Simultaneous Cell Detection and Classification in Bone Marrow Histology Images

Tzu-Hsi Song, Victor Sanchez, *Member, IEEE*, Hesham EIDaly, and Nasir M. Rajpoot, *Senior Member, IEEE*

***Abstract*—Recently, deep learning frameworks have been shown to be successful and efficient in processing digital histology images for various detection and classification tasks. Among these tasks, cell detection and classification are key steps in many computer-assisted diagnosis systems. Traditionally, cell detection and classification is performed as a sequence of two consecutive steps by using two separate deep learning networks, one for detection and the other for classification. This strategy inevitably increases the computational complexity of the training stage. In this paper, we propose a synchronized deep autoencoder network for simultaneous detection and classification of cells in bone marrow histology images. The proposed network uses a single architecture to detect the positions of cells and classify the detected cells, in parallel. It uses a curve-support Gaussian model to compute probability maps that allow detecting irregularly-shape cells precisely. Moreover, the network includes a novel neighborhood selection mechanism to boost the classification accuracy. We show that the performance of the proposed network is superior than traditional deep learning detection methods and very competitive compared to traditional deep learning classification networks. Runtime comparison also shows that our network requires less time to be trained.**

***Index Terms*—Digital pathology, Bone marrow trephine biopsy, Deep learning, Cell detection, Cell classification.**

I. Introduction

BONE marrow trephine biopsy is an important hematopathological specimen and is particularly useful in investigating specific bone marrow diseases, like myelofibrosis neoplasms (MPNs), which are neoplastic disorders of bone marrow hematopoietic stem cells. This biopsy facilitates a complete assessment of bone marrow architecture and the pattern of distribution of any abnormal infiltrate [1]. Moreover, analysis of bone marrow trephine biopsy also provides crucial information about cell morphology, which is useful for precise diagnosis. In bone marrow diseases, hematopoietic cells proliferate abnormally. Consequently, quantitative measurement of cells is an essential diagnostic tool for pathological measurements as it helps to distinguish different types of bone marrow diseases in routine clinical diagnosis. However, inter-observer variability among hematopathologists usually occurs because quantitative estimation is largely subjective and manual quantification is tedious, time intensive and expensive [2]–[5].

In order to address these challenges, nuclear/cell detection and classification methods are proposed as part of many automated histology image analysis techniques. Detection is useful to identify the local distribution of cells, while classification is useful to distinguish different cell types from the detection outcomes. Both detection and classification can help pathologists to find clinical assessment rules or models to produce an accurate diagnosis in an efficient manner and recommend efficient treatment strategies to patients. According to the guideline of the World Health Organization (WHO) diagnosis criteria of MPNs [3], [6], there is no standard and efficient quantitative clinical assessment or grading system using bone marrow hematopoietic stem cells to help the pathologists determine cellular distribution and precisely distinguish essential thrombocythemia (ET) and primary myelofibrosis (PMF), which are main subtypes of MPNs [2], [6], [7]. The main reason for this is the fact that the hematopoietic stem cells on bone marrow trephine histology images have various stain intensities, shape and texture features and are massively distributed over heterogeneous regions. Fig. 1 shows visual examples of bone marrow trephine histopathology images with Hematoxylin and Eosin (H&E) staining and the two major cell types of bone marrow hematopoietic stem cells, which are erythroid cells and myeloid cells. It can be seen that these cells tend to be massively and randomly distributed and exhibit similar morphological features. These factors increase the difficulty of not only manually identifying these types of cells but extracting cellular and tissue characteristics by using current image processing techniques.

In this paper, we propose a novel deep learning (DL) network to assist in the detection and classification of erythroid and myeloid cells in bone marrow trephine histopathology images. Specifically, we introduce a parallel DL network that can simultaneously perform detection and classification, thus reducing training time. Our network exploits the fact that the detection and classification tasks tend to use very similar DL networks, thus making it possible to perform these two tasks in parallel by a single DL network. To boost the detection and classification accuracy, our proposed network includes a novel neighborhood selection mechanism that takes into account the region surrounding the center of a detected cell.

The organization of this paper is as follows. A brief review of cell detection and classification methods is given in Section II. The motivation and rationale of the proposed DL network is discussed in Section III. Section IV details the architecture of our proposed network. A comparison of the quantitative performance of the proposed model and those of several conventional DL detection and classification frameworks is presented and discussed in Section V. Section VI concludes this paper.

T.H. Song, V. Sanchez and N.M. Rajpoot are with the Department of Computer Science, University of Warwick, UK (email: T-H.Song@warwick.ac.uk; V.F.Sanchez-Silva@warwick.ac.uk; Nasir.Rajpoot@ieee.org).

H. EIDaly is with Haematology Oncology Diagnostic Service, Addenbrookes Hospital, Cambridge, UK (email: hesham.eldaly@addenbrookes.nhs.uk.)

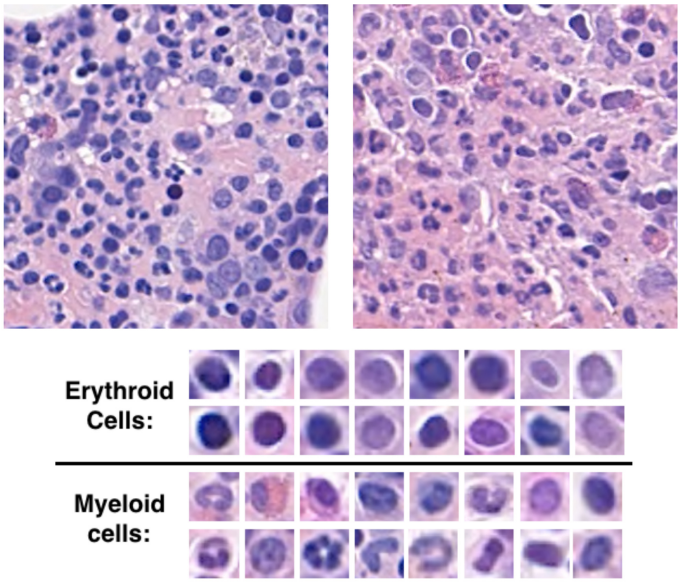


Fig. 1. Samples of bone marrow trephine histology images. The upper figures show sections of these images stained by H&E, where hematopoietic stem cells can be identified. The bottom figures show examples of the two major cell types: erythroid cells and myeloid cells. The myeloid cells include several subtypes of cells that are difficult to be visually identified by experts. Moreover, some of erythroid and myeloid cells have similar and rather fuzzy shapes or intensity features (last two columns of the bottom figure).

II. Related Work

Cell detection is the problem of determining the location of cells without accurately delineating their boundaries. This process is usually referred to as a marker or seed seeking and is performed around the cell center. The marker can be a single point or a small region inside the cell and provide support for nucleus/cell counting, tracking, and segmentation. In general, automatic detection in histology images is a very challenging task for two main reasons. First, there is a large amount of different cell types that are characterized by a large variety of morphological appearances, which are related to the high variation of biological structures, such as mitosis cells in breast cancer diagnosis. Second, cells usually overlap other cells, resulting in high number of false negatives in the detection process. In order to address these challenges, many automatic cell detection algorithms have been proposed based on nuclear or cellular hand-crafted features, such as local isotropic phase symmetry measure (LIPSyM) [8] and approaches employing convolutional neural networks (CNNs) and hand-crafted features [9]–[11].

Cell classification has been also applied on histology images for a wide range of diagnosis-related applications. In general, classification is often used as part of a segmentation or detection approach. Within this context, segmentation is useful to indicate the region of interest from which specific morphological or texture features are to be extracted to distinguish different types of cells [12], [13]. For example, Theera-Umpon *et al.* [14] use fuzzy-C mean clustering and a neural network to segment and classify the different series of myelocytes in bone marrow images. In [15], Sharma *et al.* propose a framework for nuclear segmentation and multi-cell classification using an AdaBoost classifier based on intensity, morphological and texture features in gastric cancer images. It is important to note that these approaches often require other pre-processing strategies to extract hand-crafted (morphological and intensity) features to achieve a high classification accuracy [16].

In recent years, DL-based approaches have been shown to be very successful to analyze histological images [17]–[19], including nuclear/cell detection and classification [4], [10], [19]–[21]. Generally, DL is a hierarchical learning approach that learns high-level features from pixel intensities; these high-level feature representations are then used to differentiate objects for classification purposes. DL approaches avoid the traditional design and computation of hand-crafted features and directly exploit the availability of unlabeled image data to capture high-level features. These methods directly extract multiple high-level features by using a feed-forward hierarchical structure with multiple layers [4], [17], [22], [23]. These layers compute features from the previous layer/feature representations and allow the network to gradually learn from low-level features to high-level features. This ability enables DL networks to handle high dimensional data for detection and classification. For instance, Xu *et al.* [4] use the stacked sparse autoencoder (SSAE) with a softmax classifier to learn a high-level representation of nuclear and non-nuclear objects to detect and classify nuclei. Cireşan *et al.* [10] use a CNN to detect mitosis in breast cancer histology images. Xie *et al.* [20] propose a structural regression CNN that can learn the proximity map of nuclei to provide more accurate detection results. Malon and Cosatto [24] train a CNN for mitotic and non-mitotic cell classification using color, texture and shape information. Sirinukunwattana *et al.* [19], [25] use a spatially constrained CNN (SC-CNN) with local spatial features to identify the position of nuclear centers. Despite the advantages of current DL methods used to detect and classify biological structures, detection and classification processes are usually performed by independent DL networks, which inevitably increases the training time.

In our previous works, the hybrid deep autoencoder (HDAE) network has been shown to be very efficient for nuclear/cell detection [21]. We have also previously designed a synchronized DL network to perform detection and classification simultaneously [26]. However, our previous network [26] can not efficiently reduce the intra-influence between detection and classification components to achieve the same performance as independent network process. In this paper, we propose an improved synchronized autoencoder (AE) model with a novel neighborhood selection mechanism to improve the performance of simultaneous detection and classification of erythroid and myeloid cells in bone marrow trephine histopathology images.

III. Motivation and Rationale

DL frameworks for simultaneous detection and classification in histology image analysis deal with the location and different types of nuclei/cells at the same time. Traditionally, these two problems are solved independently, thus requiring longer training time. To reduce these computational times,

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2018.2878945, IEEE Journal of Biomedical and Health Informatics

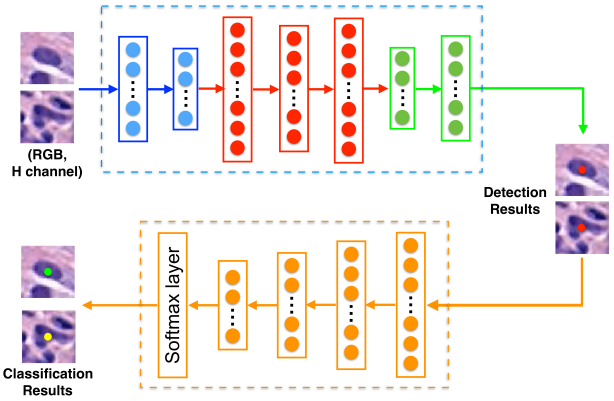


Fig. 2. Example of traditional DL framework for detection and classification. In general, there are two networks sequentially constructing for detection (upper half) such as HDAE network [21], and classification (bottom half) like CNN.

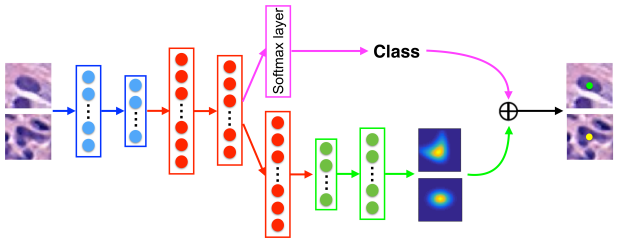


Fig. 3. Overall architecture of a synchronized DL framework for simultaneous cell detection and classification based on the HDAE network [21]. Cells are detected by employing the high-level feature representations computed from the input images and the corresponding probability maps (in green) via the connection network (in red). Also, the connection network works with the input network (in blue) and a softmax classifier (in pink) to classify different types of cells.

while keeping the same performance attained by the separate DL networks, our proposed framework combines two DL networks into one to synchronize the detection and classification mechanisms.

Let us consider a traditional DL framework for detection and classification comprising two separate DL networks. This traditional DL framework first detects objects and then classifies the detected objects. Fig. 2 shows an example of such conventional DL framework, where two networks are used to detect and classify bone marrow hematopoietic stem cells, respectively. For instance, we can use the HDAE network [21] to detect cells and a CNN to classify different cell types (Fig. 2). The classification mechanism can also use any machine learning approaches, such as a softmax classifier or Bag-of-Words dictionary learning [27], or DL methods. We note that if a conventional stacked AE network with a softmax classifier is used for classification, the first several layers of the stacked AE network and the first several layers of the HDAE network learn and process similar color and morphological features from the same data. Moreover, the architectures of these two networks are similar. Therefore, we can use the HDAE for detection and a stacked AE network for classification and merge them into a single parallel network to simultaneously detect and classify bone marrow hematopoietic stem cells. Specifically, one can merge the encoding mechanism based on the fact that the number of hidden layers and neurons used for encoding by these two networks is the same. Fig. 3 depicts the architecture of such a synchronized DL framework for cell detection and classification.

The particular DL framework in Fig. 3 includes a softmax classifier to perform cell classification because the high-level feature representations computed by AE structure also provide enough information to classify different types of cells. In other words, the connection network of this DL framework learns high-level feature representations that are useful for detection and classification. The detection is performed by employing these high-level representations and the corresponding probability maps. These maps are computed using a curve-support Gaussian model [28].

The synchronized DL framework in Fig. 3, however, may not attain accurately detection and classification results because the detection and classification mechanisms tend to affect each other. Specifically, the probability network concentrates on learning the features to compute the probability maps and does not take into account enough information acquired by the classification mechanism. When the loss function of classification mechanism is introduced in backpropagation step, it may act as noise and negatively influence the parameter optimization for detection mechanism. Similarly, the loss function of detection mechanism also affects the parameters of classification mechanism. In order to efficiently avoid and reduce the negative influence between detection and classification mechanisms, we propose a new DL network based on the synchronized DL framework in Fig. 3. We explain in detail our proposed DL network in the next section.

IV. SYN-ADHA: Synchronized Asymmetric Deep Hybrid Autoencoder

Our proposed DL network is based on the idea from Lee *et al.* [29], who proposed a concatenated asymmetric DL structure to efficiently take into account the images with distorted feature and noise influence. Let us recall that in an AE-based model, input data are compared with the reconstructed data to train the network's parameters via backpropagation, and this makes the network symmetric. If an AE-based model has one input and two different outputs (e.g., one for detection and one for classification), it is necessary to design an appropriate architecture to deal with the two results and optimize the network's parameters as much as possible. In [29], the authors design an asymmetric deep denoising network to estimate the noise and reconstruct clear features from distorted input features. Inspired by this idea, we propose a novel synchronized asymmetric DL network based on the HDAE network to perform detection and classification simultaneously.

It is important to emphasize that using similar network architectures for both detection and classification helps to reduce the overall computational complexity of the DL network. As discussed in the previous section, the stacked AE network for classification has a similar architecture to that of the HDAE network. We then adopt these two networks and design an appropriate connection network to integrate an input network,

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2018.2878945, IEEE Journal of Biomedical and Health Informatics

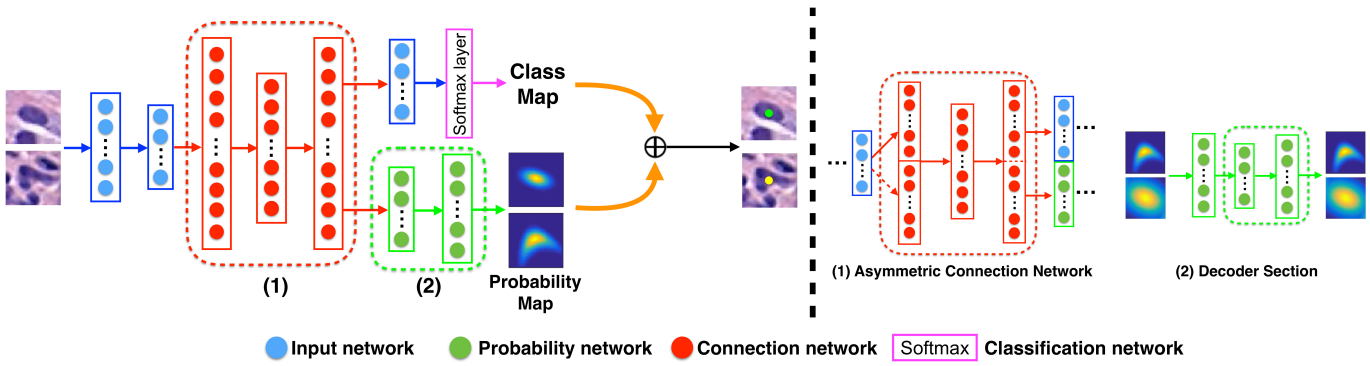


Fig. 4. (Left) Architecture of the proposed Syn-AHDA network for simultaneous detection and classification of cells. (Right) Details of the key components in the proposed network: (1) asymmetric connection network and (2) probability network. The asymmetric connection network uses one input vector and reconstructs two output vectors: the input vector used for classification and probability label vector used for detection. The decoding section of probability network is trained with the probability maps generated from curve-support Gaussian model.

a classification network, and a probability network, in order to perform detection and classification in parallel.

Fig. 4 shows the architecture of our proposed Synchronized Asymmetric Deep Hybrid AE network, which we turn on the Syn-AHDA network. There are four main components in the Syn-AHDA network: the input network $\mathcal{N}_I$, the classification network $\mathcal{N}_{Class}$, the probability network $\mathcal{N}_P$, and the asymmetric connection network $\mathcal{N}_{AC}$.

A. Input Network

The input network, $\mathcal{N}_I$, is constructed and trained in the same way as that of the HDAE network [21]. Formally, given a set of samples $X = [x_1, x_2, ..., x_N]$, the formulations of $\mathcal{N}_I$ are shown as:

$$\begin{cases} h_i^I = f\left(W_{l^I} x_i + b_{l^I}\right) \\ \hat{x}_i = f\left(W'_{l^I} h_i^I + b'_{l^I}\right) \end{cases} \tag{1}$$

where the hidden representation h^I is the intrinsic representation of $\mathcal{N}_I$; $f(\cdot)$ is the activation function that is set as a ReLU function, l^I represents the number of layer in $\mathcal{N}_I$. W and b are the weight matrix and bias terms in the encoding section, respectively, while W' and b' denote the same in the decoding section.

B. Classification Network

The classification network, $\mathcal{N}_{Class}$, is an AE-based architecture with a softmax layer and trained by using the high-level features from the input network $\mathcal{N}_I$. Given a set of class maps $C = c_1, c_2, \ldots, c_n$, the formulation of $\mathcal{N}_{Class}$ is:

$$\hat{c}_i = f\left(W_{smc} h_i^I\right), \tag{2}$$

where W_{smc} is the weight matrix of the softmax classifier and the hidden representation h^I is obtained from $\mathcal{N}_I$. We use the predicted class map, $\hat{c}_i$, and the labeled class map c_i to calculate the loss function to obtain the weight W_{smc}. It is worth noting that the detection mechanism includes more information that the classification mechanism does not need to use. In other words, the classification mechanism gets much noise from the detection mechanism. In order to tackle this problem, we use a sparsity term to train the softmax classifier in the classification mechanism.

C. Probability Network

The probability network, $\mathcal{N}_P$, is trained by the probability maps generated from curve-support Gaussian model, using the ground truth [see Fig. 4 (right)], to produce the probability maps needed for detection. The particular model helps to represent the shape and center of bone marrow stem cells.

1) Curve-Support Gaussian model: Gaussian model is widely employed for cell detection because the representation and properties of Gaussian distribution are quite similar to the morphological properties of round cells. For irregularly-shaped cells such as the myeloid cells, a Gaussian model cannot efficiently represent the corresponding morphological properties. Curve-support Gaussian model has been used to model the irregular shapes or ridges of biological structures such as dendritic trees and corneal nerve fibers [28]. This model takes into account rotation, scale and curvature to relax shape assumptions and accurately represents the target structures. To generate training labeled data, we use this model to compute the probability map of irregularly-shaped cells under the observation that the peak of a curve-support Gaussian corresponds to the center of a cell. From each input training patch, we can find the curve-support Gaussian model that best represents the shape of the cells; i.e., we compute the training probability maps as those in Fig. 5. The formulation of a curve-support Gaussian model is:

$$CG\left(\hat{x}, \hat{y}; \sigma, \theta, k\right) = \frac{1}{\sqrt{2\pi\sigma_x^2}} e^{-\frac{\hat{x}^2}{2\sigma_x^2}} \frac{1}{\sqrt{2\pi\sigma_y^2}} e^{-\frac{\left(\hat{y}+k\hat{x}^2\right)^2}{2\sigma_y^2}}, \tag{3}$$

$$\hat{x} = x cos\theta - y sin\theta, \quad \hat{y} = x sin\theta + y cos\theta, \tag{4}$$

where x, y denote the center position; σ is the scale of axis; θ and k control the orientation and the level of curve of the Gaussian, respectively. The first term of Eq. 4 controls the longitudinal Gaussian profile of the model, while the second term controls the cross-sectional Gaussian profile. When the parameter k increases, the curve of the Gaussian becomes steeper. Fig. 5 shows the process used to compute the training probability maps of three sample input training patches via curve-support Gaussian model with different levels

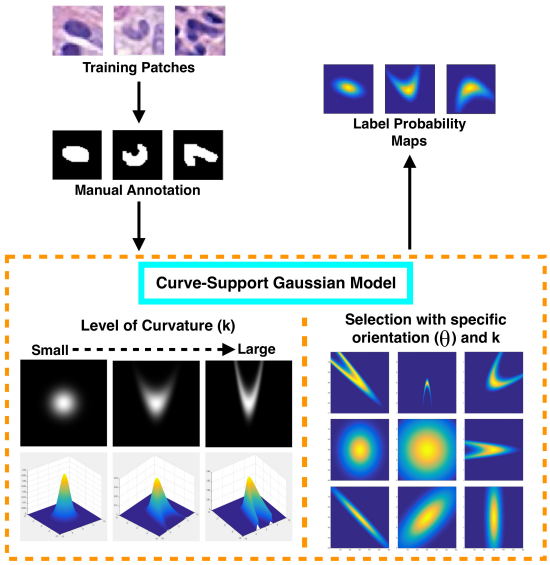


Fig. 5. Three example training patches and their corresponding label probability maps by using curve-support Gaussian model. Parameter k controls the level of curvature. Parameter θ controls the orientation of the curve-support Gaussian model. Each label probability map is generated by determining the parameters of the curve-support Gaussian model that best represent the shape and centroid of the training cell.

of curve, k, and various orientation, θ. The process estimates the parameters of the curve-support Gaussian model that best reconstruct the shape of each training cell. The shape of each training cell is provided as a binary mask computed from the ground truth.

2) Network formulation: We define a set of probability response maps $Y = [y_1, y_2, ..., y_N]$ as the input to train the probability network $\mathcal{N}_P$. The formulation of $\mathcal{N}_P$ is then as follows:

$$\begin{cases} h_i^P = f\left(W_{l^P} y_i + b_{l^P}\right) \\ \hat{y}_i = f\left(W'_{l^P} h_i^P + b'_{l^P}\right). \end{cases} \tag{5}$$

where h^P and l^P represent the hidden representation and the number of layers in $\mathcal{N}_P$, respectively.

D. Asymmetric Connection Network

The asymmetric connection network, $\mathcal{N}_{AC}$, integrates all other networks into a single architecture. According to (1) and (5), we obtain high-level feature representation, h^I, from the input images and the high-level feature representation, h^P, from the probability maps. In order for $\mathcal{N}_{AC}$ to integrate the input, probability and classification networks, two different AE sub-networks should be created to set the initial parameter of $\mathcal{N}_{AC}$. The first AE sub-network is used to obtain the weight matrix W_{AC_1} and W'_{AC_1} and bias term b_{AC_1} by using h^I, i.e., the input vectors for classification. The second AE sub-network uses h^I and h^P to compute the weight matrix W_{AC_2} and W'_{AC_2} and bias term b_{AC_2}. The second AE sub-network is a simple end-to-end structure and derived from the connect network of the HDAE model. These pre-training parameters obtained by above two sub-networks are integrated to initialize the proposed asymmetric connect network $\mathcal{N}_{AC}$. Specifically, we initialize weight matrix W_{AC_3} and bias terms b_{AC_3} of first hidden layer and weight matrix W'_{AC_3} of last hidden layer of $\mathcal{N}_{AC}$ as follows:

$$W_{AC_3} = \begin{bmatrix} W_{AC_1} & 0 \\ 0 & W_{AC_2} \end{bmatrix}, W'_{AC_3} = \begin{bmatrix} W'_{AC_1} & 0 \\ 0 & W'_{AC_2} \end{bmatrix}, \tag{6}$$

$$b_{AC_3} = \begin{bmatrix} b_{AC_1} & b_{AC_2} \end{bmatrix}^T. \tag{7}$$

We pre-train $\mathcal{N}_{AC}$ to optimize these parameters:

$$\begin{cases} h_i^{AC_3\_in} = f\left(W_{AC_3} h_i^I + b_{AC_3}\right) \\ h_i^{AC} = f\left(W_{AC} h_i^{AC_3\_in} + b_{AC}\right) \\ h_i^{\hat{AC}} = f\left(W'_{AC} h_i^{AC} + b'_{AC}\right) \\ h_i^{AC_3\_out} = f\left(W'_{AC_3} h_i^{\hat{AC}} + b'_{AC}\right), \end{cases} \tag{8}$$

where W_{AC} and b_{AC} are weight matrix and bias terms of intermediate hidden layer in $\mathcal{N}_{AC}$, and $h_i^{AC_3\_out}$ is the reconstructed feature representation of $h_i^{AC_3\_in}$. The structures of $\mathcal{N}_{AC}$ is shown in Fig. 4 (right). Then the loss function is calculated between $h_i^{AC_3\_out}$ and $\begin{bmatrix} h_i^I & h_i^{PC} \end{bmatrix}$ to optimize the parameters of $\mathcal{N}_{AC}$.

E. Training of the Syn-AHDA network

The training of the Syn-AHDA network is summarized in Algorithm 1. The parameters obtained by Algorithm 1 from input network $\mathcal{N}_I$, probability network $\mathcal{N}_P$, classification network $\mathcal{N}_{Class}$, and asymmetric connection network $\mathcal{N}_{AC}$ are used during the fine-tuning stage. The loss function of the Syn-AHDA network is formulated as follows:

$$Loss^{SynAHDA} = \sum_i \|z_i - \hat{z}_i\|^2 + \sum_i H\left(c_i, \hat{c}_i\right), \tag{9}$$

where $H\left(c_i, \hat{c}_i\right)$ is the cross-entropy cost function used to estimate the classification loss. This cost function is defined as follows:

$$H\left(c_i, \hat{c}_i\right) = -\left[c_i log\left(\hat{c}_i\right) - \left(1 - c_i\right) log\left(1 - \hat{c}_i\right)\right]. \tag{10}$$

F. Neighboring Class Selector

The proposed Syn-AHDA network produces probability maps and class maps. Therefore, it is necessary to find the local maximum points on the probability map, as these correspond to the cell centers. To detect the position of each cell center, similarly to the post-processing step of the HDAE network, a thresholding mechanism is used. The threshold, which is obtained and optimized by evaluating training images, removes those regions with low probability values of being the center and captures the local maximum points on the probability maps. To identify the class of each detected cell center, we consider the neighboring region of each detection on the class map. For each detected cell center $x_c^{(i)}$, we define a set of classes within a neighboring region with radius d_ρ as follows:

$$R\left(x_c^{(i)}\right) = \left\{x_c^{(i)} \in \wp_I : \|z\left(x_c^{(i)}\right) - z\left(x\right)\|_2 \leq d_\rho\right\}, \tag{11}$$

where x_c denotes the predicted centers of cells; $z\left(x\right)$ represent the location of point x; $\wp_I$ denotes the set of points in the

Algorithm 1 Syn-AHDA network training process

Input: X: Input data; Y: Probability maps; C: Class labels; $\mathcal{N}_I$: an input AE network; $\mathcal{N}_P$: a probability AE network; $\mathcal{N}_{Class}$: a classification network; $\mathcal{N}_{AC}$: an asymmetric connect network; W_{smc}: a softmax weight vector; W_{AC_3}, b_{AC_3} and $W^{'}_{AC_3}$: the weight and bias term of first hidden layer and the weight matrix of last hidden layer in $\mathcal{N}_{AC_3}$;

Output: The pre-trained Syn-AHDA network: $\mathcal{N}_{SynAHDA}$;

1: Use input data X to train $\mathcal{N}_I$ and extract the parameters in the encoding section of $\mathcal{N}_I$, $\mathcal{N}_{I_{encoding}}$;
2: Use probability maps Y to train $\mathcal{N}_P$ and extract the parameters in the decoding section of $\mathcal{N}_P$, $\mathcal{N}_{P_{decoding}}$;
3: Use X, the parameters of $\mathcal{N}_{I_{encoding}}$, and C to train $\mathcal{N}_{class}$ with sparsity term and then obtain W_{smc};
4: Use two sub-AE networks with the high-level features of $\mathcal{N}_{I_{encoding}}$ and $\mathcal{N}_{P_{decoding}}$ to generate W_{AC_3} and b_{AC_3} in the encoding section and $W^{'}_{AC_3}$ in the decoding section of $\mathcal{N}_{AC}$;
5: Pre-train the $\mathcal{N}_{AC}$ with W_{AC_3}, b_{AC_3} and $W^{'}_{AC_3}$ by using the output feature representation of $\mathcal{N}_{I_{encoding}}$ and integrated vector that combines the output feature representation of $\mathcal{N}_{I_{encoding}}$ and the input feature representation of $\mathcal{N}_{P_{decoding}}$ to obtain the initial fitting parameters of $\mathcal{N}_{AC}$;
6: Integrate the parameters of $\mathcal{N}_{I_{encoding}}$, $\mathcal{N}_{P_{decoding}}$, $\mathcal{N}_{AC}$ and W_{smc} into $\mathcal{N}_{SynAHDA}$, and then fine-tune this network $\mathcal{N}_{SynAHDA}$ to optimize the parameters by freezing the parameters of $\mathcal{N}_{I_{encoding}}$, $\mathcal{N}_{P_{decoding}}$, and W_{SMC};
7: **return** $\mathcal{N}_{SynAHDA}$

neighboring region on the class map of input image I. The class assigned to each detected center is:

$$C\big(x_c^{(i)}\big) = \underset{k}{argmax} \frac{\sum \big(R\big(x_c^{(i)}\big) = k\big)}{|R\big(x_c^{(i)}\big)|}, \tag{12}$$

where k is the class label and $|R\big(x_c^{(i)}\big)|$ denotes the number of elements in set $R\big(x_c^{(i)}\big)$. Note that the class selector in (12) determines which class has the highest probability to be assigned to the detected center and assigns the corresponding class label to the detected center. This class selector is necessary because the classification mechanism could be affected by the detection mechanism. To reduce the influence of the detection mechanism on the classification mechanism, we also consider the neighboring region around each detected cell center on the class map.Considering these neighboring regions for both, detection and classification, is one of the novel aspects of our network that helps to boost the accuracy.

To accurately identify the position and class of the cell centers, we limit the radius of the neighboring region used in the detection and classification mechanisms. In our experiments, we set the radius to 12 pixels for detection and $d_\rho = 2$ pixels for classification, which give the best performance by testing different values of radius parameter, so as to allow the neighboring region to cover the most of the area of the cell to be detected and classified.

V. Performance Evaluation

A. Data Collection

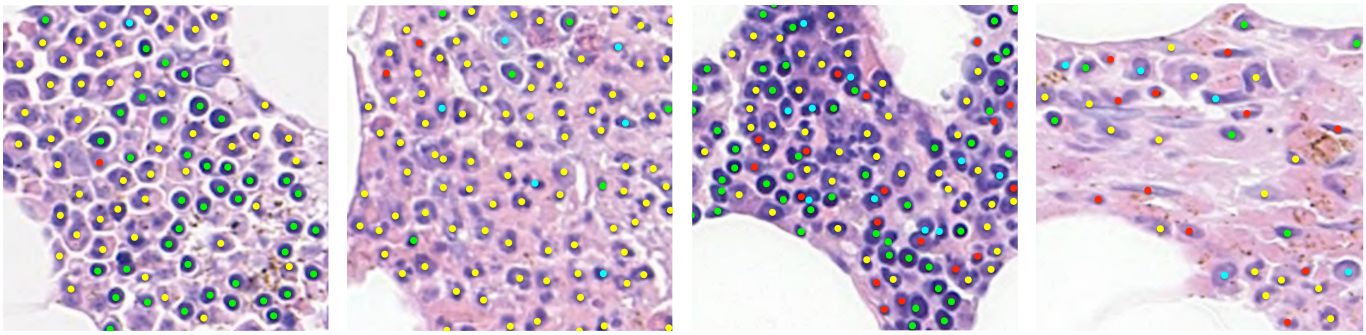
The experiment dataset includes a set of 52 H&E-stained histopathological bone marrow trephine biopsy images at 40× magnification of 5 ET and 5 prefibrotic PMF patient cases collected from University Hospitals Coventry and Warwickshire (UHCW) by using an Omnyx VL120 scanner. The annotation of 5,248 nuclear patches in our image dataset consists of 2,933 erythroid cells and 2,315 myeloid cells, which are marked by a trainee and verified by an experienced pathologist. We consider these two major types of bone marrow hematopoietic stem cells to evaluate the Syn-AHDA model because their unclear, complex, and heterogeneous cellular features make their visual identification a very complex task for pathologists. Here we use 2-fold validation to prevent over-fitting.

B. Network Setting and Training

We use the patch size of 29×29 pixels, which can efficiently capture the size of the cells to be identified at 40× magnification resolution. Each patch comprises four color channels; i.e., RGB and the Hematoxylin (H) channel. For the input network, $\mathcal{N}_I$, we set the size of input vector to $(29^2 \times 4) \times 1$, and two hidden layers with 1600 and 400 hidden units, respectively. We initialize the parameters of softmax classifier from $\mathcal{N}_{Class}$ for the classification mechanism. The probability network, $\mathcal{N}_P$, has $\big(29^2 \times 1\big)$ input units, which are generated by using curve-support Gaussian model. This network has one hidden layer with 400 hidden units. The asymmetric connect network, $\mathcal{N}_{AC}$, has 400 input units from the output of $\mathcal{N}_I$, 800 hidden units for the intermediate layer, and 800 output units, which correspond to the high-level feature representations of $\mathcal{N}_I$ and $\mathcal{N}_P$. According to Algorithm 1, the parameters of $\mathcal{N}_{AC}$ are obtained by using sub-AE networks with high-level feature representations of $\mathcal{N}_I$, $\mathcal{N}_{Class}$ and $\mathcal{N}_P$. Then we fine-tune the parameters of $\mathcal{N}_{AC}$ to optimize the parameters of Syn-AHDA network.

C. Evaluation results

Fig. 6 shows several examples of the results attained by the proposed Syn-AHDA network. These results demonstrate that our network can efficiently detect and classify erythroid and myeloid cells. In our results, a detected cell center is true positive if the detection and classification outcomes are correct according to the ground truth. We quantitatively compare our network's performance to that of other current conventional and DL approaches. First, we compare the detection performance of the Syn-AHDA network to that attained by LIPSyM [8], SSAE [4] and SC-CNN [19], HDAE network [21], and our previous Syn-AHDA network [26]. Table I lists these results in terms of precision, recall, and F1-score. These results shows that our proposed network achieves a similar performance to that attained by the HDAE network and a better performance than that attained by other methods. These results confirm that



(a) (b) (c) (d)

Fig. 6. Visual results for detection and classification of erythroid and myeloid cells via the Syn-AHDA network. Green and yellow dots represent the erythroid and myeloid cells, respectively. Red dots denote the incorrect detections and cyan dots are incorrect classifications.

the proposed network more efficiently reduces the influence from the classification mechanism in the detection mechanism and demonstrates its ability to detect the centroids of various-shaped cells.

For a fair comparison of the classification performance of the proposed Syn-AHDA network, we use the HDAE network for detection and different DL methods for classification. Specifically, we use HDAE with a sparse AE, HDAE with a SSAE network, and HDAE with a CNN. The number of hidden layers in the SSAE network is set to 6, which is the same number of hidden layers as in the proposed method. The CNN network comprises 3 convolutional layers, 3 max-pooling layers, one fully connected layer and one softmax layer. In this evaluation, we use the correctly detected cell centers according to the ground truth, to quantitatively measure the subsequent classification.

Table II lists the results of the evaluation in terms of precision, recall, and F1-score. It can be observed that our proposed network attains a high F1-score than that of other frameworks. In terms of precision, the proposed network also outperforms other DL frameworks because of the neighboring class selector. In terms of recall, our network outperforms all methods except for HDAE with SSAE. The main reason for these results is that although the detection mechanism of our network uses the location features to identify the position of the cells, the classification mechanism does not use these features to classify the cells. In other words, the location features may act as noise that affects the outcomes of the classification mechanism. Moreover, during training, since the classification mechanism does not need location features, the weight of these features tend to be reduced, which in turn affects the power of these features for detection. A similar situation occurs with the features used by the classification mechanism and not used by the detection mechanism. Although our proposed network with the neighboring selector is designed to avoid and reduce, as much as possible, this loss influence, the detection and classification mechanisms may still be negatively affected by each other. This mutual influence may also result in a set of parameters not fully optimized for either of the two mechanisms, but instead, a set of sub-optimal parameters that is most appropriate to be used by both mechanisms.

To sum up, the incorrect detection or classification results depicted in Fig. 6 occur because (1) the cell has faded stain color or the size of cell is larger than the size of the training patch; (2) the probability map is affected by the classification mechanism and the threshold is not appropriate to remove noisy points; (3) some myeloid cells have similar shape and intensity features to those of erythroid cells, making it thus difficult for the proposed network to correctly distinguish between these two types of cells; or (4) some of high-level features from the detection mechanism as noisy information affect the outcomes of classification. Overall, although the proposed Syn-AHDA network attain a better detection and classification performance than that of some conventional DL frameworks, there is room for improvement as part of our future research. For instance, because the detection and classification mechanisms affect each other, the high-level features of the detection or classification may be lost or become inaccurate during the training stage. One option is to add low-level features to high-level features to compensate for their inaccuracy.

TABLE I
DETECTION RESULTS BETWEEN THE PROPOSED MODEL AND OTHER METHODS

Methods	Precision	Recall	F1-score
LIPSyM [8]	0.7267	0.6514	0.687
SSAE [4]	0.8733	0.719	0.7887
SC-CNN [19]	**0.9517**	0.9118	0.9313
HDAE [21]	0.9273	**0.9702**	**0.9483**
Syn-AHDA	0.9266	0.9674	0.9466

TABLE II
CLASSIFICATION RESULTS BETWEEN THE PROPOSED MODEL AND OTHER DL FRAMEWORKS

Methods	Precision	Recall	F1-score
HDAE+Sparse AE	0.8267	0.8108	0.8187
HDAE+SSAE	0.8461	**0.9037**	0.874
HDAE+CNN	0.8605	0.8922	0.8761
Syn-AHDA	**0.8712**	0.8879	**0.8795**

D. Computational Times

We finish this section with some figures and remarks about the computational times for training the evaluated frameworks. We run experiments on MacOS version 10.13 with Intel CPU

TABLE III
TRAINING TIME OF VARIOUS DL FRAMEWORKS

Methods	Training Time
HDAE+Sparse AE	5-6 hours
HDAE+SSAE	7-8 hours
HDAE+CNN	6-7 hours
SC-CNN+CNN	5-6 hours
Syn-AHDA	**4-5** hours

i7 3.1GHz, 16G RAM and no GPUs. Table III tabulates the training times. Our proposed network requires the shortest amount of training time because it needs to train fewer parameters than other conventional framework due to its use of parallel architecture to implement simultaneous detection and classification. Removing the redundant learning layers and using the asymmetric connect network play a key role in the proposed network to reduce the number of fully connected layers that need to be trained.

VI. CONCLUSIONS

In this paper, we presented the Syn-AHDA network, a novel synchronized asymmetric hybrid deep autoencoder network to perform cell detection and classification simultaneously. The proposed network consists of an input network, a classification network, a probability network, and an asymmetric connection network. Our network exploits the fact that the detection and classification tasks tend to use similar high-level feature representations, thus making it possible to reduce the number of layers in the network. Our network also introduces a novel neighborhood selection mechanism that boosts the detection and classification accuracy by taking into account the region surrounding a detected cell center. Experimental results for the detection of erythroid and myeloid cells showed that the proposed Syn-AHDA network attains better detection and classification than both conventional and state-of-the-art deep learning frameworks, while requiring shorter training time. As part of our further research, we plan to extend the proposed network to perform different tasks, such as segmentation and classification, and use convolutional layer to improve the performance and training time.

REFERENCES

[1] B. J. Bain, D. M. Clark *et al.*, *Bone marrow pathology*, 4th ed. West Sussex, UK: Blackwell Science Ltd., 2010.

[2] B. S. Wilkins, W. N. Erber *et al.*, "Bone marrow pathology in essential thrombocythemia: interobserver reliability and utility for identifying disease subtypes," *Blood*, vol. 111, no. 1, pp. 60–70, 2008.

[3] J. Thiele, M. Imbert *et al.*, "Chronic idiopathic myelofibrosis. who classification of tumours: tumours of haematopoietic and lymphoid tissues," *IARC Press. Lyon*, pp. 35–38, 2001.

[4] J. Xu, L. Xiang *et al.*, "Stacked sparse autoencoder (ssae) for nuclei detection on breast cancer histopathology images," *IEEE Trans. on Medical Imaging*, vol. 35, no. 1, pp. 119–130, January 2016.

[5] H. Irshad, A. Veillard *et al.*, "Methods for nuclei detection, segmentation and classification in digital histopathology: A review. current status and future potential," *IEEE Review in Biomedical Engineering*, vol. 7, pp. 97–114, December 2014.

[6] J. W. Vardiman *et al.*, "The 2008 revision of the world health organization (WHO) classification of myeloid neoplasms and acute leukemia: rationale and important changes," *Blood*, vol. 114, no. 5, pp. 937–951, 2009.

[7] J. Thiele, H. M. Kvasnicka *et al.*, "Early stage idiopathic (primary) myelofibrosis - current issues of diagnostic features," *Leukemia & lymphoma*, vol. 43, no. 5, pp. 1035–1041, 2002.

[8] M. Kuse, M. Khan *et al.*, "Local isotropic phase symmetry measure for detection of beta cells and lymphocytes," *J. Pathol. Inform.*, vol. 2, no. 2, p. 2, 2011.

[9] H. Chen, Q. Dou *et al.*, "Mitosis detection in breast cancer histology images via deep cascaded networks." in *Thirtieth AAAI Conference on Artificial Intelligence*.

[10] D. C. Cireşan, A. Giusti *et al.*, "Mitosis detection in breast cancer histology images with deep neural networks," in *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 2013, pp. 411–418.

[11] M. N. Kashif *et al.*, "Handcrafted features with convolutional neural networks for detection of tumor cells in histology images," in *Biomedical Imaging (ISBI), 2016 IEEE 13th International Symposium on*. IEEE, 2016, pp. 1029–1032.

[12] A. Khadidos, V. Sanchez, and C. T. Li, "Path-based segmentation of overlapping cervical cells using active contour with local edge information," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2017, pp. 1058–1062.

[13] A. Khadidos, V. Sanchez, and C.-T. Li, "Weighted level set evolution based on local edge features for medical image segmentation," *IEEE Trans. on Image Processing*, vol. 26, no. 4, pp. 1979–1991, 2017.

[14] N. Theera-Umpon, "White blood cell segmentation and classification in microscopic bone marrow images," *Fuzzy systems and knowledge discovery*, pp. 485–485, 2005.

[15] H. Sharma, N. Zerbe *et al.*, "A multi-resolution approach for combining visual information using nuclei segmentation and classification in histopathological images." in *VISAPP (3)*, 2015, pp. 37–46.

[16] K. Nguyen, J. Bredno, and D. A. Knowles, "Using contextual information to classify nuclei in histology images," in *Biomedical Imaging (ISBI), 2015 IEEE 12th International Symposium on*. IEEE, 2015, pp. 995–998.

[17] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Networks*, vol. 61, pp. 85–117, January 2015.

[18] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[19] K. Sirinukunwattana, S. Raza *et al.*, "Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images," *IEEE Trans. on Medical Imaging*, vol. 35, no. 5, pp. 1196–1206, 2016.

[20] Y. Xie, F. Xing *et al.*, "Beyond classification: structured regression for robust cell detection using convolutional neural network," in *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 2015, pp. 358–365.

[21] T.-H. Song, V. Sanchez *et al.*, "Hybrid deep autoencoder with curvature gaussian for detection of various types of cells in bone marrow trephine biopsy images," in *Biomedical Imaging (ISBI 2017), 2017 IEEE 14th International Symposium on*. IEEE, 2017, pp. 1040–1043.

[22] P. Baldi, "Autoencoders, unsupervised learning, and deep architectures," *ICML unsupervised and transfer learning*, vol. 27, no. 37–50, p. 1, 2012.

[23] N. Bayramoglu and J. Heikkilä, "Transfer learning for cell nuclei classification in histopathology images," in *Computer Vision–ECCV 2016 Workshops*. Springer, 2016, pp. 532–539.

[24] C. D. Malon and E. Cosatto, "Classification of mitotic figures with convolutional neural networks and seeded blob features," *Journal of Pathology Informatics*, vol. 4, no. 1, p. 9, 2013.

[25] K. Sirinukunwattana, S. Raza *et al.*, "A spatially constrained deep learning framework for detection of epithelial tumor nuclei in cancer histology images," in *International Workshop on Patch-based Techniques in Medical Imaging*. Springer, 2015, pp. 154–162.

[26] T.-H. Song, V. Sanchez *et al.*, "Simultaneous cell detection and classification with an asymmetric deep autoencoder in bone marrow histology images," in *Medical Image Understanding and Analysis (MIUA)*, 2017, pp. 829–838.

[27] S. Ensafi, S. Lu *et al.*, "A bag of words based approach for classification of hep-2 cell images," in *Pattern Recognition Techniques for Indirect Immunofluorescence Images (I3A), 2014 1st Workshop on*. IEEE, 2014, pp. 29–32.

[28] R. Annunziata, A. Kheirkhah *et al.*, "Scale and curvature invariant ridge detector for tortuous and fragmented structures," *Medical Image Computing and Computer Assisted Inerventions (MICCAI)*, pp. 588–595, October 2015.

[29] K. H. Lee, S. J. Kang *et al.*, "Two-stage noise aware training using asymmetric deep denoising autoencoder," *Acoustics, Speech and Signal Processing (ICASSP), 2016 IEEE International Conference on*, pp. 5765–5769, 2016.