DLMI Final report
Histology images query competition

Gi-Luen Huang
National Taiwan University, Taiwan
Graduate Institute of Communication Engineering
r09942171@ntu.edu.tw

I-Hsiang Chen
National Taiwan University, Taiwan
Graduate Institute of Electrical Engineering
f09921058@ntu.edu.tw

Abstract

In this final project, our task is histology images query. Given a histology image pair, answer whether they belong to the same category. Since we don't have the label of each image, we solve this task by unsupervised learning. In the training phase, we use contrastive learning to make model learn the feature representations. After training, we could assume that each image has its unique representations produced by the CNN backbone. Therefore, we can compare the relationship between two query images to determine whether these two images are the same category or not. Here we use the cosine similarity metric to compare the relationship between the two images, and we set the threshold to determine whether the two images are similar enough. If the similarity of the two images is less than the threshold, we answer that the two images are in the same category and vice versa. After experiments, the threshold is 0.705 in this project, and we obtain an accuracy of 0.87214 in the final result.

1. Introduction

With the rapid development of modern medicine, automatic histopathological image analysis has become an important issue in medical technology. The robust self-recognition system can effectively assist the doctor when facing diagnostic analysis. The main contributions of these medical image analysis systems are as follows: (i) Rapid and large-scale preliminary screening, and priority filtering of common & obvious cases which can reduce the burden on doctors during diagnosis and make medical resources more focused on complex cases. (ii) An excellent expert system can also be used to train doctors to eliminate the diagnostic variance between doctors. (iii) Deep learning can understand specific cases that may not be recognized in the past, and it can also effectively summarize common characteristics and potential causes.

Histopathological examination is greatly helpful to the diagnosis or treatment of diseases, and the doctor can diagnosis possible diseases, diseased organs, and the cause

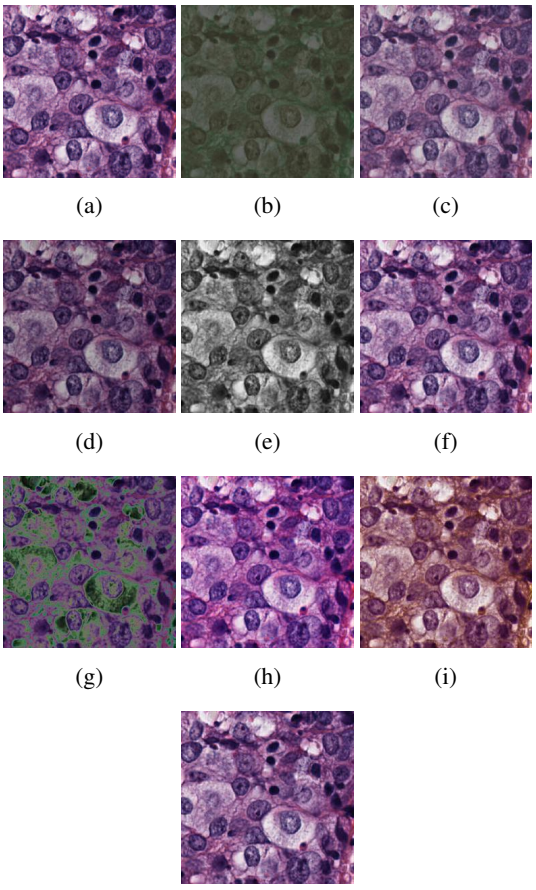

Figure 1: Example of data aumentation. (a) Original image (b) ColorJitter (c) Change of contrast value (d) Change of Brightess value (e) GrayScale (f) Horizontal Flip (g) Solarization (h) Change of Saturation value (i) Change of Hue value (j) Gaussian Blur

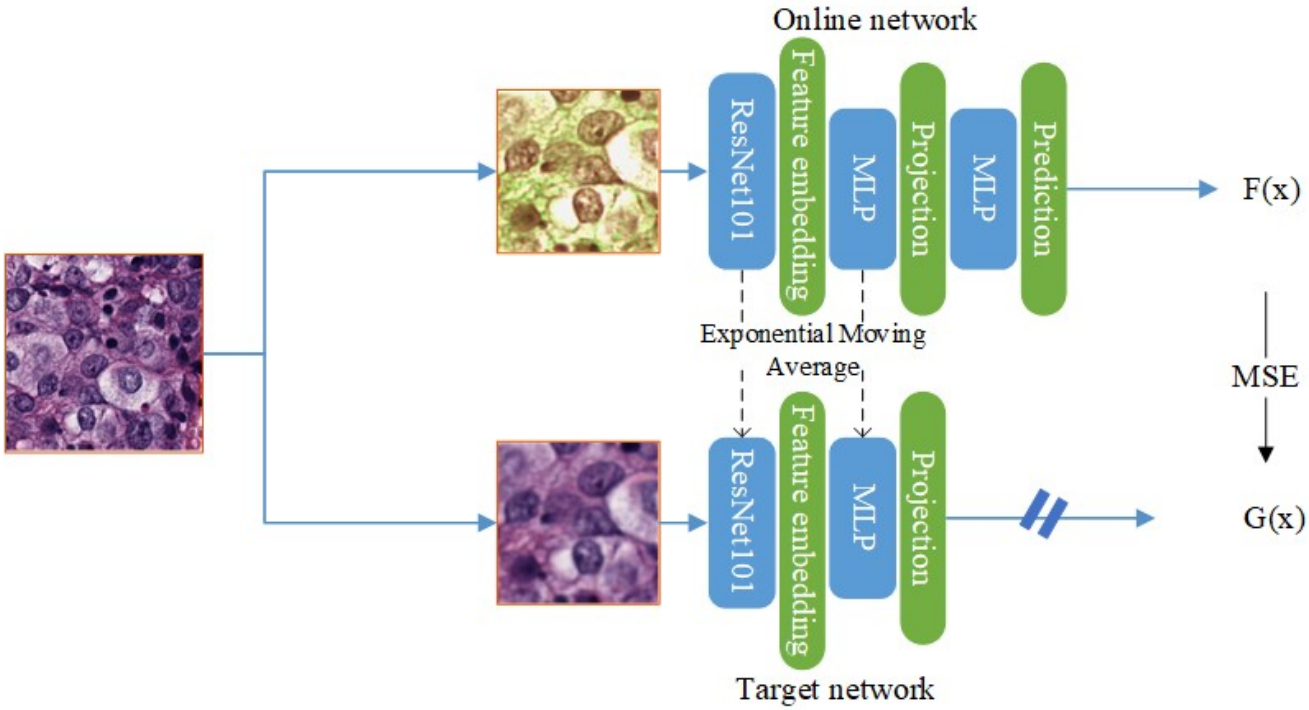

Figure 2: Our model.

of death. This information is not only helpful for understanding the disease but also improved for the diagnosis of similar diseases in the future, which provides corresponding drugs. However, that processing required the close attention of clinicians, the analysis result may be unstable when lacking complete information or long time idle. Therefore, that makes the problem of huge costs in labor costs, and the diagnosis variation on new diseases between doctors can't mitigate in a short time. That makes it is important to autonomously extractor hidden features for histopathological examination by unsupervised learning.

Unsupervised learning is a kind of machine learning, which does not need any labeled training set, and we expect the model can learn how to cluster unknown data. The main applications of unsupervised learning include cluster analysis, dimensionality reduction, and so on. Self-supervised is one branch of unsupervised learning, which uses its original data as a pseudo label, and learned how to reduce dimensionally and reconstruct the full image, such as autoencoder. In addition, contrastive learning has become the mainstream of today. It adopted different views of data as positive pairs to learn the similar feature represents, and it makes other images as contrast images to enhance distinguishing ability. The common methods include SimCLR [1], SimSiam [2], BYOL [3]. However, we find that the quality of negative samples is extremely important for contrastive learning methods, which can easily lead to 'collapsing' and destroy the stability of clustering. Therefore, we choose BYOL as our method, which doesn't require any negative samples, and it can more effectively learn the representative for each image.

Our main contributions in this task are as follows:

- We employ BYOL [3] as the main architecture for training and analysis of the impact of different data augmentation. Finally, we obtain a final result of 87.21% and 87.58% on the public and private dataset, respectively.
- We adopt projection and rank-10 to analyze the model representability and verify our method comprehensively.

2. Methods

In this work, we use the recently popular method [3]. We'll introduce how we train our model and test the result.

2.1. Training

Our model is illustrated in Fig 2. Since we don't have the annotation of the images, we cannot use the standard supervised learning to train our model. Instead, we apply data augmentation method. The image augmented by the

same image should have a similar representation. Here we use many data augmentation method, including ColorJitter, RandomGrayScale, RandomHorizontalFlip, GaussianBlur, Solarization, and RandomResizedCrop, etc. The visualization of each data augmentation method is illustrated in Fig 1. Our network comprises two parts: online network and target network. The online network consist of three stages: a feature extractor, a projector and a predictor. The target network has the same network as the online network while using a different weights. The target network provides a regression task to train the online network, and the model parameters are an exponential moving average (EMF) of the online parameters. The moving average rate is set at 0.99 in this work. Finally, the loss function can be defined as:

$$L = 2 - 2 \cdot \frac{\langle F(x), G(x) \rangle}{\|F(x)\|_2 \cdot \|G(x)\|_2} \tag{1}$$

Where $F(x)$ is the output of the online network and $G(x)$ is the output of the target network.

2.2. Inference

In this work, we have to determine whether the query images are as the same category. Therefore, we apply the feature extractor which is pretrained by the method described in section 2.1, to extract the representation of the image. Such representations are used to compare the relationship by using e.q 1. We set the threshold to determine whether the query images are similar enough. After experiments, the threshold is set at 0.705 in this work. If the similarity is less than 0.705, we determine the two query images are in the same category and vice versa.

3. Experiments

We conduct several experiments, the comparison result is illustrated in Table 1. We employ the two models: Resnext50 and Resnext101. The results demonstrate that the performance of Resnext101 is better than Resnext50. We also compute the Euclidean distance between the features of the query image and other images, and list the rank 10 images that are closest to the query image. The visualization result is illustrated in Fig. 4. The result shows that the features are very similar if the class of images is the same. Besides, we employ PCA to reduce the dimensions of the features extracted by Resnet101, and the visualization result on 2-D spaces are illustrated in Fig. 3. Fig. 3 demonstrates that the cluster of the images, and we can see that the clusters are roughly split into 3 groups.

3.1. Dataset

The dataset contains many unlabeled histology images, including 6187 training data, 3427 testing data, and 20 whole slide training data. In addition, there are 186 query image pairs from the training data to help us select the best model.

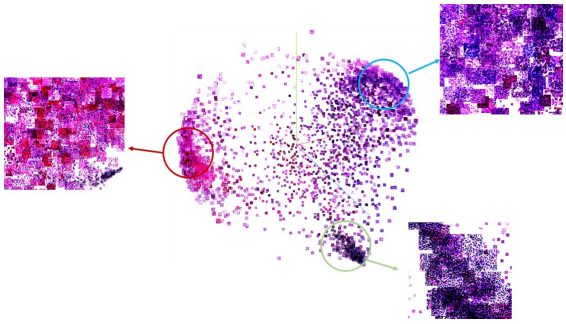

Figure 3: PCA result.

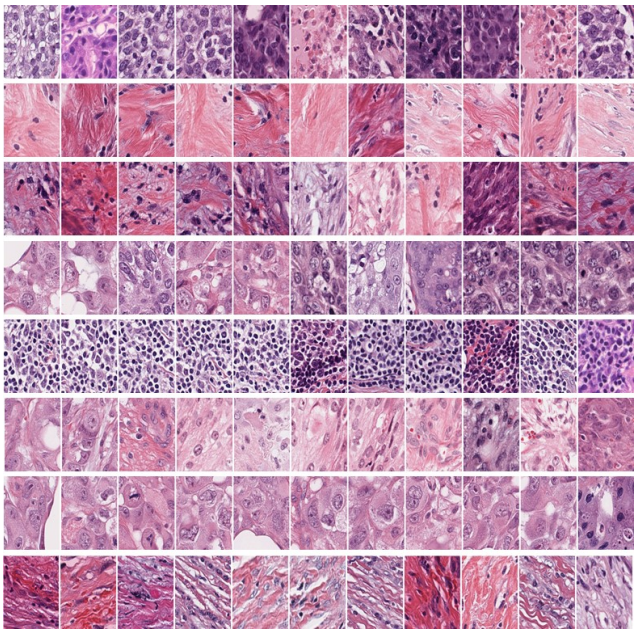

Figure 4: A query image (first column) and rank 10 images (the rest).

3.2. Implementation details.

Our method is implemented via Pytorch and is run on a server with two NVIDIA GeForce RTX 3090 GPUs and two NVIDIA Tesla V100 GPUs holding a graphics memory capacity of 24GB and 32GB for each one, respectively. We adopt the SGD optimizer with a learning rate of 0.0003 and momentum = 0.9 for the moment estimation. The batch_size and epoch are 32 and 30, respectively. In addition, the projection size and the projection hidden size are 512 and 2048, respectively.

Models	Threshold	Public(%)	Private(%)
Resnext50	0.95	82.99	83.03
Resnext50	0.8	86.54	85.98
Resnext50	0.72	86.93	86.82
Resnext101	0.65	86.93	87.19
Resnext101	0.6	86.16	86.34
Resnext101	0.63	86.61	86.87
Resnext101	0.68	87.18	87.54
Resnext101	**0.705**	**87.21**	**87.58**

Table 1: The accuracy of resnext50 and resnext101.

4. Discussion

This subsection will discuss the effectiveness of this work. In order to effectively analyze the actual effectiveness of the proposed method, the C-index has been adopted to evaluate performance on Kaggle. As the result of Table 1, we can find that Resnet101 has the more robust feature representability. And the most important factor is data augmentation, example of data augmentation can see Fig. 1, ColorJitter can provide appearance diversity to help adapt to data diversity. Horizontal Flip and RandomCrop can eliminate the scale-variance problem. The Gaussian Blur can also enhance model generalization.

On the other hand, in order to observe the effectiveness of the model. We observe the rank-10 visualization in Fig. 4. We select eight of the most representative sample as query data to sort similarities from the gallery. Then plot the first ten results to analyze. We can find the ranking result has similar histopathological appearances. Go further, we adopted PCA to project embedding feature, and we can obviously observe the clustering result, such as Fig. 3. From this, we can see that the BYOL can effectively cluster unknown domain data, and it can achieve a more robust classification ability.

5. Conclusion

To conclude, this work proved the effectiveness of BYOL for unsupervised learning on histopathological images. We also use different visualizations to prove the model's distinguishing capabilities. Finally, we achieve the second score in the public dataset. In the future, we can also combine the Variational AutoEncoder (VAE) to enhance appearance information and make the feature extractor more critical.

References

[1] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton. A simple framework for contrastive learning of visual representations. In *International conference on machine learning*, pages 1597–1607. PMLR, 2020.

[2] X. Chen and K. He. Exploring simple siamese representation learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 15750–15758, 2021.

[3] J.-B. Grill, F. Strub, F. Altché, C. Tallec, P. H. Richemond, E. Buchatskaya, C. Doersch, B. A. Pires, Z. D. Guo, M. G. Azar, et al. Bootstrap your own latent: A new approach to self-supervised learning. *arXiv preprint arXiv:2006.07733*, 2020.