EBI02077 Assignment Report

Yucheng Lu

April 25, 2023

1 Introduction

The assignment dataset contains 168 images of human HT29 colon cancer cells exhibiting 14 dramatically different morphologies. These images come from 14 samples in total, each sample provides 4 patches (*i.e.* field of view) of size 512×512. Three staining materials, namely Hoechst 33342, pH3, and phalloidin, were applied to obtain three channels labeling DNA, mitosis, and actin. The task is to train a neural network and a Gaussian mixture model (GMM) to retrieve and quantitatively characterize these morphologies.

2 Methodology

The method agreed upon by the panel and the candidate follows the design of Deep InfoMax [HFLM+18], an overview is shown in Fig. 1. This model acts as an encoder that extracts high-level features from the input image via mutual information (MI) maximization, which we refer to as *representations*. The extracted representations can be used for downstream tasks such as classification. For more information regarding the pipeline, see [HFLM+18].

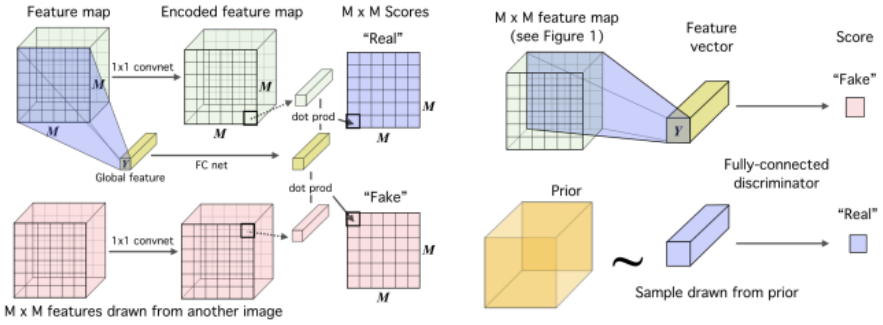

Figure 1: Overview of the unsupervised learning protocol.

2.1 Data preparation

We observed that the images in the dataset contain lots of background pixels that are not useful in discovering morphologies, which in our experiments brought the training into an undesired direction where the model learned to characterize non-morphology features. To tackle this issue, we applied binary classification to each image to identify meaningful regions by training a small bi-class GMM. Fig. 2 shows an example, note that this mask was employed for training purposes only.

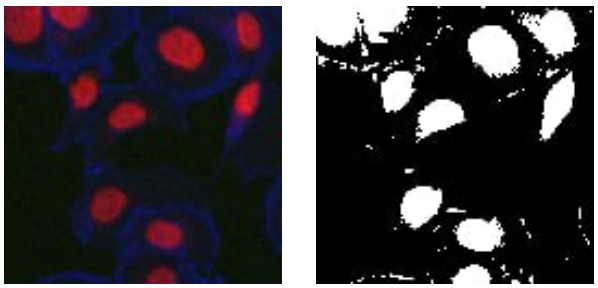

Figure 2: Example of an image (left) and its coarse binary mask (right).

2.2 Training details

We divided the sorted dataset into a training set and a testing set, and the number of samples was set to 10 and 4 (70:30), respectively. At every iteration, we loaded as input a 4D tensor of size $B \times N \times C \times H \times W$ from the dataset, each dimension in the tensor represents the batch size, fields of view, image channels, image height, and image width, respectively.

Different from [HFLM+18], the positive and negative pairs used to compute MI loss were reconstructed as follows:

1. To encourage the model to predict similar representations for the same morphology, all local representations and global representations from the *same sample*, no matter which fields of view they come from, were used to make positive pairs.
2. For negative MI loss, we only considered any paired field of views from *two different samples* as negative pairs.

Besides, when computing total MI loss, local representations from the background regions were ignored. This was achieved by identifying their classes from the pooled binary mask obtained in Section 2.1.

We designed two shallow networks to extract local and global representations and adopted the MI discriminator and the prior discriminator from [HFLM+18] in network training. More details can be found on Github.

2.3 Testing details

The well-trained models were used to extract representations of the testing set. As the image size in testing was different from that in training, we used a sliding window to run the inference on the input multiple times to obtain a collection of representations. A merit of doing so is that more prediction samples will help improve the robustness of GMM clustering. We then performed dimension reduction [MHM18] to these high-dimension representations and run GMM clustering to assign labels for all testing images.

3 Results and discussions

We provide the visualization of the dimension-reduced results of the testing set in Fig. 3(a). It can be seen that all 4 samples were distantly separated by the model, meaning that it could efficiently discover the morphologies of cells even if they were absent during training. The trained GMM successfully grouped all testing samples by their morphologies.

Fig. 3(b) shows the dimension-reduced results of the training set. Even though most of the morphology types were well-separated, there were two morphologies (0933x and 1357x) with overlapped representations that can't be easily distinguished. This might be caused by two reasons: First, the batch-norm layers in the model changed the distribution of features, making the model insensitive to signal intensity which could be one of the differences between the two morphologies; Second, the multiple max-pooling layers in the model heavily downsampled the feature map size and only transferred the most significant response to the following layers, which omitted the density of instances that might be also crucial in disguising the two classes.

Based on the above analysis, future work might take the intensity of feature response as well as the density of cells within a fixed-size valid region into account to help improve the quality of the extracted representations.

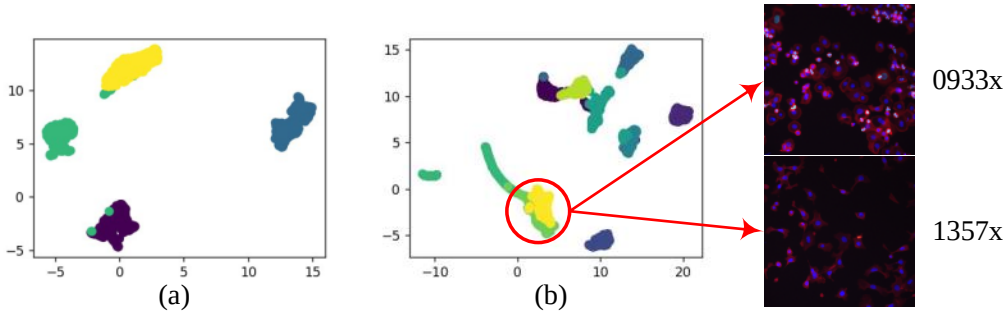

Figure 3: Visualization of the dimension-reduced results. (a) testing dataset; (b) training dataset.

References

[HFLM+18] R Devon Hjelm, Alex Fedorov, Samuel Lavoie-Marchildon, Karan Grewal, Phil Bachman, Adam Trischler, and Yoshua Bengio. Learning deep representations by mutual information estimation and maximization. *arXiv preprint arXiv:1808.06670*, 2018.

[MHM18] Leland McInnes, John Healy, and James Melville. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*, 2018.