A Deep Convolutional Neural Network Trained on Representative Samples for Circulating Tumor Cell Detection

Anonymous WACV submission

Paper ID 232

Abstract

The number of Circulating Tumor Cells (CTCs) in blood indicates the tumor response to chemotherapeutic agents and disease progression. In early cancer diagnosis and treatment monitoring routine, detection and enumeration of CTCs in clinical blood samples have significant applications. In this paper, we design a Deep Convolutional Neural Network (DCNN) with automatically learned features for image-based CTC detection. We also present an effective training methodology which finds the most representative training samples to define the classification boundary between positive and negative samples. In the experiment, we compare the performance of auto-learned features from DCNN and hand-crafted features, in which the DCNN outperforms hand-crafted features. We also prove that the proposed training methodology is effective in improving the performance of DCNN classifiers.

1. Introduction

Malignant cells may break away from the primary tumor during the process of cancer metastasis. They may enter the blood or lymphatics system, and then form a secondary tumor at a distant organ site. Several studies have proved that malignant cells are found circulating in blood in the early stages of solid tumor progression, before overt clinical signs [1, 2, 5, 6]. The number of Circulating Tumor Cells (CTCs) in blood can predict disease progression. It indicates tumor response to chemotherapeutic agents [5, 6]. Thus, the development of routine techniques for detection and enumeration of CTCs in clinical blood samples are in need. An automated technique for CTC detection may help in the early diagnosis of cancer even before tumors are visible using traditional imaging approaches. It may also help avoid the use of more invasive diagnosis techniques such as tumor biopsy.

1.1. Related Work

Because CTCs in blood are infrequent, 1 per 1 billion normal cells found in the blood [1], routine detection of CTCs poses a significant challenge. Several methods have been developed to quantify and capture CTCs from human blood. Many of these approaches depend on surface markers expressed on tumor cells which can be exploited for use in positive selection. One widely used marker for detection of carcinoma cells in the blood is epithelial cell adhesion molecule (epCAM) [7]. In some enrichment techniques magnetic beads with immobilized anti-epCAM [8], and other anti-tumor antibodies [9], are used for immunomagnetic separation of malignant cells from the normal blood cell population. Immunomagnetic-based selection of CTCs is attractive because of its simplicity and the availability of the needed tools and reagents. Commercially available systems, based on epCAM-positive selection, have been successfully implemented in CTC evaluation in certain types of carcinoma. Other approaches exploit differences in tumor cell physical properties such as size, density or adhesiveness [10]. To the best of our knowledge, little work has been done on image-based CTC detection.

Figure 1. Visualize Circulating Tumor Cells (CTCs) using Phase Contrast Microscopy Imaging. (a) An image containing CTCs. (b) CTC samples with different shapes and sizes, some of which cluster together. (c) Samples from non-CTC background, which are similar to CTCs in intensity and shape.
1.2. Motivation and Contributions

Methods that use antibodies against tumor cells require prior knowledge of the markers which vary widely according to different types and stages of cancer. These antibody-based systems that efficiently detect carcinomas will miss many other types of malignancies including leukemia, lymphoma and non-epithelial tumors. Svensson et al. [16] use a Bayesian classifier based on a probabilistic generative mixture model to detect CTCs. However, their system is based on fluorescent microscopy images which is an invasive approach. There is a great interest in the CTC community to develop a noninvasive method that is not dependent on tumor cell markers and capable of detection across a wide range of cancer types.

A phase contrast microscopy image containing some CTCs is shown in Fig. 1(a). The CTCs exhibit large variations in shape and size and they overlap with each other (Fig. 1(b)). Some non-CTC background has similar appearance to the CTCs (Fig. 1(c)). It is very hard to distinguish CTCs from the background by simple intensity thresholding or morphological operation.

Therefore, reliable image features are needed to detect CTCs. Deep Convolutional Neural Network (DCNN) has shown its effectiveness on object detection and classification in recent years [3, 4]. It has been proved to be an effective tool in several biomedical applications such as mitosis cell detection [13, 14]. In a DCNN architecture, it has several layers of convolutional filters with each layer followed by either max or mean pooling operations to produce abstract and useful representations of the input object. The parameters of kernels are automatically learned without any human effort. The learned convolutional kernels are more effective feature extractors compared to other human-designed feature descriptors.

The balance between the number of positive and negative samples is quite important in the DCNN training. However, CTCs in blood are infrequent so that it is hard to acquire a large amount of CTC image patches for DCNN training. Meanwhile, there are much more negative samples from the background with redundancies so that it is infeasible to include every possible negative sample in training. Thus, a training methodology which is able to collect the most representative training samples from limited training images is needed to avoid the class imbalance problem.

The above three needs (noninvasive microscopy imaging, image feature extraction, and training with representative samples) motivate us to develop a CTC detection system with the following contributions:

- We propose an image-based CTC detection system based on DCNN. The proposed system is non-invasive without staining markers that damage the viabilities of CTCs.
- An effective training methodology is proposed. It finds the most representative samples to better define the classification boundary between positive and negative samples.

2. Methodology

2.1. Overview

The diagram of our framework is shown in Fig. 2. In the $i$th iteration, DCNN detector $D_i$ is trained from positive and negative training dataset plus the false positives generated from detectors $D_1$ to $D_{i-1}$, and the $i$th detector $D_i$ generates a set of false positives $FP_i$. $FP_i$ is added to the
training dataset to train the detector \( D_{i+1} \) in the \( i + 1 \) iteration. The iteration stops when the performance converges. This sequence of iterations is defined as one ROUND of training in the paper. Then, we apply all \( D_i \)'s (\( i \in [1, N] \)) in one ROUND of \( N \) iterations to all the collected false positive samples during \( N \) iterations. The confident scores \( S_i \) are the output values of \( D_i \) to label the false positive samples as positive. Since we have \( N \) iterations in one ROUND, each false positive will have a \( N \times 1 \) feature vector. K-mean clustering method is applied to classify these false positives based on their feature vectors into two groups: easy samples and hard samples. Only hard samples are added to the original negative training dataset to start another ROUND of iteratively training. We obtain one DCNN detector eventually after multi-ROUNDs of training (each ROUND has multi-iterations), i.e., the final trained detector is the DCNN in the last iteration of the last ROUND.

### 2.2. Data Acquisition

![figure 3](image)

Figure 3. Staining and fluorescence imaging are used to obtain the ground truth. (a) A phase contrast microscopy image containing three CTCs; (b) The corresponding fluorescence image shows the location of CTCs.

We labeled MCF-7 breast cancer cells with a red fluorescence cell-tracker dye for 30 minutes. The MCF-7 cells were mixed with purified sheep red blood cells at a ratio of 1:10,000 (note that, in reality, the CTCs are very rare and the ratio of CTCs to red blood cells is as low as 1:10^9). We increase the ratio to have more positive samples for the purpose of training and validation. Then the CTC samples were mounted onto glass slides using an 18x18 mm coverslip. Fluorescence and phase contrast image sets were acquired using a Leica DMIRE2 epifluorescence microscope equipped with a 10X objective and 12-bit monochrome CCD camera as shown in Fig.3. The bright regions in fluorescence image indicate where are the CTC cells. Note that, the invasive staining process and fluorescence imaging are used to obtain ground truth for training and evaluating our computational algorithms. In our non-invasive CTC detection, the CTCs will not be stained so fluorescence imaging will not be used.

### 2.3. DCNN Architecture

We design a DCNN as shown in Fig.4. The input image patch to DCNN is normalized to 40 \times 40 pixels. In the first layer, it has 6 different convolutional filters with size 5 \times 5 applied on the input images. The convolution operation is formulated as

\[
y^j = \text{sigm}(b^j + \sum_i k^{ij} \ast x^i)
\]

where \( x^i \) and \( y^j \) are the \( i \)-th input map and \( j \)-th output map, respectively, \( b^j \) is the bias term and \( k^{ij} \) is the convolutional kernel between \( x^i \) and \( y^j \). The sigmoid function, \( \text{sigm} \), maps output values to the range of [-1,1].

The first layer is followed by a max-pooling layer which is used to extract local signal in every 2 \times 2 region. Max-pooling function is expressed as

\[
z^j_{p,q} = \max_{0 \leq m,n \leq 2} \{ y^j_{2x(p+m,2x(q+n)} \}
\]

where the pixel at \( (p,q) \) of the output map \( z^j \) pools over a 2 \times 2 region in \( y^j \).

The third layer is a convolutional layer with 12 kernels. The fourth layer is a max-pooling layer. The last layer is fully connected to the output layer by performing the dot product between the weight vector and input vector. The weighted sum is then passed to a sigmoid function.

All the parameters in kernels, bias terms and weight vectors are automatically learned by back propagation with the learning rate equaling to 0.1.

### 2.4. Training Methodology

The locations of positive training samples are automatically obtained around the bright regions in fluorescence images. In order to enhance the tolerance to variations of intensity and rotation, we rotate the phase contrast microscopy images every 30 degrees and automatically crop positive samples from them.

It is important to build a comprehensive negative training dataset in order to precisely define the classification boundary between positive and negative samples. But collecting negative samples which cover every possible location in the background may introduce a lot of repetitive samples and cause a large class imbalance between positive and negative samples. Thus, how to collect a representative set of negative samples becomes crucial.

We propose a bootstrapping method to collect representative negative samples from limited training images.
Figure 5. Iteratively training a detector by adding the false positives in the previous iteration to the training dataset for the next iteration. (a) ROC of each iteration; (b) The 1st iteration (c) The 3rd iteration (d) The 5th iteration. Red: false positive; Green: true positive; Blue: miss detection.

Unlike other training methodologies which train classifiers with all the found false positives until the performance converges, our approach continues to refine the classification boundary by training with the most representative samples among the false positives.

Traditional boosting training method trains the detector iteratively. After one iteration, the detector will collect false positives and add them to negative training dataset, and then start a new training iteration. As shown in Fig. 5(a), the traditional iterative training ends when the performance converges. However, the detector after the iterative training still contains quite some false positives (Fig. 5(d)).

When we apply the trained detector of each iteration on all the false positives, a large amount of false positives generate low responses as shown in the first ROUND training in Fig. 6, which means they can be classified as negative samples relatively easily. As shown in Fig. 7, these relatively easy false positive samples are close to the classification boundary. But the rest small amount of false positive samples with relatively high responses are hard samples far away from the classification boundary. To train a classifier to better classify those hard samples, they should gain more weights in the training.

Suppose we have $N$ iterations in one ROUND of iterative training, then we apply these $N$ detectors on all the false positives collected from all iterations. For each false positive sample, it has a $N \times 1$ confidence score feature vector. The confidence score is the output of a classifier which indicates how likely a false positive sample is classified as positive. The higher the confidence score is, the more likely the false positive sample is classified as positive. We simply apply k-mean clustering method to classify
these false positives based on confidence score feature vectors into two groups: easy samples which have low confidence scores and hard samples which have high confidence scores. To enhance the influence of these hard samples on the training, we start another ROUND of iterative training by only adding these hard samples to the previous negative training dataset. By this iterative training, only a small number of false positives will be collected. As shown in Fig. 6, the number of false positive samples reduces from 11000 in the first ROUND to 2500 in the second ROUND. The proportion of samples with relatively high scores in the second ROUND is larger than that in the first ROUND. Thus hard false positive samples gain more weights in the new training ROUND.

### 3. Experimental Results

#### 3.1. Evaluation Metric

We acquired 45 phase contrast microscopy images, each of which has its corresponding fluorescence image as the ground truth. We randomly select 35 images for training and the rest 10 for testing. To avoid bias, we repeat this random experiment 5 times. The evaluation result is based on the average performance of 5 trials. As defined in PASCAL [15], a detection is a True Positive (TP) if the area of the intersection between the detection window and the ground truth exceeds 50 percent of their union area, otherwise it is a False Positive (FP). If one cell is not detected, it is missed (False Negative, FN). We define precision as $P = \frac{|TP|}{|TP| + |FP|}$, recall as $R = \frac{|TP|}{|TP| + |FN|}$, and F score as the Harmonic mean of precision and recall.

#### 3.2. Comparison of Hand-Crafted Features and Features Learned by DCNN

The Histogram-of-Gradient (HoG, [11]) feature can be used to extract regional gradient information, capturing the shape of objects. The Histogram of Color (HoC) of image patches may be considered as an feature to separate cells from the background. We distribute the color of image patches into 32 bins. In the experiment we feed the HoG + HoC to Support Vector Machine (SVM, [12]) to compare with DCNN.

The average number of positive training samples during the five trials is 1400. Training DCNN classifier takes 2 hours and training SVM takes around 0.5 hour. Note: we only use fluorescence images as ground truth. No information from fluorescence image is extracted as image features for CTC detection.

As shown in Tab. 1, the F score of SVM + HoG + HoC is 78.4% and that of DCNN is to 97%. The F score of DCNN is larger than that of SVM + HoG by 18.6 percentage points in the second ROUND. This result indicates that DCNN finds the better feature than HoG + HoC to detect CTCs. Some detection examples of DCNN are shown in Fig. 9.

#### 3.3. Validation of the Proposed Training Method

We evaluate our training methodology for both SVM and DCNN classifiers. Fig. 8 shows the F score in every iteration of 2 ROUNDs. Both the SVM and DCNN classifiers converge in 5 iterations in each ROUND. The performance of both DCNN and SVM + human-designed feature improve after the first ROUND, which shows that our training method is effective in finding representative samples. Without our training method, the DCNN achieves F score of 91.2%. The F score of SVM + HoG + HoC is 75.4%. The F scores increase to 97% and 78.4% respectively with our training method, as summarized in Table 1.
4. Conclusion and Discussion

In this paper, we proposed an image-based CTC detection system based on DCNN. We also proposed an effective training method which targets at finding the most representative training samples. Comparison of DCNN and SVM classifiers shows that our DCNN classifier works better and the proposed training method is able to improve the performance of classifiers by reducing the redundancy in negative samples. The high performance of DCNN on a challenging dataset shows that it is promising to solve the problem of automated CTC detection in a non-invasive way. Our image-based CTC detection is not dependent on cell marker expression, and is not limited to any particular cancer type. Our detection approach could be adapted to microfluidics devices for accurate and rapid enumeration of CTCs in clinical blood samples for early diagnosis and treatment monitoring.

References


