Multi-Cell Detection and Classification using a Generative Convolutional Model

Florence Yellin
Johns Hopkins University

Benjamin D. Haeffele
Johns Hopkins University

Sophie Roth
IMEC

René Vidal
Johns Hopkins University

Abstract

Detecting, counting, and classifying various cell types in images of human blood is important in many biomedical applications. However, these tasks can be very difficult due to the wide range of biological variability and the resolution limitations of many imaging modalities. This paper proposes a new approach to detecting, counting and classifying white blood cell populations in holographic images, which capitalizes on the fact that the variability in a mixture of blood cells is constrained by physiology. The proposed approach is based on a probabilistic generative model that describes an image of a population of cells as the sum of atoms from a convolutional dictionary of cell templates. The class of each template is drawn from a prior distribution that captures statistical information about blood cell mixtures. The parameters of the prior distribution are learned from a database of complete blood count results obtained from patients, and the cell templates are learned from images of purified cells from a single cell class using an extension of convolutional dictionary learning. Cell detection, counting and classification is then done using an extension of convolutional sparse coding that accounts for class proportion priors. This method has been successfully used to detect, count and classify white blood cell populations in holographic images of lysed blood obtained from 20 normal blood donors and 12 abnormal clinical blood discard samples. The error from our method is under 6.8% for all class populations, compared to errors of over 28.6% for all other methods tested.

1. Introduction

Object detection and classification in images of biological specimens has many potential applications in diagnosing disease and predicting patient outcome. However, due to the wide range of possible imaging modalities, biological data can potentially suffer from low resolution images or significant biological variability from patient to patient. Moreover, many state-of-the-art object detection and classification methods in computer vision require large amounts of annotated data for training, but such annotations are often not readily available for biological images, as the annotator must be an expert in the specific type of biological data. Additionally, many state-of-the-art object detection and classification methods are designed for images containing a small number of object instances per class, while biological images can contain thousands of object instances.

One particular application that highlights many of these challenges is holographic lens-free imaging (LFI). LFI is often used in medical applications of microscopy due to its ability to produce images of cells with a large field of view (FOV) with minimal hardware requirements. However, a key challenge is that the resolution of LFI is often low when the FOV is large, making it difficult to detect and classify cells [1] [6] [4]. The task of cell classification is further complicated due to the fact that cell morphologies can also vary dramatically from person to person, especially when disease is involved. Additionally, annotations are typically not available for individual cells in the image, and one might only be able to obtain estimates of the expected proportions of various cell classes via the use of a commercial hematology blood analyzer.

In prior work [15, 16], LFI images have been used for counting fluorescently labeled white blood cells (WBCs), but not for the more difficult task of classifying WBCs into their various subtypes, e.g., monocytes, lymphocytes, and granulocytes. In [10], the authors suggest using LFI images of stained WBCs for classification, but they do not provide quantitative classification results. Existing work on WBC classification uses high-resolution images of stained cells from a conventional microscope and attempts to classify cells using hand-crafted features and/or neural networks [2, 11, 9, 3]. However, without staining and/or high resolution images, the cell details (i.e., nucleus and cytoplasm) are not readily visible, making our task of WBC classification significantly more difficult. Furthermore, purely data-driven approaches, such as neural networks, typically require large amounts of annotated data to succeed, which is not available for lens free images of WBCs.

Paper contributions. This paper considers the problem of detecting, counting, and classifying various subcategories of WBCs, *e.g.* monocytes, lymphocytes, and granulocytes, in reconstructed lens free images. Unlike typical computer vision problems, each image has hundreds to thousands of instances of each object category and each training image

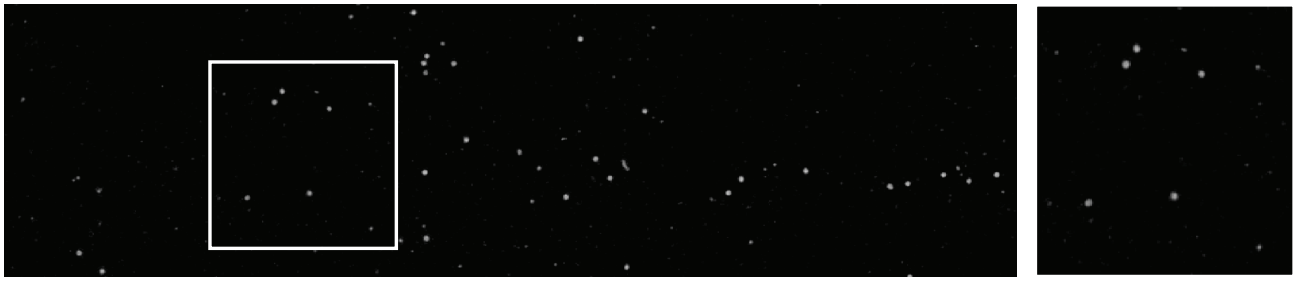

Figure 1: Sample image of WBCs containing a mixture of granulocytes, lymphocytes, and monocytes, in addition to lysed red blood cell debris (left). The region in the white box (zoomed-in, right) shows a typical region of the image, where cells belonging to different classes are sparsely distributed.

is annotated only with the expected number of object instances per class in an image. Thus, a key challenge is that there are no bounding box annotations for any object instances.

We address this challenge by developing a new probabilistic generative model of an image. Conditioned on the total number of objects, the model generates the number of object instances for each class according to a prior model for the class proportions. Then, for each object instance, the model generates the object's location as well as a convolutional template describing the object's appearance. An image is then generated as the superposition of the convolutional templates associated with all object instances.

Given the model parameters, we show that the problem of detecting, counting and classifying object instances in new images can be formulated as an extension of the convolutional sparse coding problem, which can be solved in a greedy manner, similar to [12, 8, 14]. However, unlike the above methods, our generative model utilizes class proportion priors, which greatly enhances our ability to jointly classify multiple object instances, in addition to providing a principled stopping criteria for determining the number of objects for our greedy method. We also address the problem of learning the model parameters from known cell type proportions, which we formulate as an extension of convolutional dictionary learning with priors on class proportions.

We evaluate our proposed convolutional sparse coding method with class proportion priors on LFI images of human blood samples. Our experiments for the task of estimating the proportions of WBCs show that our method clearly outperforms not only standard convolutional sparse coding but also support vector machines and convolutional neural networks. Furthermore, we test the proposed method on blood samples from both healthy donors and donors with abnormal WBC concentrations due to various pathologies which are rare events in the prior model, demonstrating that our method is able to provide promising results across a wide range of biological variability and for cases that are not likely *a priori* under our prior model.

Paper outline. The remainder of the paper is organized as follows. Section 2 describes the probabilistic generative model for WBC images. Section 3 discusses the inference problem for detecting, counting, and classifying cells, and Section 4 discusses the problem of learning the model parameters. Section 5 presents experimental results in counting and classifying WBCs and compares the proposed method to standard convolutional sparse coding and other methods. Section 6 gives the conclusions.

2. Generative Model for Cell Images

Let I be an observed image containing N WBCs, where each cell belongs to one of C distinct classes. Cells from all classes are described by a collection of K class templates $\{d_k\}_{k=1}^K$ that describe the variability of cells within each class. Figure 1 shows a typical LFI image of human blood diluted in a lysing solution that causes the red blood cells to break apart, leaving predominately just WBCs and red blood cell debris. Note that the cells are relatively spread out in space, so we make the assumption that each cell does not overlap with a neighboring cell and that a cell can be well approximated by a single cell template, each one corresponding to a single, known class. The cell templates can thus be used to decompose the image containing N cells into the sum of N images, each containing a single cell. Specifically, the image intensity at pixel (x, y) is generated as

$$I(x,y) = \sum_{i=1}^{N} \alpha_i [d_{k_i} \star \delta_{x_i,y_i}](x,y) + \epsilon(x,y), \qquad (1)$$

where (x_i, y_i) denotes the location of the i^{th} cell, δ_{x_i,y_i} is shorthand for $\delta(x - x_i, y - y_i)$, $\star$ is the 2D convolution operator, k_i denotes the index of the template associated with the i^{th} cell, the coefficient α_i scales the template d_{k_i} to represent the i^{th} cell, and the noise $\epsilon(x, y) \sim N(0, \sigma_I^2)$ is assumed to be i.i.d. zero-mean Gaussian noise with standard deviation σ_I at each pixel (x, y). Under this model, the probability of generating an image I, given that there are N

cells at locations $\mathbf{x} = \{x_i, y_i\}_{i=1}^N$ described by K templates with indices $\mathbf{k} = \{k_i\}_{i=1}^N$ and strengths $\boldsymbol{\alpha} = \{\alpha_i\}_{i=1}^N$ is given by the multivariate Gaussian

$$p(I \mid \mathbf{k}, \boldsymbol{\alpha}, \mathbf{x}, N) = (2\pi\sigma_I^2)^{-\frac{P_I}{2}} \exp\left(-\frac{\|I - \sum_{i=1}^N \alpha_i d_{k_i} \star \delta_{x_i,y_i}\|_F^2}{2\sigma_I^2}\right), \quad (2)$$

where P_I denotes the number of pixels in image I.

To complete the model, we need to define a prior for the distribution of the cells in the image $p(\mathbf{k}, \boldsymbol{\alpha}, \mathbf{x}, N)$. To that end, we assume that the template indices, strengths and locations are independent given N, i.e.,

$$p(\mathbf{k}, \boldsymbol{\alpha}, \mathbf{x}, N) = p(\mathbf{k} \mid N)p(\boldsymbol{\alpha} \mid N)p(\mathbf{x} \mid N)p(N). \quad (3)$$

Therefore, to define the prior model, we simply need to define each one of the terms in the right hand side of (3). Note that this assumption of conditional independence makes sense when the cells are of similar scale and the illumination conditions are relatively uniform across the FOV, as is the case for our data.

To define the prior model on template indices, we model each template d_k as corresponding to one of the C classes, denoted as $\text{class}(k)$. Therefore, given k_i and N, the class s_i of the i^{th} cell is a deterministic function of the template index, $s_i = \text{class}(k_i)$. Next, we assume that all templates associated with one class are equally likely to describe a cell from that class. That is, we assume that the prior distribution of the template given the class is uniform, i.e.,

$$p(k_i \mid s_i) = \frac{1(\text{class}(k_i) = s_i)}{t_{s_i}}, \quad (4)$$

where t_c is the number of templates for class c. We then assume that the prior probability that a cell belongs to a class is independent of the number of cells in the image, i.e., $p(s_i = c \mid N) = p(s_i = c)$. Here we denote the probability of a cell belonging to class c as

$$p(s_i = c) = \mu_c, \quad (5)$$

where $\sum_{c=1}^C \mu_c = 1$. Next, we assume that the classes of each cell are independent from each other and thus the joint probability of all cells being described by templates $\mathbf{k}$ and belonging to classes $\mathbf{s} = \{s_i\}_{i=1}^N$ can be expressed as

$$p(\mathbf{k}, \mathbf{s} \mid N) = \prod_{i=1}^N p(k_i \mid s_i)p(s_i) = \prod_{i=1}^N \frac{\mu_{s_i}}{t_{s_i}} 1(\text{class}(k_i) = s_i) = \prod_{c=1}^C \left(\frac{\mu_c}{t_c}\right)^{n_c} 1(\text{class}(\mathbf{k}) = \mathbf{s}), \quad (6)$$

where $n_c = \sum_{i=1}^N 1(s_i = c)$ is the number of cells in class c. The above equation, together with the constraint $\text{class}(\mathbf{k}) = \mathbf{s}$, completes the definition of $p(\mathbf{k} \mid N)$ as

$$p(\mathbf{k} \mid N) = \prod_{i=1}^N \frac{\mu_{\text{class}(k_i)}}{t_{\text{class}(k_i)}}. \quad (7)$$

To define the prior on the strengths of the cell detections, $\boldsymbol{\alpha}$, we assume that they are independent and exponentially distributed with parameter η,

$$p(\boldsymbol{\alpha}|N) = \frac{1}{\eta^N} \exp\left(\frac{-\sum_{i=1}^N \alpha_i}{\eta}\right), \quad (8)$$

and we note that this is the maximum entropy distribution for the detections under the assumption that the detection parameter is positive and has mean η.

To define the prior on the distribution of the cell locations, we assume a uniform distribution in space, i.e.,

$$p(\mathbf{x} \mid N) = \prod_{i=1}^N \frac{1}{P_I} = \frac{1}{P_I^N}. \quad (9)$$

To define the prior on the number of cells in the image, we assume a Poisson distribution with mean λ, i.e.,

$$p(N) = e^{-\lambda} \frac{\lambda^N}{N!}. \quad (10)$$

Both assumptions are adequate because the imaged cells are diluted, in suspension and not interacting with each other.

In summary, the joint distribution of all the variables of the proposed generative model (see Figure 2 for dependencies among variables) can be written as follows:

$$\begin{aligned} &p(I, \mathbf{k}, \boldsymbol{\alpha}, \mathbf{x}, N) = \\ &p(I \mid \mathbf{k}, \boldsymbol{\alpha}, \mathbf{x}, N)p(\mathbf{k}|N)p(\boldsymbol{\alpha}|N)p(\mathbf{x}|N)p(N) \\ &= \frac{\lambda^N}{e^\lambda (2\pi\sigma_I^2)^{\frac{P_I}{2}} (P_I\eta)^N N!} \exp\left(\frac{-\sum_{i=1}^N \alpha_i}{\eta}\right) \\ &\quad \exp\left(-\frac{\|I - \sum_{i=1}^N \alpha_i d_{k_i} \star \delta_{x_i,y_i}\|_F^2}{2\sigma_I^2}\right) \prod_{i=1}^N \frac{\mu_{\text{class}(k_i)}}{t_{\text{class}(k_i)}}. \end{aligned} \quad (11)$$

3. Inference for Cell Detection, Classification, and Counting

Given an image, we need to detect, count, and classify all the cells and then predict cell proportions. In order to do this inference task, we maximize the log likelihood,

$$\begin{aligned} (\hat{\mathbf{k}}, \hat{\boldsymbol{\alpha}}, \hat{\mathbf{x}}, \hat{N}) &= \arg\max_{\mathbf{k}, \boldsymbol{\alpha}, \mathbf{x}, N} p(\mathbf{k}, \boldsymbol{\alpha}, \mathbf{x}, N \mid I) \\ &= \arg\max_{\mathbf{k}, \boldsymbol{\alpha}, \mathbf{x}, N} \log p(I, \mathbf{k}, \boldsymbol{\alpha}, \mathbf{x}, N). \end{aligned} \quad (12)$$

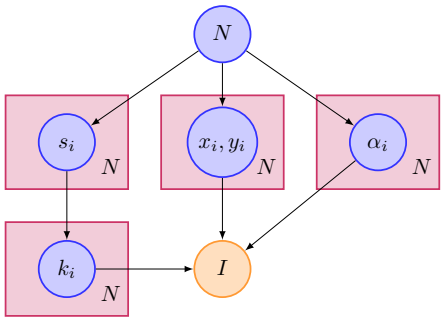

Figure 2: Generative model dependencies for an image.

Assuming the parameters of the modeled distributions are known, the inference problem is equivalent to

$$\min_{\mathbf{k},\boldsymbol{\alpha}>0,\mathbf{x},N}\Bigg[\frac{\|I-\sum_{i=1}^{N}\alpha_i d_{k_i}\star\delta_{x_i,y_i}\|_F^2}{2\sigma_I^2}+\frac{1}{\eta}\sum_{i=1}^{N}\alpha_i - \sum_{i=1}^{N}\log\left(\frac{\mu_{\text{class}(k_i)}}{t_{\text{class}(k_i)}}\right)+N\log\left(\frac{\eta P_I}{\lambda}\right)+\log(N!)\Bigg]. \tag{13}$$

3.1. Cell Detection and Classification

Assume for now that the number of cells N in an image is known. To perform cell detection and classification, we would like to solve the inference problem in Equation (13) over $\mathbf{x}$, $\mathbf{k}$, and $\boldsymbol{\alpha}$. Rather than solving for all N cell detections and classifications in one iteration, we employ a greedy method that uses N iterations, in which each iteration solves for a single cell detection and classification.

We begin by defining the residual image at iteration i as

$$R_i = I - \sum_{j=1}^{i}\alpha_j d_{k_j}\star\delta_{x_j y_j}. \tag{14}$$

Initially, the residual image is equal to the input image, and as each cell is detected, its approximation is removed from the residual image. At each iteration, the optimization problem for $\mathbf{x}$, $\mathbf{k}$, and $\boldsymbol{\alpha}$ can be expressed in terms of the residual as

$$\min_{x_i,y_i,\alpha_i>0,k_i}\Bigg[\|R_{i-1}-d_{k_i}\star\alpha_i\delta_{x_i,y_i}\|_F^2+\frac{2\sigma_I^2}{\eta}\alpha_i - 2\sigma_I^2\log\left(\frac{\mu_{\text{class}(k_i)}}{t_{\text{class}(k_i)}}\right)\Bigg]. \tag{15}$$

Given x_i, y_i and k_i, the solution for $\hat{\alpha}_i$ is given by

$$\hat{\alpha}_i = \frac{\mathcal{S}_{\frac{\sigma_I^2}{\eta}}\left((d_{k_i}\odot R_{i-1})(x_i,y_i)\right)}{\|d_{k_i}\|_F^2}, \tag{16}$$

where $\mathcal{S}_\tau(\alpha)=\max\{\alpha-\tau,0\}$ is the shrinkage thresholding operator and $\odot$ is the correlation operator. We can then solve for the remaining variables in (15) by plugging in the expressions for $\hat{\alpha}_i(x_i,y_i,k_i)$ and simplifying, which leads to

$$(\hat{x}_i,\hat{y}_i,\hat{k}_i)=\arg\max_{x_i,y_i,k_i}\Bigg[\frac{\left((d_{k_i}\odot R_{i-1})(x_i,y_i)-\frac{\sigma_I^2}{\eta}\right)^2}{\|d_{k_i}\|_F^2} +2\sigma_I^2\log\left(\frac{\mu_{\text{class}(k_i)}}{t_{\text{class}(k_i)}}\right)\Bigg]. \tag{17}$$

Note that although at first glance Equation (17) appears to be somewhat challenging to solve as it requires searching over all object locations and templates, the problem can, in fact, be solved very efficiently by employing a max-heap data structure and only making local updates to the max-heap at each iteration, as discussed in [14].

3.2. Cell Counting

Cell counting amounts to finding the optimal value for the number of cells in the image, N, in (13). The objective function for N, plotted in Figure 3 (left), at each iteration is

$$f(N)=\frac{\|R_N\|_F^2}{2\sigma_I^2}+\frac{1}{\eta}\sum_{i=1}^{N}\alpha_i-\sum_{i=1}^{N}\log\left(\frac{\mu_{\text{class}(k_i)}}{t_{\text{class}(k_i)}}\right) +N\log\left(\frac{\eta P_I}{\lambda}\right)+\log(N!). \tag{18}$$

Notice that in the expression for $f(N)$, the residual's norm $\|R_N\|_F^2$ should be decreasing with each iteration as cells are detected and removed from the residual image. Note also that α_i is positive, and $\mu_{si_i}/t_{s_i}<1$, so assuming that $\eta P_I>\lambda$ (which is typically easily satisfied), all terms in the expression for $f(N)$ except the residual term should be increasing with N. This suggests that we stop searching for cells when $f(N)$ begins to increase, i.e., $f(N)>f(N-1)$.

The above condition can be expressed as

$$\frac{\alpha_N}{\eta}-\frac{2R_N\odot d_{k_N}\alpha_N+\|d_{k_N}\|_F^2\alpha_N^2}{2\sigma_I^2}+\log\left(\frac{\eta P_I N t_{s_N}}{\lambda\mu_{s_N}}\right)>0. \tag{19}$$

Moreover, if $R_N\odot d_{k_N}\geq\frac{\sigma_I^2}{\eta}$, it follows from (16) that $R_N\odot d_{k_N}=\alpha_N\|d_{k_N}\|_F^2+\frac{\sigma_I^2}{\eta}$. Substituting this into (19) leads to the following stopping criteria

$$\alpha_N^2<\frac{2\sigma_I^2}{\|d_{k_N}\|_F^2}\log\left(\frac{\eta P_I N t_{s_N}}{\lambda\mu_{s_N}}\right). \tag{20}$$

That is, we should stop cell counting when the square of the strength of the detection decreases below the stopping condition. Notice that the stopping condition is class-dependent, as both μ_c and t_c will depend on which class c is selected to describe the N^{th} cell. Although the stopping criteria for different classes might not fall in the same range, the iterative process will not terminate until the detections

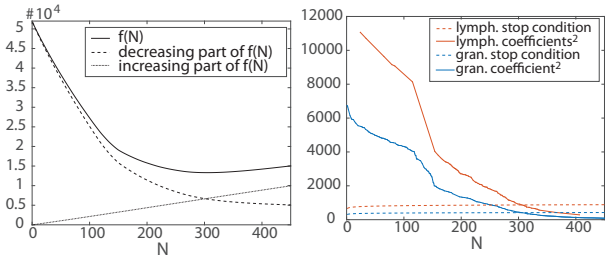

Figure 3: (Left) The greedy cell counting scheme stops at the minimum of $f(N)$. (Right) The stopping condition is class dependent. Only two WBC classes, lymphocytes (lymph.) and granulocytes (gran.), are shown for ease of visualization. The stopping condition is the right hand side of Equation 20, and the squared coefficients are α^2. Both classes reach their stopping condition at around the same iteration, despite having different coefficient values.

from all classes are completed. For example, notice in Figure 3 that although the coefficients for one class (shown in orange) are larger than those for a second class (shown in blue), both cell classes reach their respective stopping conditions at around the same iteration.

The class-dependent stopping condition is a major advantage of our model, compared to standard convolutional sparse coding. Indeed, notice that if the class proportion prior term is eliminated from (18), then the stopping criteria in (20) does not depend on the class because w.l.o.g. one can assume that the dictionary atoms are unit norm, i.e., $\|d_k\| = 1$. As a consequence, the greedy procedure will tend to select classes with larger cells because they reduce the residual term $\|R_N\|_F^2$ more. Our model alleviates this problem because when μ_c is small, the threshold in (20) increases and so our method stops selecting cells from class c.

In summary, the greedy method described by Equations (14), (17) for detecting and classifying cells, together with the stopping condition in Equation (20) for counting cells give a complete method for doing inference in new images.

4. Parameter Learning

In the previous section we described the method used for inferring the latent variables, $\{\boldsymbol{\alpha}, \mathbf{k}, \mathbf{x}, N\}$, of our generative convolutional model in (11) given an image I. However, before we can do inference on new images, we must first learn the parameters $\{\sigma_I, \{d_k\}_{k=1}^K, \eta, \lambda, \{\mu_c\}_{c=1}^C\}$ of our model. In typical object detection and classification models, this is usually accomplished by having access to training data which provides manual annotations of many of the latent variables (for example, object locations and object class). However, our application is uniquely challenging in that we do not have access to manual annotations, so instead we exploit using two datasets for learning our model parameters: 1) a complete blood count (CBC) database of approximately $300,000$ patients of the Johns Hopkins hospital system and 2) LFI images taken of cells from only one WBC subclass obtained by experimentally purifying a blood sample to isolate cells from a single subclass.

Population Parameters. First, to learn the model parameters that correspond to the expected number of cells and the proportions of the various subclasses we utilize the large CBC database, which provides the total number of WBCs as well as the proportion of each subclass of WBC (i.e., monocytes, granulocytes, and lymphocytes) for each of the approximately $300,000$ patients in the dataset. From this, we estimate λ and $\{\mu_c\}_{c=1}^C$ as

$$\lambda = \frac{1}{J_{cbc}} \sum_{j=1}^{J_{cbc}} N^j, \quad \mu_c = \frac{\sum_{j=1}^{J_{cbc}} n_c^j}{\sum_{j=1}^{J_{cbc}} N^j} \tag{21}$$

where $J_{cbc} \approx 300,000$ is the number of patient records in the dataset and (N^j, n_c^j) are the total number of WBCs and number of WBCs of class c, respectively, for patient j (appropriately scaled to match the volume and dilution of blood that we image with our LFI system).

Imaging Parameters. With these population parameters fixed, we are now left with the task of learning the remaining model parameters which are specific to the LFI images $\theta = \{\sigma_I, \{d_k\}_{k=1}^K, \eta\}$. To accomplish this task, we employ a maximum likelihood scheme using LFI images of purified samples which contain WBCs from only one of the subclasses. Specifically, because the samples are purified we know that all cells in an image are from the same known class, but we do not know the other latent variables, so to use a maximum likelihood scheme, one needs to maximize the log likelihood with respect to the model parameters, θ, by marginalizing over the latent variables $\{\boldsymbol{\alpha}, \mathbf{k}, \mathbf{x}, N\}$,

$$\hat{\theta} = \arg\max_{\theta} \sum_{j=1}^{J} \log p(I^j) = \arg\max_{\theta} \sum_{j=1}^{J} \log(\triangle) \tag{22}$$
$$\triangle = \sum_{\mathbf{k}^j, N^j} \iint p(I^j, \boldsymbol{\alpha}^j, \mathbf{k}^j, \mathbf{x}^j, N^j) \, \mathrm{d}\boldsymbol{\alpha}^j \, \mathrm{d}\mathbf{x}^j,$$

where J denotes the number of images of purified samples.

However, solving for the $\hat{\theta}$ parameters directly from 22 is difficult due to the integration over the latent variables $\{\boldsymbol{\alpha}, \mathbf{k}, \mathbf{x}, N\}$. Instead, we use an approximate expectation maximization (EM) algorithm to find the optimal parameters by alternating between updating the latent variables, given the parameters and updating the parameters, given the latent variables. Specifically, note that the exact EM update step for new parameters θ, given current parameters $\hat{\theta}$, is

$$\theta_{EM} = \arg\max_{\theta} \sum_{j=1}^{J} \sum_{\mathbf{k}^j, N^j} \iint \left[p_{\hat{\theta}}(\boldsymbol{\alpha}^j, \mathbf{k}^j, \mathbf{x}^j, N^j | I^j) \right.$$
$$\left. \log \left(p_{\theta}(I^j, \mathbf{x}^j, N^j, \boldsymbol{\alpha}^j, \mathbf{k}^j) \right) \right] \mathrm{d}\boldsymbol{\alpha}^j \, \mathrm{d}\mathbf{x}^j, \tag{23}$$

which can be simplified by approximating with a delta function $p_{\hat{\theta}}(\boldsymbol{\alpha}, \mathbf{k}, \mathbf{x}, N|I) = \delta(\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}}, \mathbf{k} - \hat{\mathbf{k}}, \mathbf{x} - \hat{\mathbf{x}}, N - \hat{N})$, as in [7], where

$$(\hat{\boldsymbol{\alpha}}, \hat{\mathbf{k}}, \hat{\mathbf{x}}, \hat{N}) = \arg\max_{\boldsymbol{\alpha}, \mathbf{k}, \mathbf{x}, N} p_{\hat{\theta}}(\boldsymbol{\alpha}, \mathbf{k}, \mathbf{x}, N|I). \quad (24)$$

The above assumption leads to the approximation

$$\hat{\theta}_{approx} = \arg\max_{\theta} \sum_{j=1}^{J} \log p_{\theta}(I^j, \hat{\boldsymbol{\alpha}}^j, \hat{\mathbf{k}}^j, \hat{\mathbf{x}}^j, \hat{N}^j). \quad (25)$$

Using this approximate EM framework, we then alternate between updating the latent variables given the old parameters and updating the parameters, given the latent variables:

$$(\hat{\boldsymbol{\alpha}}^j, \hat{\mathbf{k}}^j, \hat{\mathbf{x}}^j, \hat{N}^j) = \arg\min_{\boldsymbol{\alpha}^j > 0, \mathbf{k}^j, \mathbf{x}^j, N^j} \left[\frac{\|I^j - \sum_{i=1}^{N^j} \hat{d}_{k_i^j} \star \alpha_i^j \delta_{x_i^j, y_i^j}\|_F^2}{2\hat{\sigma_I}^2} + \frac{\sum_{i=1}^{N^j} \alpha_i^j}{\hat{\eta}} + N^j \log\left(\frac{\eta P_I}{\lambda}\right) + \log(N^j!) \right] \quad (26)$$

$$\text{subject to} \quad \text{class}(k_i^j) = s^j \quad \forall (i,j)$$

and

$$\max_{\theta} \sum_{j=1}^{J} \left[-\frac{\|I^j - \sum_{i=1}^{N^j} \hat{\alpha}_i^j d_{\hat{k}_i^j} \star \delta_{\hat{x}_i^j, \hat{y}_i^j}\|_F^2}{2\sigma_I^2} - \frac{\sum_{i=1}^{N^j} \hat{\alpha}_i^j}{\eta} - \frac{P_I}{2} \log\left(2\pi\sigma_I^2\right) - N^j \log\left(P_I \eta\right) \right]. \quad (27)$$

Note that the latent variable inference in (26) is equivalent to the inference described in Section 3 except that because we are using purified samples we know the class of all cells in the image, s^j, so the prior $p(\mathbf{k}|N)$ is replaced by the constraint on the template classes.

Unfortunately, the optimization problem in Equation 27 that was obtained via approximation is not well defined, since the objective goes to infinity when $\eta \to 0$ and $\hat{\alpha} \to 0$ with the norm of the templates, $\{d_k\}_{k=1}^K$, going to ∞. To address these issues, we fix the signal to noise ratio (SNR) of $\frac{\eta}{\sigma_I^2}$ to a constant and constrain the ℓ_1 norms of the templates to be equal to enforce that the mean value of a pixel for any cell is the same regardless of the class type[1]. Subject to these constraints, we solve (27) for η and the templates by

$$\eta = \frac{\sum_{j=1}^{J} \sum_{i=1}^{N^j} \hat{\alpha}_i^j}{\sum_{j=1}^{J} \hat{N}^j}, \quad d_l = \frac{\sum_{(i,j) \in W} z_i^j}{\sum_{(i,j) \in W} \hat{\alpha}_i^j}, \quad (28)$$

[1]Our images are non-negative, so with our template update scheme the templates are also always non-negative. As a result the ℓ_1 norm is proportional to the mean pixel value of the template.

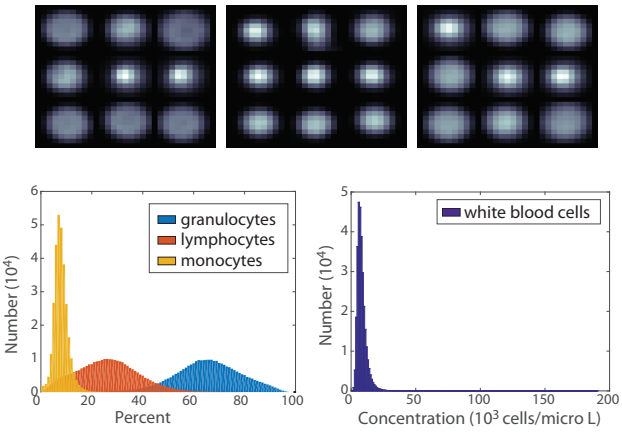

Figure 4: (Top) Learned templates of WBCs. Each template belongs to either the granulocyte (top left), lymphocyte (top middle), or monocyte (top right) class of WBCs. (Bottom) Statistical training data obtained from the CBC dataset. The overlaid histograms of class proportions (bottom left) show that most patients have many more granulocytes than monocytes or lymphocytes. Notice that the histogram of concentrations of WBCs (bottom right) has a long tail.

where $W = \{(i,j) : \hat{k}_i^j = l\}$ and z_i^j is a patch with the same size as the templates, extracted from I^j centered at $(\hat{x}_i^j, \hat{y}_i^j)$. The templates are then normalized to have unit ℓ_1 norm and σ_I is set based on the fixed signal-to-noise ratio, $\sigma_I^2 = \frac{\eta}{\text{SNR}}$, where the SNR is estimated as the ratio of ℓ_2 norms between background patches of the image and patches containing cells. Note that because all of the dictionary updates decouple by training image and each training image contains only one cell class, our procedure is equivalent to learning a separate dictionary for each cell class independently.

5. Results

The proposed cell detection, counting and classification method was tested on reconstructed holographic images of lysed blood, which contain three sub-populations of WBCs (granulocytes, lymphocytes and monocytes) as well as lysed red blood cell debris, such as the image shown in Figure 1. The recorded holograms were reconstructed into images using the sparse phase retrieval method [5], and the absolute value of the complex reconstructed image was used for both training and testing.

5.1. Training Results

Using the purified cell images, we learned the templates shown in Figure 4. Notice that the lymphocyte templates are smaller than the granulocyte and monocyte templates, consistent with what is known about WBCs. The templates have low resolution due to the low resolution, large field of view images obtained with lens-free imaging. To learn

the prior class proportions and the mean number of cells per image, we utilize the database of CBC results. Figure 4 shows histograms of the class proportions of granulocytes, lymphocytes, and monocytes, in addition to a histogram of the total WBC concentrations, from the CBC database.

5.2. Detection, Counting, and Classification Results

Cell detection, counting, and classification with the proposed method was tested on a dataset consisting of lysed blood for 32 donors. The blood comes from both healthy volunteer donors and clinical discards from hospital patients. The clinical discards were selected for having abnormal granulocyte counts, which often coincides with abnormal lymphocyte, monocyte, and WBC counts as well due to various pathologies. We were therefore able to test our method on both samples that are well described by the mean of the probability distribution of class proportions as well as samples that lie on the tail of the distribution.

Our proposed method shows promising results. Figure 5 (left) shows a small region of an image overlaid with detections and classifications predicted by our method. Because we lack ground truth detections and classifications for individual cells in our testing data, we turn to counting and classification results for cell populations to evaluate our method's performance. Each donor's blood was divided into two parts– one part was imaged with a lensfree imager to produce at least 20 images, and the other portion of blood was sent for analysis in a standard hematology analyzer. The hematology analyzer provided ground truth concentrations of WBCs and ground truth cell class proportions of granulocytes, lymphocytes, and monocytes for each donor. By estimating the volume of blood being imaged and the blood's dilution in lysis buffer, we extrapolated ground truth WBC counts per image from the known concentrations.

A comparison of the cell counts obtained by our proposed method and the extrapolated counts obtained from the hematology analyzer is shown in Figure 5 (right). Note that all of the normal blood donors have under 1000 WBCs per image, while the abnormal donors span a much wider range of WBC counts. Observe there is a clear correlation between the counts from the hematology analyzer and the counts predicted by our method. Also note that errors in estimating the volume of blood being imaged and the dilution of blood in lysis buffer could lead to errors in the extrapolated cell counts.

Figure 6 (right) shows a comparison between the class proportion predictions obtained from our method and the ground truth proportions for both normal and abnormal blood donors. As before, we do not have ground truth for individual cells, but for the entire blood sample. Notice once again that the abnormal donors span a much wider range of possible values than do the normal donors. For example,

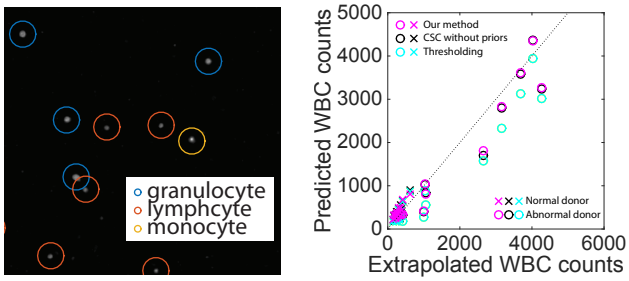

Figure 5: (Left) A a zoomed in region of one image, overlaid with detections and classifications produced by our proposed method. (Right) Results of cell counting. Cell counts estimated by various methods are compared to results extrapolated from a hematology analyzer. The methods shown are thresholding (cyan), CSC without priors (black) and our method (magenta). Results are shown for 20 normal blood donors (x) and 12 abnormal clinical discards (o).

normal donors contain at least 15% lymphocytes, but abnormal donors contain as few as 2% lymphocytes. Despite abnormal donors having WBC differentials widely varying from the distribution mean learned by our model, we are still able to predict their differentials with promising accuracy. Finally, note that WBC morphology can vary from donor to donor, especially among clinical discards. Having access to more purified training data from a wider range of donors would likely improve our ability to classify WBCs.

5.3. Comparison with other methods

To quantify our method, we compare the counting and classification ability of our method to standard convolutional sparse coding (CSC) without priors as described in [14], as well as to support vector machine (SVM), and convolutional neural networks (CNN) classifiers. The SVM and CNN algorithms operate on extracted image patches of detected cells, where the cells were detected via thresholding, filtering detections by size (i.e., discarding objects that were smaller or larger than typical cells).

Figure 5 shows the counting results and Figure 6 shows the classification results obtained by the various methods. Templates used for CSC without priors are trained from purified WBC populations, and the class assigned to each detected cell corresponds to the class of the template that best describes that cell. In terms of total WBC counts, standard CSC performs similarly to our proposed method. This is not surprising, as both methods iteratively detect cells until the coefficient of detection falls beneath a threshold. However, an important distinction is that with standard CSC this threshold is selected via a cross validation step, while in the proposed method the stopping threshold is provided in closed form via (20). Likewise, simple thresholding also achieves very similar but slightly less accurate counts compared to the convolutional encoding methods.

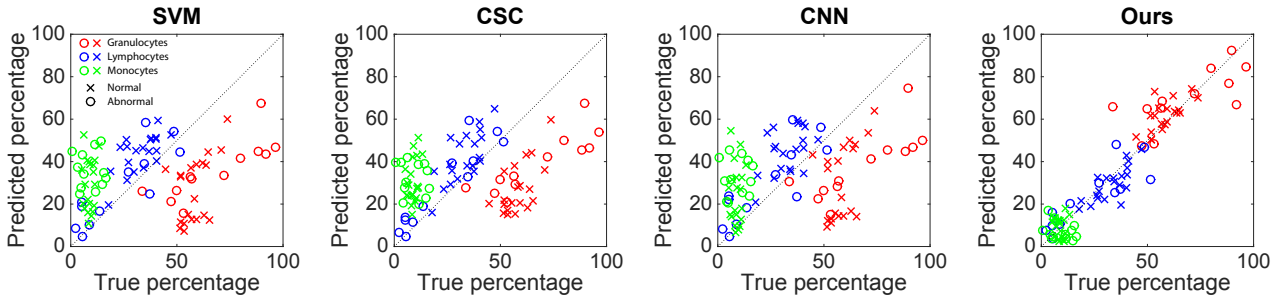

Figure 6: The percentages of granulocytes (red), lymphocytes (blue), and monocytes (green) predicted by various methods are compared to results from a hematology analyzer. The methods, from left to right, are: SVM on patches extracted from images via thresholding, CSC without statistical priors, CNN on patches extracted from images via thresholding, and our method. Results are shown for 20 normal blood donors (x) and 12 abnormal clinical discards (o).

Although in simply counting the number of WBCs per image, the various methods all perform similarly, a wide divergence in performance is observed in how the methods classify cell types as can be seen in the classification results in Table 1. CSC without a statistical model for the class proportions is unable to reliably predict the proportions of granulocytes, lymphocytes, and monocytes in an image, while our proposed method does a much better job. For only normal donors, the proposed method is able to classify all cell populations with absolute mean error under 5%, while standard CSC mean errors are as large as 31% for granulocytes. For the entire dataset, which contains both normal and abnormal blood data, the proposed method achieves on average less than 7% absolute error, while the standard CSC method results in up to 30% average absolute error.

In addition to standard CSC, we also used the cell detections from thresholding to extract cell patches centered at the detections and then classified the extracted cell patches using both a support vector machine (SVM) and a convolutional neural network (CNN). The SVM performed a one-versus-all classification with a Gaussian kernel using cell patches extracted from the images taken from purified samples to train the SVM. Additionally, we implemented a CNN similar to that described in [3]. Specifically, we kept the overall architecture but reduced the filter and max-pooling sizes to account for our smaller input patches, resulting in a network with 3 convolutional layers fed into 2 fully-connected layers with a max-pooling layer between the second and third convolutional layer. Each convolutional layer used ReLU non-linearities and a 3x3 kernel size with 6, 16, and 120 filters in each layer, respectively. The max-pooling layer had a pooling size of 3x3, and the intermediate fully-connected layer had 84 hidden units. The network was trained via stochastic gradient descent using the cross-entropy loss on 93 purified cell images from a single donor. Note that the CNN requires much more training data than our method, which requires only a few training images.

Both the SVM and CNN classifiers perform considerably worse than our proposed method, with the SVM producing errors up to 32%. The CNN achieves slightly better performance than the SVM and standard CSC methods, but errors still reach up to 29%.

Mean Absolute Error	Ours	CSC	SVM	CNN
Granulocytes - normal	4.5	31.1	31.6	27.8
Lymphocytes - normal	4.6	9.5	11.1	12.8
Monocytes - normal	4.7	21.9	20.4	15.9
Granulocytes - all	6.8	30.1	31.8	28.6
Lymphocytes - all	5.6	8.3	10.1	11.6
Monocytes - all	5.5	22.3	22.8	18.9

Table 1: Mean absolute error between ground truth and predicted results for classification are shown for only normal donors and for all donors. Classification results for the three WBC classes are shown for our proposed method, CSC, SVM, and CNN. Note results are for population proportions.

6. Conclusion

We have proposed a generative model for images of populations of WBC together with an algorithm for detection, counting and classification of WBCs based on a convolutional sparse coding method that incorporates statistical population priors. Our method has been tested on and produced promising results on WBC images from both normal donors and clinical discards.

Acknowledgements. The authors thank Evelien Mathieu, Marco Peca, Christian Pick, and Ariel Spence for collecting some of the data used in our experiments, Joshua Betz and Yukari Manabe for sharing WBC class proportions, and Lin Zhou, Murali Jayapala, and Stuart Ray for insightful discussions. This work was funded by miDIAGNOSTICS.

References

[1] O. A and D. U. Ultra wide-field lens-free monitoring of cells on-chip. *Lab Chip*, 8(1):98–106, 2008.

[2] S. F. Bikhet, A. M. Darwish, H. A. Tolba, and S. I. Shaheen. Segmentation and classification of white blood cells. In

Acoustics, Speech, and Signal Processing, 2000. ICASSP'00. Proceedings. 2000 IEEE International Conference on, volume 4, pages 2259–2261. IEEE, 2000.

[3] M. Habibzadeh, A. Krzyżak, and T. Fevens. White blood cell differential counts using convolutional neural networks for low resolution images. In *ICAISC*, pages 263–274. Springer, 2013.

[4] B. Haeffele, S. Roth, L. Zhou, and R. Vidal. Removal of the twin image artifact in holographic lens-free imaging by sparse dictionary learning and coding. In *IEEE International Symposium on Biomedical Imaging*, pages 741–744, 2017.

[5] B. Haeffele, R. Stahl, G. Vanmeerbeeck, and R. Vidal. Efficient reconstruction of holographic lens-free images by sparse phase recovery. In *Medical Image Computing and Computer Assisted Intervention*, pages 109–117. Springer, 2017.

[6] S. B. Kim, H. Bae, K. Koo, M. R. Dokmeci, A. Ozcan, and A. Khademhosseini. Lens-free imaging for biological applications. *Journal of Lab Automation*, 17(1):43–49, 2013.

[7] B. A. Olshausen and B. J. Field. Sparse coding with an overcomplete basis set: a strategy employed by V1? *Vision Research*, 1997.

[8] M. Pachitariu, A. M. Packer, N. Pettit, H. Dalgleish, M. Hausser, and M. Sahani. Extracting regions of interest from biological images with convolutional sparse block coding. In *Advances in Neural Information Processing Systems*, pages 1745–1753, 2013.

[9] V. Piuri and F. Scotti. Morphological classification of blood leucocytes by microscope images. In *Computational Intelligence for Measurement Systems and Applications, 2004. CIMSA. 2004 IEEE International Conference on*, pages 103–108. IEEE, 2004.

[10] S. Seo, S. O. Isikman, I. Sencan, O. Mudanyali, T.-W. Su, W. Bishara, A. Erlinger, and A. Ozcan. High-throughput lens-free blood analysis on a chip. *Analytical chemistry*, 82(11):4621–4627, 2010.

[11] N. Sinha and A. Ramakrishnan. Automation of differential blood count. In *TENCON 2003. Conference on Convergent Technologies for the Asia-Pacific Region*, volume 2, pages 547–551. IEEE, 2003.

[12] A. Szlam, K. Kavukcuoglu, and Y. LeCun. Convolutional matching pursuit and dictionary training. *arXiv preprint arXiv:1010.0422*, 4, 2010.

[13] L. Tao. *Learning Discriminative Feature Representations for Action Recognition and Segmentation*. PhD thesis, Johns Hopkins University, 2017.

[14] F. Yellin, B. Haeffele, and R. Vidal. Blood cell detection and counting in holographic lens-free imaging by convolutional sparse dictionary learning and coding. In *IEEE International Symposium on Biomedical Imaging*, pages 650–653, 2017.

[15] H. Zhu, S. Mavandadi, A. F. Coskun, O. Yaglidere, and A. Ozcan. Optofluidic fluorescent imaging cytometry on a cell phone. *Analytical chemistry*, 83(17):6641–6647, 2011.

[16] H. Zhu, I. Sencan, J. Wong, S. Dimitrov, D. Tseng, K. Nagashima, and A. Ozcan. Cost-effective and rapid blood analysis on a cell-phone. *Lab on a Chip*, 13(7):1282–1288, 2013.