Deep Metric Learning Based Histopathological Image Classification

Gang Zhang¹ and Hui Xu²*

¹ School of Automation, Guangdong University of Technology, Guangzhou 510006, China
² School of Economic and Management, Shanghai Maritime University, Shanghai 201306, China
Email: huixu@shmtu.edu.cn

Abstract. With the rapid development of deep learning, histopathological image classification models have made remarkable progress. Recent deep learning-based methods have been applied to raw histopathological images to construct end-to-end models, which avoid hand-craft feature engineering. To construct a model that can capture the intrinsic pattern of the histopathological image dataset, we design a model based on deep metric learning which embeds data points into a Euclidean space. The proposed model trains a deep neural network, which embeds an input image into a Euclidean space where dissimilar images are located far away to each other and vice versa. We adopt a BN-Inception network pretrained on ImageNet as the embedding model. Then it is retrained on target datasets with some triplet loss function. A weighted distance-based triplet sampling strategy is designed to generate hard triplets for the training procedure. Evaluations on benchmark datasets indicate that our deep metric learning-based method outperforms recent successful deep learning models.

1. Introduction
Histopathological image (HI) classification is significant in computer-aid diagnosis (CAD) system [1]. In recent years, methods and models of HI analysis have made remarkable progress [2-3]. A large body of study of HI classification, recognition, grading and segmentation focus on constructing end-to-end models, in which deep learning methods are applied to raw image data and the model outputs the classification labels. End-to-end deep models avoid hand-craft feature engineering. However, end-to-end models cannot be directly applied to applications other than classification, such as unsupervised clustering, ranking, image retrieval, outlier analysis, etc. To construct a model that can represent the intrinsic pattern of HI datasets, we propose a deep metric learning (DML)-based model which provides an embedding function, mapping an input HI as a point of Euclidean space. In such Euclidean space, the points representing images from the same category are close to each other and vice versa. We construct a deep model to map input images to points in some feature space. The main idea of our method is sketched in figure 1.

In figure 1, there are three images belonging to two categories, namely connective and muscular. The images are fed into a deep CNN which encode them as points. After mapping, the Euclidean distance between blue points (black line) are shorter than that between blue and red point (green line). Machine learning tasks can be easily accomplished in the target space.
The network is optimized by gradient descendent and loss backpropagation, like most deep network training. We adopt triplet loss as the optimization target. The triplet consists of two positive and a negative instance, denoted as \((a, p, n)\), in which \(a\) and \(p\) are belonging to same class while \(a\) is belonging to some different class. We call \((a, p)\) positive pair (PP) and \((a, n)\) negative pair (NP). Metric learning with triple loss aims to learn a metric that ensures the distance of each PP is smaller than that of NP.

The training of the model relies on the positive and negative pairs of training histopathological images. It is infeasible to train with the whole set of triplets, whose order is cubic in the number of images. Hence a sampling strategy should be designed, aiming to pick out the most difficult triplets, i.e. the triplets having more information than the others. Wu et al. [4] stated that sampling strategies are as significant as loss functions when training models. They proved that randomly sampling strategy contributes little to the embedding function.

The remainder of our paper is structured as follows. Section 2 reviews the work closely related to this study. Section 3 presents the proposed DML method, including loss function definition, sampling method and deep CNN design. Section 4 reports the evaluation results of the proposed method. Finally, we conclude the paper in section 5.

2. Literature Review

DML has been widely studied in general image classification, face verification, image clustering and recognition, etc. Schroff et al. [5] proposed deep embedding model for face verification. They defined triplet loss to ensure that the face images of the same person are located closely in a feature space and vice versa. And an online triplet selection strategy is designed to select hard positive and hard negative pairs for fast convergence. Song et al. [6] proposed an embedding method which lifted the vector of pair-wise distance to the matrix of pair-wise distance which provided a better pair sampling strategy to avoid local optima. Wang et al. [7] proposed a DML framework with an angular loss function. Different from contractive loss and traditional triplet loss, the angular loss has scale invariance. Angular loss function can capture additional local structure by introducing a third-order geometric constraint. Qian et al. [8] proposed SoftTriple loss for deep metric learning. The replaced the max operator in the output layer to softmax operator which evaluating the distances from the input sample to the predefined multiple centers. Their method can avoid triplets sampling in a batch. Recent progress and related problems of deep metric learning can be found in Refs. [9, 10].

3. Deep Metric Learning

We give formal definition of the problem of DML. Let \(D = (x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\) be a set of samples, in which \(x_i \in R^d\) is the feature vector and \(y_i\) is the concept label of the sample. The learning goal of DML is to learn a mapping \(f : R^d \rightarrow R^m\). Under \(f\), similar samples (have the same concept labels) are closer to each other and vice versa. In traditional metric learning, the function \(f\) is often parametrized by a matrix and the embedding can be regarded as a Mahalanobis distance, as shown in equations (1) and (2).
Linear transformation limits the representation ability of the learned mapping. Deep metric learning adopts deep convolutional neural network as the mapping function $f$. To train, a triplet $(x_a, x_p, x_n)$ of training images is fed into the network and the outputs are $(f(x_a), f(x_p), f(x_n))$. In the triplet, $x_a$ and $x_p$ are similar while $x_a$ and $x_n$ are dissimilar. The loss function evaluates the distance between $||x_a - x_p||^2_2$ and $||x_a - x_p||^2_2$. And then an optimizer calculates the gradients and tunes the parameters of the layers backward. We pose a L2 normalization on the network output to make $||f(x_i)||_2 = 1$. This means that the embedding points spread on a hyper unit sphere. Figure 2 shows the main framework of our DML framework.

![Figure 2. The proposed DML framework.](image)

A deep CNN model, a loss function and a sampling strategy are three main parts in deep metric learning. Recent study [10] shows that these three have prominent impact on metric learning. A good deep CNN model can provide excellent representation and generalization ability. The network model is trained with a loss function for optimal feature expression to distinguish different classes of objects. Due to the use of sample pairs for training, selecting the most informative sample pairs for training can make the model converge faster. The following subsections elaborate on these three aspects.

3.1. Deep CNN

We choose the model with excellent performance in ImageNet image recognition competition in recent years as the embedded learning model. The architecture of the model follows ResNet and DenseNet. In ResNet model, gradients can reach any previous layer by fast connections, so that the problem of gradient disappearance is avoided. DenseNet proposes a more radical dense-join mechanism: all layers are interconnected, i.e. each layer accepts all the previous layers as its additional input. There are two types of basic blocks in ResNet-152. The first is identity block, whose input and output dimensions are identical so they can be concatenated. The second is convolutional block, whose input and output dimensions are different. Figure 3 shows the core blocks of both models.

![Figure 3. Residual block and dense block.](image)
By adjusting the depth of the model, the parameter size of the two models is kept at a similar level. We add an L2 regularization after the last fully connected layer so that its L2-norm is equal to 1. This causes the points encoded by the model to be distributed on a unit hypersphere. Limited by computing resources, we retrain the models trained in ImageNet.

3.2. Triplet Loss Function
The advantage of triplet loss function over the contractive loss is that it can tolerate some inner class variance. The main idea of triplet loss is that after embedding, the distance between two samples of the same kind should be greater than the distance between two samples of different kinds. The model only needs to pay attention to this difference in value size, and does not need to care about the real distance value. Equation (3) shows the general form of triplet loss function:

$$L_{\text{triplet}} = \max \left( 0, ||h_w(x) - h_w(x^p)||_2 - ||h_w(x) - h_w(x^n)||_2 + \alpha \right)$$

(3)

In equation (3), \(h_w(\cdot)\) stands for the embedding function parameterized by \(w\). \(x, x^p\) and \(x^n\) stand for the anchor, positive and negative samples. And \(\alpha\) is a margin value. When optimizing the model, the parameter \(w\) is adjusted with a gradient descent process in order to minimize the triplet loss on the training dataset. Recent work shows that the triplet loss can be relaxed by replacing the max function with a softmax function [8]. However, our triplet loss function is readily comprehensible and easily to be optimized. The margin value \(\alpha\) is a hyperparameter when optimizing the model. Figure 4 shows the main idea of triplet loss function.

![Figure 4. Triplet loss function.](image)

The value of \(\alpha\) will vary for different training data sets. We propose an adaptive strategy to set the value of \(\alpha\) based on the samples in each batch, as shown in equation (4).

$$\alpha = \frac{d_{\text{max}}}{d_{\text{max}} + d_{\text{min}}} ||C^p - C^n||$$

(4)

In equation (4), \(d_{\text{max}}\) and \(d_{\text{min}}\) mean the max and min distances among all samples in a batch. \(C^p\) and \(C^n\) stand for the embedded centroids of negative and positive classes.

3.3. Sampling Strategy
An important problem of our DML model is to build the triplet training dataset. Recent research [4] indicated that hard triplets can train the model more effectively. For a hard triplet, the distance between \(x\) and \(x^p\) is large and the distance between \(x\) and \(x^n\) is small. However, it is expensive to pick out hardest triplets from the whole training dataset. Each embedding point is located on a unit hypersphere. The distribution of the distance between an arbitrary pair of samples approaches \(N(\sqrt{2}, 1/2d)\) when the dimension \(d\) becomes large [11]. Hence in a probabilistic sense, random sampling cannot generate hard triplets. We adopt a distance weighted sampling strategy to generate triplets. Given a training image dataset, we sample according to the distance of a pair, as shown in equation (5):

$$p(n|p) \propto min(\beta, d^{-1}_n)$$

(5)
\[ d_{np} = | |h(n) - h(p)| | \] stands for the distance of the pair in the embedding space. We set the batch size to a relatively large value and confine the sampling process in a batch to avoid combination explosion.

4. Evaluations
We evaluate our method, denoted as DMLC, on real histopathological image datasets compared with recent proposed state-of-the-art methods. The first dataset is BreakHis (Breast Cancer Histopathological Database) [12], in which there are images and pathological diagnoses for 82 patients. Totally there are 4872 images of breast tumour tissue, in which 2412 are benign and 2460 are malignant. The image size is 700*460 at 200X magnification. All images are in PNG format. The second dataset is HISTissue Dataset, consisting of four types of basic tissue images. Table 1 shows the distribution of these four types.

| Tissue  | Images |
|---------|--------|
| nervous | 1026   |
| connective | 484   |
| epithelial | 804   |
| muscular | 514    |

We write the evaluation code using python 3 and use the Google Tensorflow 2.x to implement the deep network. The evaluation code runs on a workstation with Intel i7 central process unit, 32GB random access memory, NVIDIA 1080ti 11GB graphics accelerator. The methods for comparison were proposed in Refs. [6, 8], denoted as M1 and M2 respectively. The motivation of choosing these two methods is twofold. On the one hand, these two works adopt the same network architectures as this one. On the other hand, M1 adopts an end-to-end deep network and outputs the classification results through a softmax layer placed behind the last fully connected layer. M2 performs DML with a soft triplet loss function without sampling strategy when building training triplets. M1 and M2 have good performance on their evaluation datasets.

We use 10-fold cross-validation to evaluate the methods. In order to get stable results, we repeat the random division 10 times and record the mean and standard deviation of the evaluation criteria. Both datasets are single-label, i.e. each sample attaches to only one concept label. Hence, we adopt ACC, Recall and F1 score as evaluation criteria. See equations (6)-(8) for their definitions.

\[
ACC = \frac{TP}{TP+TN+FP+FN} \tag{6}
\]
\[
Recall = \frac{TP}{TP+FN} \quad Precision = \frac{TP}{TP+FP} \tag{7}
\]
\[
F_1 = \frac{2}{1/Recall+1/Precision} \tag{8}
\]

\( TP, FP \) and \( FN \) stand for the numbers true positive, false positive and false negative images. The deep networks for metric learning are ResNet-152 [13] and DenseNet-BC [14]. For DMLC and M2, either of them outputs an optimal embedding for an input \( x \). A gradient boosting decision tree (GBDT) is adopted to obtain the final classification results. Tables 2 and 3 show the results of both datasets.

We highlight the lines on which the model performs best in both tables. It can be seen that the proposed DMLC model achieves the best performance. The deep metric learning based models with GBDT classifier perform better than pure deep neural network image classification model.
Table 2. The evaluation results of BreakHis.

| Model               | ACC       | Recall    | Precision | \(F_1\)    |
|---------------------|-----------|-----------|-----------|------------|
| M1                  | 82.5%±2.5%| 86.2%±2.7%| 81.0%±2.4%| 0.418±0.009|
| M2 ResNet-152       | 88.1%±2.8%| 84.6%±3.4%| 89.1%±2.9%| 0.434±0.010|
| M2 DenseNet-BC      | 87.2%±3.1%| 90.2%±2.0%| 85.1%±2.1%| 0.438±0.012|
| DMLC ResNet-152     | 89.6%±2.2%| 87.1%±1.9%| 92.0%±2.4%| 0.447±0.012|
| DMLC DenseNet-BC    | 90.2%±2.7%| 90.1%±2.3%| 90.8%±2.2%| 0.452±0.011|

Table 3. The evaluation results of HISTissue.

| Model               | ACC       | Recall    | Precision | \(F_1\)    |
|---------------------|-----------|-----------|-----------|------------|
| M1                  | 89.7%±3.4%| 86.1%±2.6%| 90.7%±2.9%| 0.442±0.015|
| M2 ResNet-152       | 92.1%±2.9%| 94.1%±2.8%| 90.6%±3.2%| 0.462±0.017|
| M2 DenseNet-BC      | 93.0%±2.6%| 95.6%±2.9%| 92.4%±3.6%| 0.470±0.013|
| DMLC ResNet-152     | 94.9%±1.9%| 95.8%±2.2%| 93.7%±2.4%| 0.474±0.013|
| DMLC DenseNet-BC    | 94.7%±2.0%| 91.2%±3.1%| 95.6%±2.5%| 0.467±0.021|

5. Conclusion
In this paper we proposed a deep metric learning for histopathological image classification. We describe the deep CNN, triplet loss function and sampling strategy in detail. Evaluation results on benchmark datasets show that the proposed model outperforms the recent successful models. The output embedding of deep metric learning model can easily be applied to other machine learning tasks. Future work includes designing effective sampling strategy, improving the loss function for fast convergence and optimizing deep CNN models.

Acknowledgements
This work is supported by the National Natural Science Foundation of China (No. 71702100, 81373883, 81573827), the Science and Technology Planning Project of Guangdong Province (No. 2016A030310340) and the College Student Career and Innovation Training Plan Project of Guangdong Province (2019111845125, xj2019111845084, xj2019111845128).

References
[1] Gurcan M N, Boucheron L E, Can A, Madabhushi A and Yener B 2009 Histopathological image analysis: A review IEEE Reviews in Biomedical Engineering 2 147-171.
[2] Xie J, Liu R, Luttrel I V J and Zhang C 2019 Deep learning based analysis of histopathological images of breast cancer Frontiers in Genetics 10 80.
[3] Korbar B, Olofson A M, Mirafior A P, Nicka C M, Suriawinata M A, Torresani L, Suriawinata A A and Hassanpour S 2017 Deep learning for classification of colorectal polyps on whole-slide images Journal of Pathology Informatics 8 abs/1703.01550.
[4] Wu C Y, Mannathar R, Smola A J and Krahenbuhl P 2017 Sampling matters in deep embedding learning Proceedings of the IEEE International Conference on Computer Vision (Venice: Italy) pp 2840-2848.
[5] Schroff F, Kalenichenko D and Philbin J 2015 Facenet: A unified embedding for face recognition and clustering Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (Boston: MA, USA) pp 815-823.
[6] Song H Oh, Xiang Y, Jegelka S and Savarese S 2016 Deep metric learning via lifted structured feature embedding Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition pp 4004-4012.
[7] Wang J, Zhou F, Wen S, Liu X and Lin Y 2017 Deep metric learning with angular loss Proceedings of the IEEE International Conference on Computer Vision (Venice: Italy) pp 2593-2601.
[8] Qian Q, Shang L, Sun B, Hu J, Li H and Jin R 2019 SoftTriple Loss: Deep metric learning without triplet sampling*Proceedings of the IEEE International Conference on Computer Vision* (Seoul: South Korea) pp 6450-6458.

[9] Kaya M and Bilge H Ş 2019 Deep metric learning: A survey *Symmetry* **11** (9) 1066.

[10] Musgrave K, Belongie S and Lim S N 2020 A metric learning reality check *arXiv preprint abs/2003.08505*.

[11] Lehnen A and Wesenberg G E 2017 *The Sphere Game in n Dimensions* http://faculty.madisoncollege.edu/alehnen/sphere/hyps.htm.

[12] Motlagh N H, Jannesary M, Aboulkheyr H, Khosravi P, Elemento O, Totonchi M and Hajirasouliha I 2018 Breast cancer histopathological image classification: A deep learning approach *bioRxiv.* 1 242818.

[13] He K, Zhang X, Ren S and Sun J 2016 Deep residual learning for image recognition *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* pp 770-778

[14] Huang G, Liu Z, Pleiss G, Van Der Maaten L and Weinberger K 2019 Convolutional networks with dense connectivity *IEEE Transactions on Pattern Analysis and Machine Intelligence* p 1.