CS5242 Project Report

Le Tan Dang Khoa
NUSNET ID: E1124464
Student ID: A0274718U

December 24, 2024

1 Introduction

1.1 Background

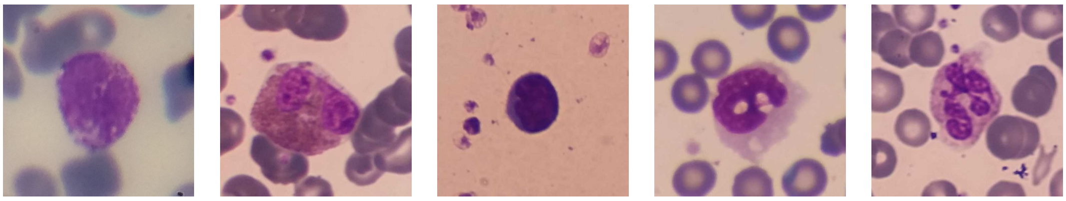
White blood cells (WBCs), or leukocytes are vital components of human immune system. When the body is under attacked from bacteria or viruses, they are produced and released into bloodstream to perform a wide range of functions [1]. Accurately classifying them helps doctors identify diseases more quickly and easily, which could lead to early treatment and intervention. Furthermore, it allows doctors to prevent further complications, thus better outcomes for patients. Developing new methods also broadens our understanding of the immune system, could potentially save lives and improve the well-being of many people. Such a system possibly reduces the manual workload for laboratory technicians, freeing up their time for more complex task and improving productivity.

Reliable white blood cell classification methods are a subject of ongoing research. Al-Dulaimi et al. [1] provided a comprehensive survey on techniques for handling WBCs data and presented the design of existing systems. [8] proposed to combine VGG features with an improved Swarm optimization to select the most relevant features for classification. In the context of medical imaging, a variant of convolutional auto-encoder named UNet is proposed [7], which has achieved remarkable results in various segmentation tasks. On the other hand, the Generalized Dice Loss (GDL) is introduced in [9] to address the problem of imbalance mask ground truths.

The final project aims to develop a classification model to identify five types of WBCs. In this report, I propose a variant of UNet network that jointly optimizes the reconstruction error, the label information and the mask of each cell. The next section briefly summarizes available data, following by the details of the proposed method and the evaluation results.

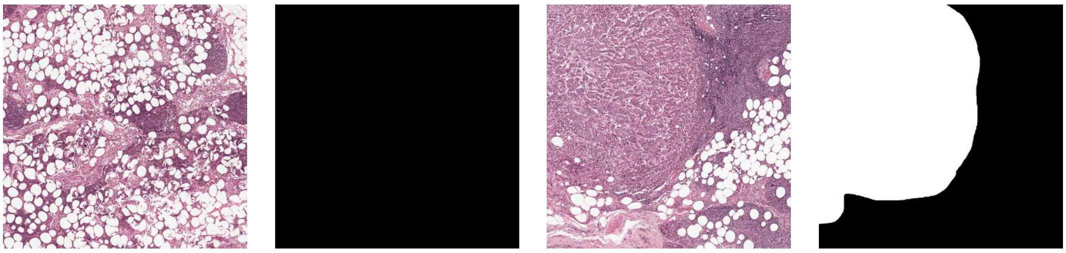
1.2 Datasets

The Raabin-WBC (WBC) dataset [5] contains microscopic images of 5 types of white blood cells that we want to classify, namely **basophils**, **eosinophils**, **lymphocytes**, **monocytes**,



(a) Basophil (b) Eosinophil (c) Lymphocyte (d) Monocyte (e) Neutrophil

Figure 1: Sample data from WBC dataset.



(a) Normal cell. (b) Normal cell mask (c) Tumor cell (d) Tumor cell mask

Figure 3: Sample data from CAM16 dataset.

and **neutrophils**. Each image of 575 × 575 resolution (Figure 1) contains one or two WBCs aligned in the center. About 10% of data also have a supplementary mask corresponding to the cell. The full training set (WBC_100) contains 8447 images with 842 masks while the validation set has 1728 images with no mask provided. In addition, there are 3 other variants: WBC_1, WBC_10, and WBC_50 corresponding to 1%, 10%, and 50% segregation of the WBC_100 dataset.

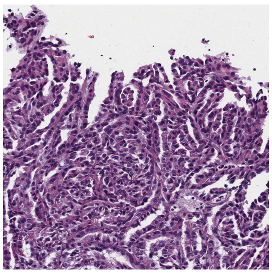


Figure 2: An image from the pRCC dataset.

The pRCC dataset [3] contains 1491 microscopic images of papillary renal cell carcinoma prepared by experienced pathologists. Each image is 2000 × 2000 resolution (Figure 2) with no label or annotated masks. The dataset is served as the source for pretraining the model.

The CAMELYON16 (CAM16) dataset [6] contains 1081 whole-slide images of 384 × 384 resolution from normal and tumor lymph nodes. It is split into training, validation and test sets with each of them having 757, 108, 216 images, respectively. Approximately 10% of the data have annotated masks. However, only tumor cells are annotated (Figure 3).

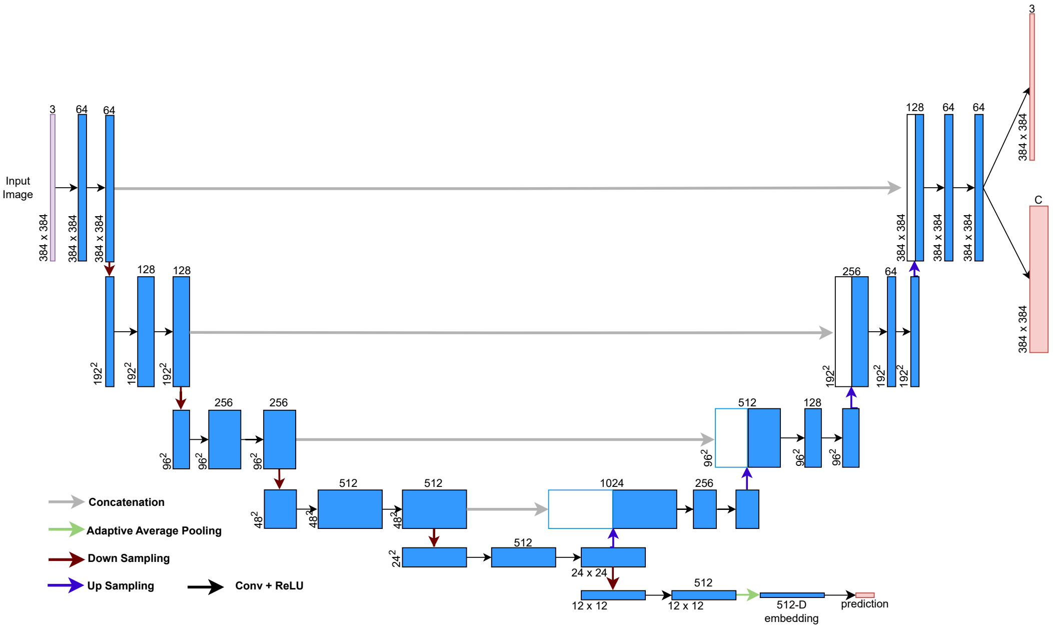


Figure 4: The proposed network design. The upper text and lower left of the block shows the number of channels and the size of each feature maps.

2 Proposed Method

2.1 Model Design

Inspired by UNet [7], the proposed network is an auto-encoder convolutional network in which the encoder extracts high-level features from the input while the decoder gradually recovers spatial information and generate a precise reconstruction. As depicted in Figure 4, during the decoding process, feature maps with the same resolution in the encoder are concatenated with the counterpart in the decoder. This works as a skip connection mechanism, allowing the network to combine high-level semantic information with fine-grained spatial information. Since the number of training images are limited, and to reduce the effect of overfitting, an upscaling layer is employed to enlarge the feature map instead of transposed convolution layers while the max pooling of size 2×2 is used for down sampling operators.

There are three ground truth types provided from the aforementioned datasets: (1) the original content of the image, (2) the label, and (3) the associated mask. Because the decoder already produces the reconstruction, two new convolutional branches are attached to the model to learn the others:

1. The classification branch is integrated to the embedding space. An additional down-sampling layer reduces the feature size to 12×12, followed by the global average pooling to extract a 512-dimensional embedding. This embedding serves as an input to a fully connected layer, which acts as a classifier, generating the final class predictions.

2. If the mask is available, the decoder generates additional output for the Generalized Dice Loss (GDL). The output feature map has size of $C \times H \times W$ where C is the number of classes in the dataset and $H \times W$ is the input size.

2.2 Loss Function

As described in the previous section, each dataset provides a set of unique groundtruths: while pRCC does not contain any labels, CAM16 and WBC have both labels and partial masks. To fully utilize all of them, a jointly cost function comprised of 3 losses is proposed as follows:

- The standard mean-squared loss L_{MSE} is used minimizing the differences between input images and their reconstruction. This loss can be used on all 3 provided datasets.
- The Cross-Entropy Loss $L_{\text{CE}} = -\sum_i^C y_i \log(f(s)_i)$, where $f(s)$ is a softmax function and s is the unnormalized probabilities, is employed to classify C types of cells from the CAM16 and WBC datasets.
- The Generalized Dice Loss from [9] is incorporated into the training to optimize the mask prediction. Derived from the Dice Loss, GDL adjusts the contribution of each class by assigning higher weights to less prevalent classes.

  $$L_D = 1 - 2\frac{\sum_{l=1}^{C} w_l \sum_i \hat{y}_{li} y_{li}}{\sum_{l=1}^{C} w_l \sum_i (\hat{y}_{lo} + y_{li})} \tag{1}$$

  the term $w_l = \frac{1}{(\sum_i r_{li})^2}$ to balance the loss across different classes by the inverse of class volume while $\hat{y}$ is the feature map outputs after applying sigmoid from the $C \times H \times W$ output branch and y is the mask target output.

In the end, the total loss of the training is formulated as follows:

$$L_{\text{total}} = L_{MSE} + \lambda_{\text{CE}} L_{\text{CE}} + \lambda_D L_D \tag{2}$$

In Equation 2, λ_{CE} and λ_D are hyperparameters for controlling the CE and GD losses, respectively.

3 Experiments

3.1 Experimental Details

Training deploys the Adam optimizer [4] for 32 epochs with step scheduling (step size = 12). Gradient clipping is employed to stabilize the training due to model and loss complexity. The input size of the model is set at 384×384 pixels. Data augmentation involves randomly cropping to match input size, random horizontal and vertical flips for both the images and their corresponding masks. Emphasizing classification over segmentation learning, the training is forced to prioritize L_{CE} over L_D, thus $\lambda_{\text{CE}} = 2.0$ and $\lambda_D = 1.0$.

3.2 Pretrain Performance

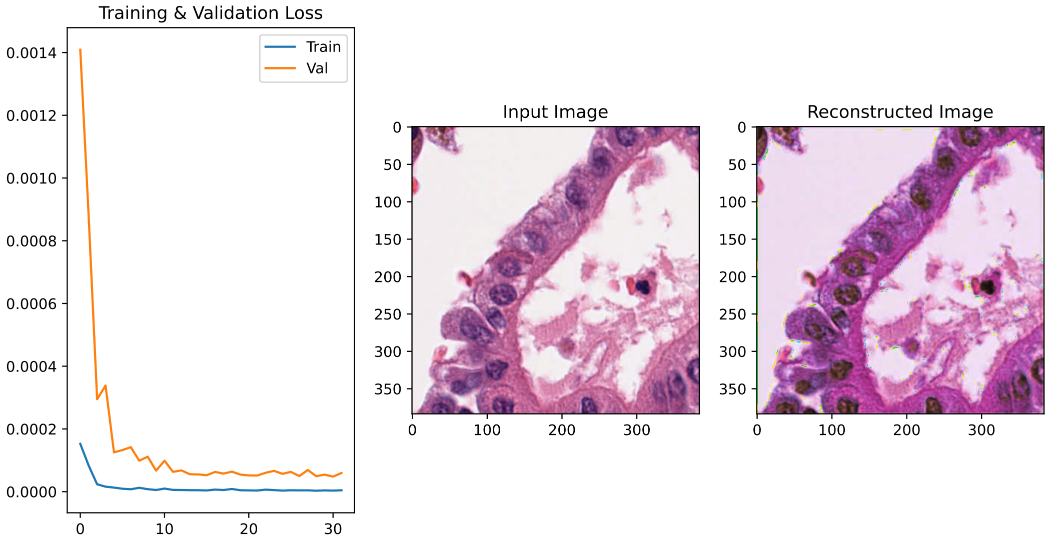


Figure 5: Train/Validation losses and the reconstruction image of the pRCC model.

To measure the performance on pRCC dataset, 10% of the training images are utilized as a validation set. As depicted in Figure 5, both training and validation losses show a substantial decline after 10 epochs, suggesting that the proposed model effectively learns from the provided dataset. Moreover, the reconstructed image closely resembles the original input, but with minor artifacts such as erroneous pixels at cell boundaries and dimmer backgrounds.

Method	1	2	3	4
CE Loss	✓	✓	✓	✓
pRCC		✓		✓
Dice Loss			✓	✓
Accuracy	0.9259	0.9167	0.9259	0.9213

Table 1: Pretrain accuracy on CAM16 testset.

Dataset	w/o pretrain		with pretrain	
	w/o mask	with mask	w/o mask	with mask
WBC_1	0.8860	0.8958	0.7228	0.7240
WBC_10	0.9491	0.9554	0.9271	0.9375
WBC_50	0.9745	0.9774	0.9705	0.9653
WBC_100	0.9792	0.9826	0.9740	0.9722

Table 2: Model accuracy on WBC's test set.

Since CAM16 comes with labels, the classification branch is added to the model. In order to observe the benefit of both pRCC dataset and GDL, models are trained with and without them. Table 1 shows the accuracy $\text{acc} = \frac{\#\text{correct predictions}}{\#\text{samples}}$ of the CAM16 testset. Although all 4 models train on CAM16 achieves more than 90% accuracy, the one using only Cross-Entropy loss outperforms the other settings with negligible margin.

Table 2 presents the validation accuracy of models on the WBC_100 dataset when training on 4 settings: with and without pretrained pRCC+CAM16 dataset, with and without mask training. When training on limited WBC data such as WBC_1 or WBC_10, the models without pretrain significantly outperform others with large margins. One explanation is that because WBC images from WBC_1 and WBC_10 are scarce, the model still overfits to the pretrained datasets, i.e. pRCC and CAM16 datasets, leading to subpar performance. Still, we can see the benefits of GDL when it helps to gain small margins on models training from scratch. It is clear that increasing the training size improves the accuracy across all 4 settings. The best performance belongs to the model training from scratch with GDL.

To better understand how well the model learns the visual features of WBC data, t-SNE [10] is utilized to project embedding features (Figure 4) from 512 dimensions to 2 dimensions for visualization purpose. As demonstrated in Figure 6, the model is capable of distinguish between different type of cells, particularly basophils, eosinophils, and neutrophils. However, there are still a few overlapping regions, such as a few neutrophils located near the lymphocyte region and a portion of monocyte region overlapping with the lymphocyte region.

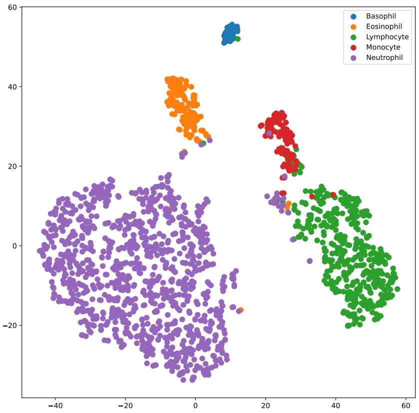


Figure 6: Embedding visualization from WBC_100 validation set via t-SNE.

4 Conclusion

In this project, a variant of the UNet architecture is proposed for identification of five types of white blood cells. Experimental results demonstrated the benefits of utilizing the UNet architectures, which effectively combines the idea of auto-encoder and skip connections. While multiple task learning further enhanced model performance, the observed benefits were limited due to the scarcity of mask ground truths and the model complexity. Future research could explore various avenues for performance improvement and address the issue of lacking high-quality annotated data, such as self-supervised learning [2], more powerful network designs and better training schemes.

References

[1] Khamael Abbas Khudhair Al-Dulaimi, Jasmine Banks, Vinod Chandran, Inmaculada Tomeo-Reyes, and Kien Nguyen Thanh. Classification of white blood cell types from microscope images: Techniques and challenges. *Microscopy science: Last approaches on educational programs and applied research (Microscopy Book Series, 8)*, pages 17–25, 2018.

[2] Shekoofeh Azizi, Basil Mustafa, Fiona Ryan, Zachary Beaver, Jan Freyberg, Jonathan Deaton, Aaron Loh, Alan Karthikesalingam, Simon Kornblith, Ting Chen, et al. Big self-supervised models advance medical image classification. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 3478–3488, 2021.

[3] Zeyu Gao, Bangyang Hong, Xianli Zhang, Yang Li, Chang Jia, Jialun Wu, Chunbao Wang, Deyu Meng, and Chen Li. Instance-based vision transformer for subtyping of papillary renal cell carcinoma in histopathological image. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part VIII 24*, pages 299–308. Springer, 2021.

[4] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[5] Zahra Mousavi Kouzehkanan, Sepehr Saghari, Sajad Tavakoli, Peyman Rostami, Mohammadjavad Abaszadeh, Farzaneh Mirzadeh, Esmaeil Shahabi Satlsar, Maryam Gheidishahran, Fatemeh Gorgi, Saeed Mohammadi, et al. A large dataset of white blood cells containing cell locations and types, along with segmented nuclei and cytoplasm. *Scientific reports*, 12(1):1123, 2022.

[6] Geert Litjens, Peter Bandi, Babak Ehteshami Bejnordi, Oscar Geessink, Maschenka Balkenhol, Peter Bult, Altuna Halilovic, Meyke Hermsen, Rob van de Loo, Rob Vogels, et al. 1399 h&e-stained sentinel lymph node sections of breast cancer patients: the camelyon dataset. *GigaScience*, 7(6):giy065, 2018.

[7] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, pages 234–241. Springer, 2015.

[8] Ahmed T Sahlol, Philip Kollmannsberger, and Ahmed A Ewees. Efficient classification of white blood cell leukemia with improved swarm optimization of deep features. *Scientific reports*, 10(1):2536, 2020.

[9] Carole H Sudre, Wenqi Li, Tom Vercauteren, Sebastien Ourselin, and M Jorge Cardoso. Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: Third International Workshop, DLMIA 2017, and 7th International Workshop, ML-CDS 2017, Held in Conjunction with MICCAI 2017, Québec City, QC, Canada, September 14, Proceedings 3*, pages 240–248. Springer, 2017.

[10] Laurens Van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of machine learning research*, 9(11), 2008.