SKIN CANCER ANALYSIS WITH DEEP VISION

Gia-Bao Le[1,2,*], Van-Tien Nguyen[1,2,*], Trung-Nghia Le[1,2] and Minh-Triet Tran[1,2,3]

[1] *University of Science, VNU-HCM, Vietnam*
[2] *Vietnam National University, Ho Chi Minh City, Vietnam*
[3] *John von Neumann Institude, VNU-HCM, Vietnam*

Abstract

We propose a novel method that utilize the benefit of contrastive learning in skin cancer analysis of Whole-Slide Tissue Images.

1. Introduction

Clinical cancer diagnosis requires pathologists to inspect histopathological images. Whole-Slide Tissue Images (WSI), which are digital slides created by virtual microscopy on glass slides, are widely used in modern pathology. However, the entire process is time-consuming, hard work, and challenging for pathologists. Therefore, there is an urgent demand for computer-aided diagnosis to aid in this process. Annotating skin regions in whole slide images (WSI) is a crucial step in the computer-aided diagnosis of skin cancer. Accurate identification of skin regions in pathological images plays a critical role in the detection and classification of skin lesions. However, pathological images often contain a variety of tissues, making it challenging to identify skin regions accurately.

Self-supervised learning is a promising path to advance machine learning [1] because it has the ability to extract useful features from vast unlabeled data. The performance of self-supervised asymptotically approaches the supervised in many domains (image, text, video, audio, time series).

Contrastive learning is a pearl of the deep metric family of self-supervised. It origins from contrative loss.

To avoid the abandonment of negative samples i.e. large batch size, non-contrastive is a self-supervised learning technique that uses a special architecture e.g. clustering, pseudo-labels, stop-gradient, momentum.

In this paper, there are four main contributions:

- Construct train, validation, and test set from the raw dataset.
- Propose method
- Benchmark

2. Related Work

Contrastive learning collect attention from the research community recently.

SimCLR [2] is a method that treats two augmentation views of an image as a positive pair and views of different images as negative samples. By applying contrastive loss, the positive pair is pulled closer, and negative samples are separated. SimCLR has achieved success in many natural image domains by augmentations and large batch size which means a large number of negative samples.

BYOL [3] and SimSiam[4], have the same general architecture. Two views x_1 and x_2 of the same image are generated with two different augmentations, and x_1 is passed into the online

[*] Corresponding author.

[†] These authors contributed equally.

backbone network on the left, while x_2 is passed into the target backbone network on the right. The outputs of these two backbone networks are passed into the corresponding projection MLPs, and then a prediction MLP is used to predict the projected representation of x_2 from the projected representation of x_1. SimSiam uses the same network for the online and target backbone and projection networks and uses a stop-gradient to prevent gradient signal from propagating through the second branch. BYOL also uses a stop-grad, but additionally uses an exponential moving average (EMA) to update the target backbone and projection networks.

SimTriplet [5] is a volunteer that applies non-contrastive method into the skin lesion domain. It bases on the SimSiam architecture but with two positive pairs, one for two augmentations of same center patch, and another is a center patch and nearby patch. Similar to SimSiam, SimTriplet have worse performance on unbalanced datasets, however, still archive better result than SimSiam on digital histology image [5]

SupCon [6]is a supvised version of SimCLR, takes both the advantage of contrastive method and the information of label. The contrastive loss used in SupCon have more positive terms which from same-class images in a batch.

3. Proposed Method

Our method takes advantage of the two methods SimTriplet, Supcon. Instead of label imgages like Supcon, we sample n nearby patches of an input image and assign it as a positive sample The nearby patches are used as same-class image

3.1. Loss function

Within a two-view batch, let $i \in I \equiv \{1, ..., 2B\}$ be the index of arbitrary augmented samples.

$$\mathcal{L} = \sum_{i\in I} \mathcal{L}_i = -\sum_{i\in\mathcal{I}} \left(\log \frac{\exp(z_i.z_a(i)/\tau)}{\sum_{j\in\mathcal{J}(i)} \exp(z_i.z_j/\tau)} + \frac{1}{|\mathcal{N}(i)|} \sum_{n\in\mathcal{N}(i)} \log \frac{\exp(z_i.z_n/\tau)}{\sum_{j\in\mathcal{J}(i)} \exp(z_i.z_j/\tau)} \right)$$

Here, $\tau \in \mathbb{R}^+$ is a scalar temperature parameter, and $\mathcal{J}(i) \equiv I \backslash \{i, a(i)\}$. The index i is called the anchor, index $a(i)$ is called the positive, and the other $2(B-1)$ $(\{k \in \mathcal{N}(i)\})$ indices are called the negatives.

4. Experimental Results

4.1. Dataset

CAnine CuTaneous Cancer Histology (CATCH) dataset [7] provides a large size of the dataset and extensive annotations that can be utilized for segmentation or classification tasks. Due to various homologies between the species that are shown in many previous works, canine cutaneous tissue can serve as a model for human samples. For the training set, we sampled 238 WSI and then random cropped 1000 images from each image, collecting 238000 images for the set. For the validation set, we sample 49 WSI and collect 3000 images for each class(Background, Tumor, Dermis, Epidermis, Inflamed/Necrosis, Subcutis). For the test set, we sample 35 WSI and collect 5000 images for each class. In all our training setups, images for model training and testing are resized to 128×128 resolution. We used SGD as optimizer with a learning rate $lr = 0.05$, momentum $= 0.9$, and weight decay of $= 0.0001$. Mixed Precision Training technique [8] is also applied for reducing computation expense and training time; therefore, the new learning rate is $lr \times BatchSize/256$.

Method	Class	Accuracy	
		Epoch 150	Epoch 300
SimCLR(512)	Dermis	72.1	71.46
	Epidermis	82.73	85.24
	Inflamn-Necrosis	73.55	72.38
	Subcutis	68.72	69.36
	Tumor	80.20	79.13
	Background	99.30	99.23
SimCLR(1024)	Dermis	71.8	70.32
	Epidermis	82.13	83.68
	Inflamn-Necrosis	70.61	73.32
	Subcutis	70.16	71.81
	Tumor	76.5	77.13
	Background	99.47	98.99
SupCon(512)	Dermis	74.07	75.13
	Epidermis	88.01	88.75
	Inflamn-Necrosis	82.31	80.87
	Subcutis	77.52	80.02
	Tumor	79.23	80.77
	Background	99.1	99.32

Table 1
Accuracy per class in 200 and 300 epoch

4.2. Linear evaluation

The pre-trained ResNet-50 backbone is frozen and followed by a linear layer with bias. When finetuning with annotated data, only the extra linear layer is trained. The SGD optimizer is also used with learning rate $lr = 0.005$, momentum $= 0.9$, and weight decay $= 0$. We use 5-fold cross-validation: four of five for training, the remaining one for validation. The linear classifiers are trained for 30 epochs.

4.3. Results

The result 1 show that our method archived better accuracy on each class of test set: Inflamn-Necrosis and Subcutis increased about 10%, Tumor class remained the same, and the others classé increased 2-5 %. The next testing phase include adding finetune 1 %,10 %,25 %,50 %, also divide the test set into 7 types of tumors. We have not added SimTriplet and SimSiam yet but the result is outperformed by the models in 1.

5. Conclusion

In this project, We present a method that can take the advantage of WSI image to create a simple and effective SSL method for unsupervised learning. By using the nearby patches of each image in the train set, the model can pull the inter sample similarly togetherfor better convergence and performance result

References

[1] R. Balestriero, M. Ibrahim, V. Sobal, A. Morcos, S. Shekhar, T. Goldstein, F. Bordes, A. Bardes, G. Mialon, Y. Tian, A. Schwarzschild, A. G. Wilson, J. Geiping, Q. Garrido, P. Fernandez, A. Bar, H. Pirsiavash, Y. LeCun, M. Goldblum, A cookbook of self-supervised learning, 2023. arXiv:2304.12210.

[2] T. Chen, S. Kornblith, M. Norouzi, G. Hinton, A simple framework for contrastive learning of visual representations, 2020. arXiv:2002.05709.

[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, B. Piot, K. Kavukcuoglu, R. Munos, M. Valko, Bootstrap your own latent: A new approach to self-supervised learning, 2020. arXiv:2006.07733.

[4] X. Chen, K. He, Exploring simple siamese representation learning, 2020. arXiv:2011.10566.

[5] Q. Liu, P. C. Louis, Y. Lu, A. Jha, M. Zhao, R. Deng, T. Yao, J. T. Roland, H. Yang, S. Zhao, L. E. Wheless, Y. Huo, Simtriplet: Simple triplet representation learning with a single gpu, 2021. arXiv:2103.05585.

[6] P. Khosla, P. Teterwak, C. Wang, A. Sarna, Y. Tian, P. Isola, A. Maschinot, C. Liu, D. Krishnan, Supervised contrastive learning, 2021. arXiv:2004.11362.

[7] F. Wilm, M. Fragoso, C. Marzahl, C. Bertram, R. Klopfleisch, A. Maier, M. Aubreville, K. Breininger, Canine cutaneous cancer histology dataset, 2022. URL: https://wiki.cancerimagingarchive.net/x/DYITBg. doi:10.7937/TCIA.2M93-FX66.

[8] P. Micikevicius, S. Narang, J. Alben, G. Diamos, E. Elsen, D. Garcia, B. Ginsburg, M. Houston, O. Kuchaiev, G. Venkatesh, H. Wu, Mixed precision training, 2018. arXiv:1710.03740.