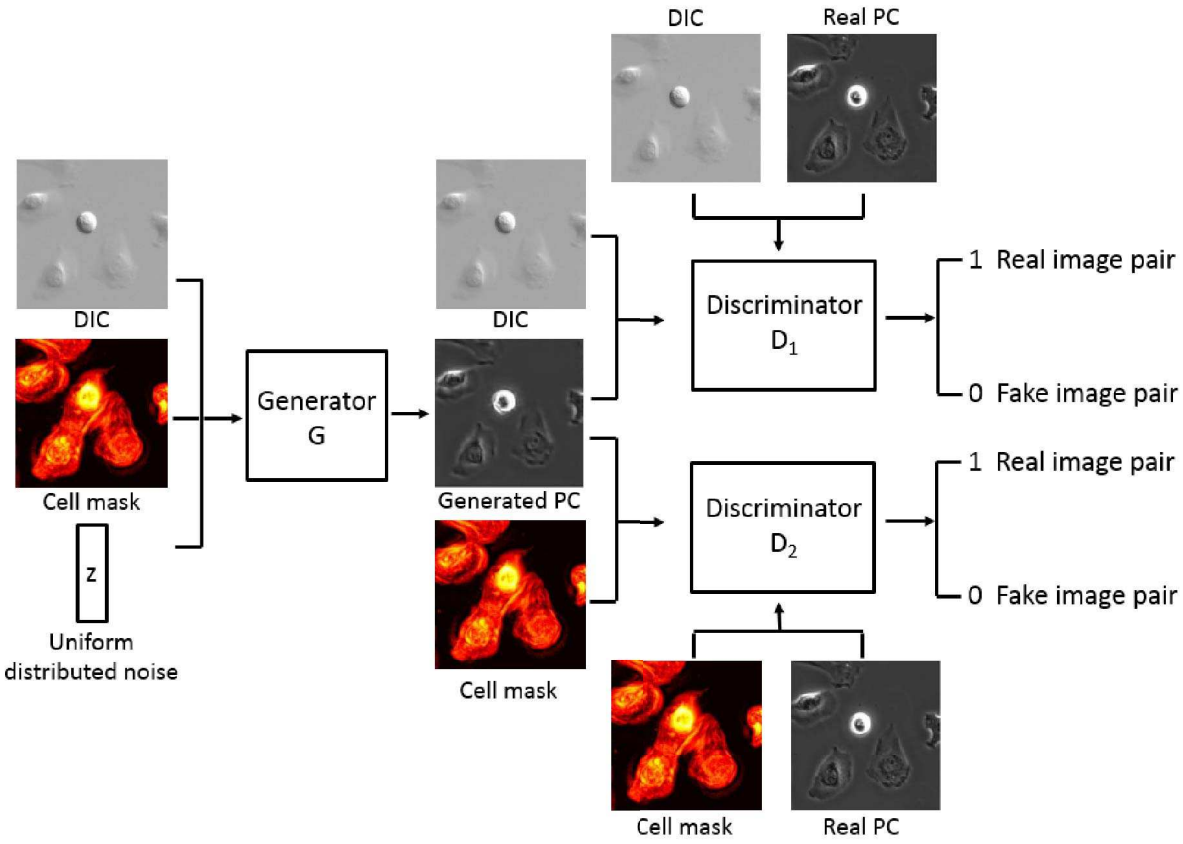

Figure 4. We take the microscopy image transfer from DIC to Phase Contrast (PC) as an example to illustrate the structure of our conditional GAN. Given the ground truth DIC image, cell mask image and uniform distributed noise z as input, the generator G is trained to generate corresponding Phase Contrast image. The discriminator D_1 takes image pair of DIC image and generated Phase Contrast image, and image pair of DIC image and real Phase Contrast image as input, and classify the generated one and real one. The discriminator D_2 takes image pair of cell mask image and generated Phase Contrast image, and image pair of cell mask image and real Phase Contrast image as input, and classify the generated one and real one.

get our final objective

$$(G^*, D_1^*, D_2^*) = \\ arg\min_G \max_{D_1,D_2} \Big(\mathcal{L}_{cGAN}(G, D_1, D_2) + \mu\mathcal{L}_{\ell_1}(G)\Big), \quad (5)$$

The discriminators try to distinguish the realistic images generated by G from the real ones by performing binary classification, while the generator tries to generate images not only fool the discriminators but also similar to the real ones in an ℓ_1 manner. The structure of our conditional GANs is shown in Fig.4.

3.3. Network Design

As the generator and discriminators in our conditional GAN model are all neural networks, we describe their designs in this section.

3.3.1 Generator Architecture

An encoder-decoder network architecture is adapted to design the generator of the conditional GANs in many previous works [12, 6, 10, 13]. In an encoder-decoder network, the input information is sent to pass through a series of progressively downsampling layers. When the transmitting information encounters a bottleneck layer, the process will be reversed, and the transmitting information will pass through a series of progressively upsampling layers (Fig.5 (a)). As we can see, the downsampling layers and the upsampling layers in the encoder-decoder networks are usually symmetric, and the input information will pass through all the layers, including the bottleneck layer.

In our problem, the source image and objective image are DIC microscopy image and Phase Contrast microscopy image that image the same set of specimens, therefore, the source image and objective image should share the same underlying image structure. Taking this constraint into consid-

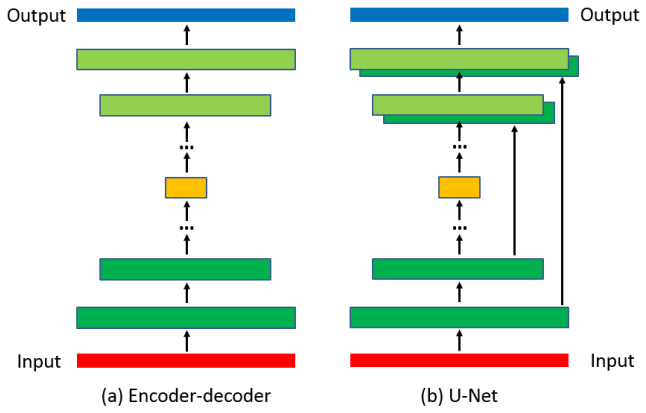

Figure 5. Comparison of Encoder-decoder and "U-Net ". The only difference is that there are skip connections between the symmetric layers in the "U-Net "architecture.

eration, we add skip connections into the encoder-decoder, and form a "U-Net"[14] structure for our generator (Fig.5 (b)). More specifically, suppose there are n layers in the network, between each layer i and layer $n-i$, an skip connection is added, which concatenates all channels at layer i with those at layer $n-i$. The structure of generator is shown in Fig.6 and Tab.1. After the last layer of the decoder, we apply a convolution to map to the number of output channels (3 in our problem), with a Tanh function followed. In the encoder, all the ReLUs are leaky with a slope of 0.2, while in the decoder ReLUs are not leaky.

It is worth noting that we do not provide the noise vector z in the very beginning layers of the generator in addition to x_1 and x_2 as it is proved that the generator will simply learn to ignore the noise. Instead, we add some noise only to several layers of our generator in the form of dropout.

3.3.2 Discriminator architecture

The discriminators D_1 and D_2 in our conditional GANs are multi-layer network architectures, and at the end of each discriminator is a sigmoid function which converts the output value into a $[0, 1]$ range. We can do binary classification easily by thresholding the output of the sigmoid function.

In our final objective function, we use an ℓ_1 constraint to force the generated image to be similar to the ground truth. Unfortunately, the ℓ_1 constraint cannot capture high-frequency correctness perfectly, though it has good ability to encourage low frequencies. This motivates us to find a way to force the conditional GAN discriminator to model the high-frequency structure of the image, based on the fact that the ℓ_1 constraint can capture the low-frequencies.

In order to model the high-frequency information of the image, we borrow the idea of $Patch$GAN from [9], in which the discriminator only penalizes structure at the scale of image patches. Specifically, the input of the discriminator is a pair of image patches with size of $N \times N$ instead of a whole image pair. The discriminator convolutionally runs across a whole image, and all the responses are averaged to provide a final output of this image. The structure of the discriminator is shown in Tab.2. After the last layer of each discriminator, we apply a convolution to map to a one dimensional output, with a Sigmoid function followed. All ReLUs in the discriminators are leaky with a slope of 0.2.

3.4. Optimization and inference

We follow the iterative, numerical approach from [5] to optimize our conditional GAN, i.e., we alternate between one gradient descent step on optimizing D_1 and D_2, and one gradient descent step on optimizing G. In our optimization process, the minibatch Stochastic Gradient Descent (SGD) is used, and the Adam solver is applied [15]. The optimization process is summarized in Alg.1.

After the conditional GAN is trained, we apply the generator onto the source image (e.g., DIC image) to generate the objective image (e.g., Phase Contrast image), achieving our goal of transferring microscopy image modalities.

4. Experiments

To test the effectiveness of our proposed conditional GAN model, we perform some experiments to transfer the microscopy image from one image domain (e.g., DIC image) to another image domain (e.g., Phase Contrast image) on two datasets, and evaluate the quality of the generated images qualitatively and quantitatively.

Dataset: We evaluate our algorithm on two datasets. For each dataset, we collect 1,600 pairs of time-lapse DIC and Phase Contrast images on the same set of specimens over time, and obtain the corresponding 1,600 cell mask images with the method in [11]. 800 triplets consisting with DIC image, Phase Contrast image, and the corresponding cell mask image are selected as the training set, another 400 triplets are chosen as the validation set, and the rest 400 triplets are regarded as the testing set. All images are presented at 256×256 resolution, and the image has pixel value range $[0, 255]$.

Evaluation Metric: Evaluating the quality of generative models is known to be very challenging. Here we use the Normalized Root Mean Square Error $(NRMSE)$ and Structural Similarity Index $(SSIM)$ [16] to quantitatively evaluate our generative model. $NRMSE$ is defined as

$$NRMSE(x, y) = \frac{\sqrt{\frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}(y_{ij}^{*} - y_{ij})^2}}{max - min} \quad (6)$$

where max and min are the maximum and minimum intensities of the ground truth (y_{ij}), respectively. The normalization denominators make this error metric insensitive to the

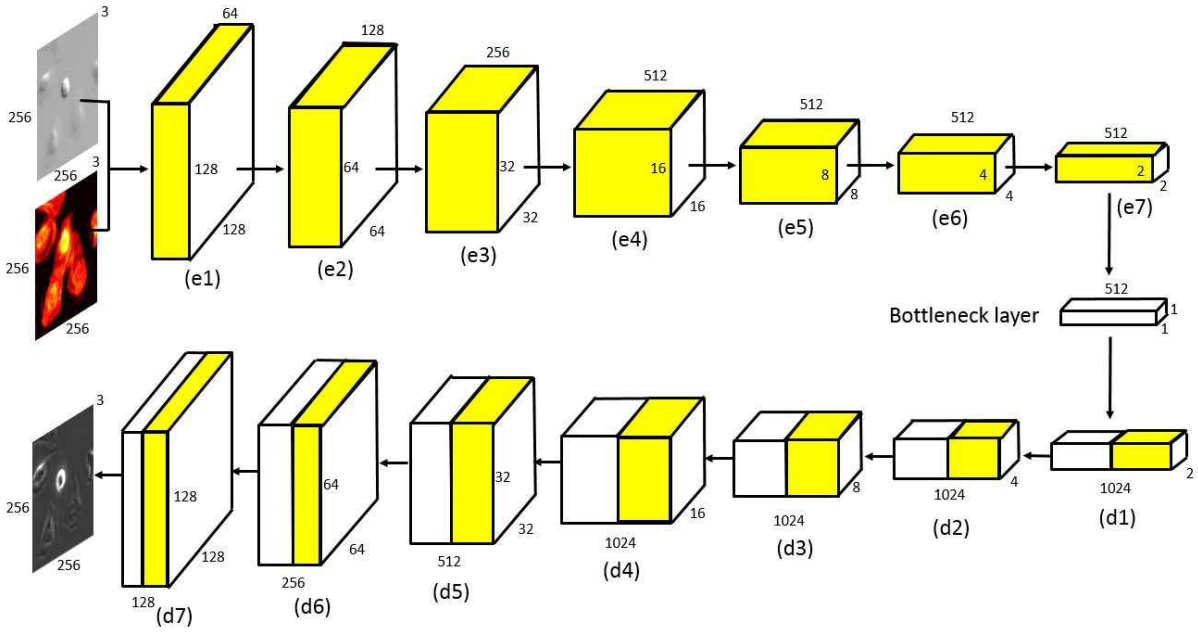

Figure 6. Structure of the generator in our conditional GAN. ei stands for the ith layer of the encoder of the "U-Net ", and dj represents the jth layer of the decoder of the "U-Net ". In the decoder, two symmetric layers are concatenated together. All convolutions are 4×4 spatial filters applied with stride 2. convolutions downsample bu a factor of 2 in the encoder of the "U-Net ", and upsample by a factor of 2 in the decoder of the "U-Net". Dropout noise is added to d1, d2, and d3.

Layer	fc	e1	e2	e3	e4	e5	e6	e7
Module	-	CR	CBR	CBR	CBR	CBR	CBR	CBR
Kernel Number	-	64	128	256	512	512	512	512
Stride	-	2d	2d	2d	2d	2d	2d	2d

Layer	d1	d2	d3	d4	d5	d6	d7	fc
Module	CDBR	CDBR	CDBR	CBR	CBR	CBR	CBR	-
Kernel Number	512	512	512	512	256	128	64	-
Stride	2u	2u	2u	2u	2u	2u	2u	-

Table 1. Structure of the encoder and decoder in the generator. "fc "means fully connected layer, ei stands for the ith layer of the encoder, and dj represents the jth layer of the decoder. "CR "denotes Convolution-ReLU without BatchNorm, "CBR "is Convolution-BatchNorm-ReLU, "CDBR "denotes Convolution-BatchNorm-ReLU with a dropout rate of 50%. "2d "means filtering with stride 2 and downsampling, and "2u "means filtering with stride 2 and upsampling.

Layer	fc	D1	D2	D3	D4
Module	-	CR	CBR	CBR	CBR
Kernel Number	-	64	128	256	512
Stride	-	2d	2d	2d	1

Table 2. Structure of the discriminators D_1 and D_2. "fc "means fully connected layer, Di stands for the ith layer of the discriminator. "CR "denotes Convolution-ReLU without BatchNorm, "CBR "is Convolution-BatchNorm-ReLU. "2d "means filtering with stride 2 and downsampling.

image value scale and image size. $SSIM$ is defined as

$$SSIM(x,y) = l(x,y) \cdot c(x,y) \cdot s(x,y) \qquad (7)$$

where $l(x,y) = \frac{2\mu_x\mu_y+C_1}{\mu_x^2+\mu_y^2+C_1}$, $c(x,y) = \frac{2\sigma_x\sigma_y+C_2}{\sigma_x^2+\sigma_y^2+C_2}$, $s(x,y) = \frac{\sigma_{xy}+C_3}{\sigma_x\sigma_y+C_3}$, $C_1 = (0.01 * L)^2$, $C_2 = (0.03 * L)^2$, and $C_3 = C_2/2$. μ_x, μ_y, σ_x, σ_y, and σxy are the local means, standard deviations, and cross-covariance for images x and y. L is the specified Dynamic Range value of the image.

Parameter Setup: In our experiments, we set batch size to 1, noting little difference between experimental results with different batch sizes as long as enough training iterations. The number of training iterations is 300. We use the Adam optimizer in our experiments, the learning rate for Adam is $2e^{-4}$, and the exponential decay rate for the moment estimate is 0.5. As we use $Patch$GAN, the size of the input patch for the two discriminators is 70×70. We use 5-fold cross validation on the validation set to select the

Algorithm 1 Training of conditional GAN

Require: Training set, weight parameter μ.

for number of training iterations **do**

- Sample minibatch of one noise sample $\{z\}$ from noise prior $p_z(z)$.
- Randomly sample minibatch of one example $\{x_1, x_2, y\}$ from training set.
- Update the discriminator D_1 by ascending its stochastic gradient with Adam optimizer:

$$\nabla_{\theta_{D_1}}[logD_1(x_1, y) + log(1 - D_1(x_1, G(z|x_1, x_2))]$$

- Update the discriminator D_2 by ascending its stochastic gradient with Adam optimizer:

$$\nabla_{\theta_{D_2}}[logD_2(x_2, y) + log(1 - D_2(x_2, G(z|x_1, x_2))]$$

- Update the generator G by descending its stochastic gradient with Adam optimizer:

$$\nabla_{\theta_G}[log(1 - D_1(x_1, G(z|x_1, x_2)) + log(1 - D_2(x_2, G(z|x_1, x_2))]$$

end for

Ensure: Trained generator G, discriminators D_1 and D_2.

weight paramater μ in our final objective function as 100.

Evaluation and Discussion: Fig.7 shows the results of transferring microscopy image modalities from DIC image to Phase Contrast image, and Fig.8 presents the results of transferring from Phase Contrast image to DIC image. These two figures qualitatively evaluate the performance of our conditional GAN compared with a previous state-of-the-art method [9]. Fig.8 shows that our method generates the comparable results with the method from [9]. We can hardly distinguish which one is better. In Fig.7, it is easy to see that our method outperforms the previous method from [9], especially in the area marked by the red circles.

The quantitative evaluations of the proposed conditional GAN and the one in [9] with $NRMSD$ and $SSIM$ are summarized in Tab.3 and Tab.4, respectively. From these two tables, we also can find that our method beats the previous one when transferring microscopy image from DIC to Phase Contrast, and achieves almost the same performance when transferring image from Phase Contrast to DIC.

The reason that our method cannot outperform the prevous conditional GAN when transferring microscopy image from Phase Contrast to DIC is that most of the Phase Contrast images in our dataset have very clear cell edges and the cells can easily be segmented out of the background, which means that the cell mask image cannot offer much useful additional information to determine the cell region. When transferring microscopy image from DIC to Phase Contrast, some cell edges in the DIC images are not easy to be detected, and some cell region cannot be segmented from the background well. In this case, the cell mask images offer us very useful additional information to detect the cell region, and as a result, our conditional GAN based on one generator and two discriminators can benefit from these additional information and generate better Phase Contrast images.

Direction	DIC → PC	PC → DIC
Method of [9]	0.0420	0.0942
Our method	0.0371	0.0912

Table 3. Quantitative evaluation with $NRMSD$.

Direction	DIC → PC	PC → DIC
Method of [9]	0.8686	0.8709
Our method	0.9056	0.8705

Table 4. Quantitative evaluation with $SSIM$.

5. Conclusions

In this paper, we present a conditional GAN, which contains one generator and two discriminators to do microscopy image modality transfer. The generator in the proposed GAN can take additional object segmentation information into consideration which helps to train the generator better. Two discriminators in the GAN can distinguish the generated image from the real one with higher possibility, which means the GAN is forced to generate images much more similar to the real ones. It is worth noting that our model can be generally extended to include more discriminators, which will further help to train a better generator. We both qualitatively and quantitatively evaluate our method on two datasets, and the preliminary experimental results show that the proposed approach can do very well on transferring the microscopy image modality.

Acknowledgement

This project was supported by NSF CAREER award IIS-1351049 and NSF EPSCoR grant IIA-1355406.

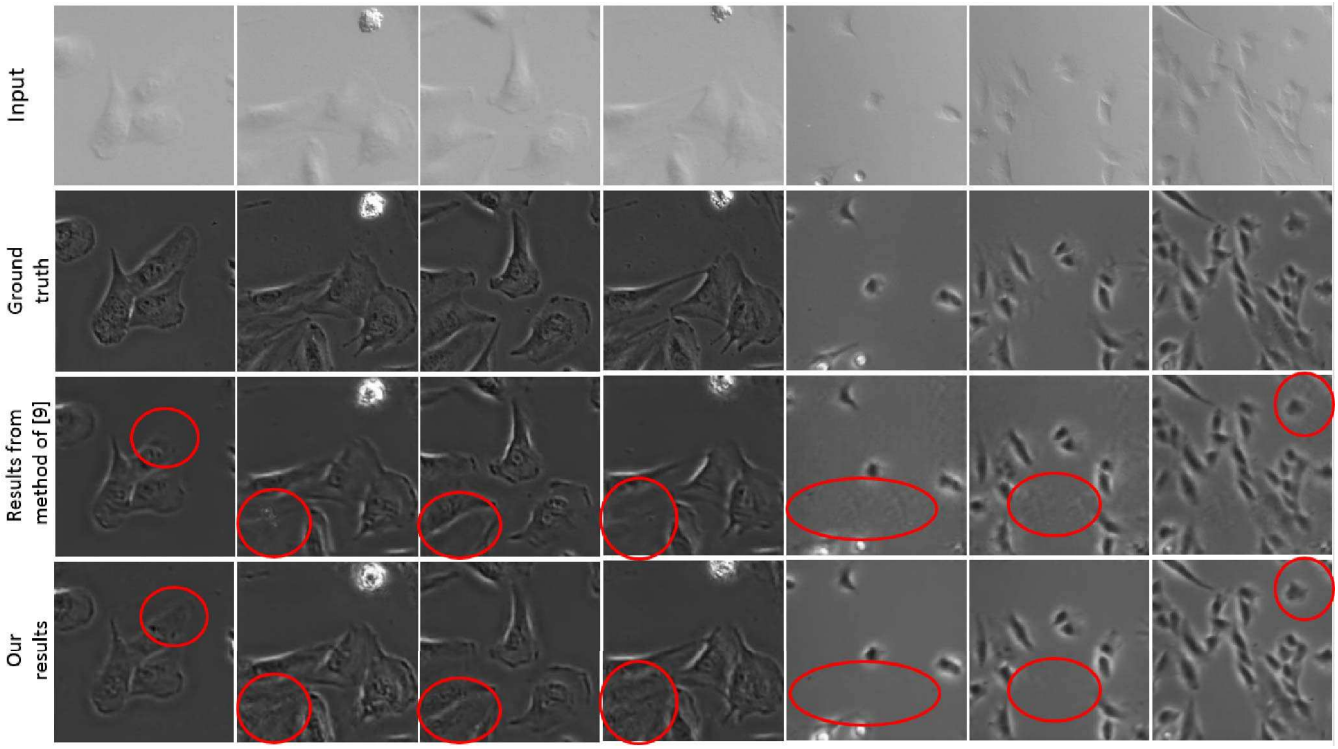

Figure 7. Qualitative evaluation of transferring microscopy image from DIC to Phase Contrast.

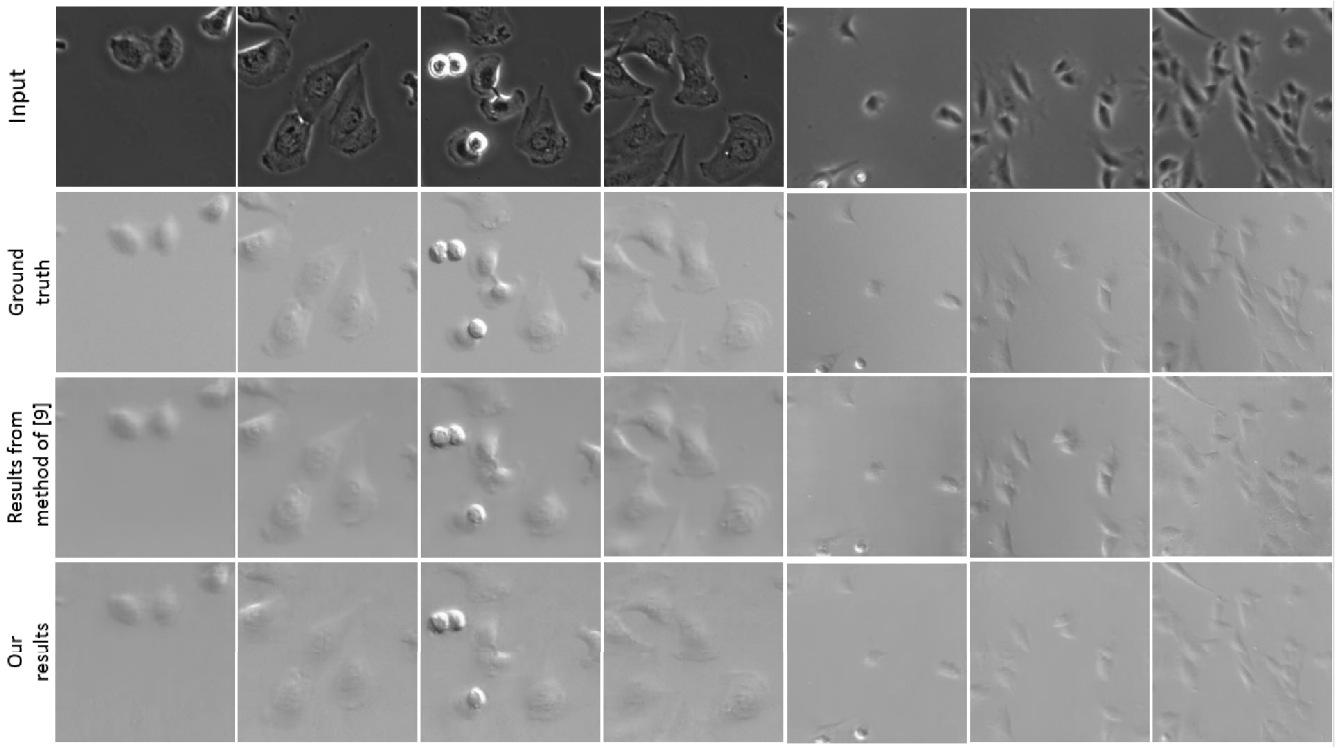

Figure 8. Qualitative evaluation of transferring microscopy image from Phase Contrast to DIC.

References

[1] F. Zernike, "How I discovered phase contrast," *Science*, 1955. 1

[2] D. B. Murphy, "Fundamentals of light microscopy and electronic imaging," *John Wiley and Sons*, 2001. 1

[3] "http://www.olympusmicro.com/primer/techniques/dic/dicphasecomparison.html." 2

[4] "http://www.microscopyu.com/tutorials/java/phasedicmorph/index.html." 2

[5] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," *Advances in neural information processing systems*, 2014. 2, 5

[6] X. Wang and A. Gupta, "Generative image modeling using style and structure adversarial networks," *European Conference on Computer Vision*, 2016. 2, 4

[7] H. Dong, P. Neekhara, C. Wu, and Y. Guo, "Unsupervised image-to-image translation with generative adversarial networks," *arXiv preprint arXiv:1701.02676*, 2017. 2

[8] D. Yoo, N. Kim, S. Park, A. S. Paek, and I. S. Kweon, "Pixel-level domain transfer," *European Conference on Computer Vision*, 2016. 2

[9] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," *arXiv preprint arXiv:1611.07004*, 2016. 2, 5, 7

[10] D. Pathak, P. Krahenbuhl, J. Donahue, T. Darrell, and A. A. Efros, "Context encoders: Feature learning by inpainting," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016. 3, 4

[11] W. Jiang and Z. Yin, "Restoring the invisible details in differential interference contrast microscopy images," *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2015. 3, 5

[12] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," *European Conference on Computer Vision*, 2016. 4

[13] Y. Zhou and T. L. Berg, "Learning temporal transformations from time-lapse videos," *European Conference on Computer Vision*, 2016. 4

[14] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2015. 5

[15] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014. 5

[16] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image qualifty assessment: From error visibility to structural similarity," *IEEE Transactions on Image Processing*, 2004. 5