Mixture Density Networks for Existence-Aware Nuclei Detection

1 Introduction

Quantitative characterization of histology images is crucial for clinical and research applications. Despite advances in methods developed for analyzing these images, they mostly rely on the existence of large data. We address this by proposing simple yet effective modeling of the problem, where a less number of parameters need optimization, enabling the model to deal with small data.

Mixture Density Networks (MDNs) [1] provide a framework with both mixing coefficients and conditional probabilities conditioned on the input, and can also perform data augmentation by their generative nature. We use MDNs to explicitly model the distribution of high level features extracted by a plugged in deep neural network to address the cell detection problem. This problem is conventionally treated by classifying image pixels through networks which provide limited control on the complexity of their underlying models. We use synthetic images to show that the proposed method retains its performance with reduced data size and also use samples of two real image datasets to illustrate it can achieve baseline performance (Fig. 1).

2 Method

For each image $\mathbf{x}_i$ in the set $\mathbf{X} = \{\mathbf{x}_i\}$, we want to find the set of coordinates $\mathbf{t}_i$ that contains the locations of nuclei in $\mathbf{x}_i$. Koohbanani *et al.* have used MDNs for cell detection [2], however they have modeled the existence of nuclei in the image through adding a heuristic term in the loss function. Here, we use the Bernoulli distribution to generate the joint probability distribution of the model. The Bernoulli variable $e(\mathbf{x}_i)$ (denoted by e_i for simplicity) determines the *existence* of a nucleus in $\mathbf{x}_i$. Thus, given the image $\mathbf{x}_i$, the joint probability is written as $p(\mathbf{t}_i, e_i|\mathbf{x}_i) = p(\mathbf{t}_i|e_i, \mathbf{x}_i)p(e_i|\mathbf{x}_i)$. e_i is conditioned on the input image through $p(e_i|\mathbf{x}_i) = \gamma^{e_i}(e_i - \gamma)^{1-e_i}$, where γ is the mean of the Bernoulli distribution.

Considering a mixture of Gaussian functions for the image, the joint probability distribution for the n^{th} nucleus in image $\mathbf{x}_i$ is written as follows.

$$p(\mathbf{t}_n i, e_i|\mathbf{x}_i) = (\gamma \sum_{k=1}^{K} \pi_k(\mathbf{x}_i)\mathcal{N}(\mathbf{t}_{ni}|\mu_k(\mathbf{x}_i), \sigma_k^2(\mathbf{x}_i)\mathbf{I}))^{e_i}(\frac{1-\gamma}{s_i})^{1-e_i}. \quad (1)$$

In Eq. 1, s_i denotes the image area. If the image does not contain any nuclei, the Bernoulli distribution will be zero, $e_i = 0$. This results in $p(\mathbf{t}_n i, e_i|\mathbf{x}_i) = \frac{1-\gamma(\mathbf{x}_i)}{s_i}$.

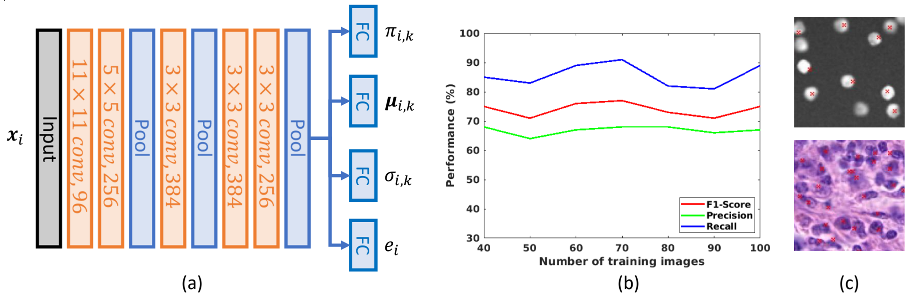

Fig. 1. Panel (a) shows the network structure. Panel (b) shows the performance of the algorithm on the synthetic image set when different training sizes were chosen. Panel (c) shows sample images from the two real image datasets [3, 4].

Rewriting the joint probability for N_i nuclei in the image $\mathbf{x}_i$ and deriving the negative log likelihood generates the following loss function.

$$\begin{aligned} E(\mathbf{w}) = -\sum_{i=1}^{N} e_i \sum_{n=1}^{N_i} \ln \sum_{k=1}^{K} \pi_k(\mathbf{x}_i, \mathbf{w}) \mathcal{N}(\mathbf{t}_{ni} | \mu_k(\mathbf{x}_i, \mathbf{w}), \sigma_k^2(\mathbf{x}_i, \mathbf{w})\mathbf{I}) \\ - \sum_{i=1}^{I} N_i e_i \ln \gamma - \sum_{i=1}^{I} N_i (1 - e_i) \ln \frac{1-\gamma}{s_i}. \end{aligned} \tag{2}$$

3 Results and Discussion

Fig. 1 shows the network structure and results of the method on synthetic and sample real images. At this point we have not made a quantitative comparison to the baseline methods but showed the method's performance on sample real images and its robustness to the training size on the synthetic images.

We derived a new mathematical modeling for using MDNs with pre-trained deep neural nets to address limited data availability for detection. Using Bernoulli distribution, the input data can include images without nuclei in them. We used AlexNet [5] for simplicity, while there is no constraint on choosing the pre-trained network. This work can be extended to perform a joint classification-detection algorithm using categorical distribution for classification prior to detection. MDNs can be used with arbitrary probability distributions and can also incorporate priors through Bayesian MDNs.

References

1. C M Bishop. Mixture density networks. 1994.
2. N A Koohababni and *et al.* Nuclei detection using mixture density networks. In *MLMI*, pages 241–248. Springer, 2018.
3. K Sirinukunwattana and *et al.* Locality sensitive deep learning for nuclei detection and classification in colon cancer histology. *IEEE TMI*, 35(5):1196–1206, 2016.
4. A Lehmussola and *et al.* Synthetic images of high-throughput microscopy for validation of image analysis methods. *Proceedings of the IEEE*, 96(8):1348–1360, 2008.
5. F N Iandola and *et al.* Squeezenet: Alexnet-level accuracy with 50x fewer parameters and 0.5 mb model size. *arXiv preprint arXiv:1602.07360*, 2016.