Generative Adversarial Networks for Peripheral Cell Blood Images Standardization

A. Pérez[1], S. Alférez[2], K. Barrera[1], A. Merino[3] and J. Rodellar[1]

[1] Department of Mathematics, Barcelona Est Engineering School, Technical University of Catalonia, Barcelona, Spain, alejo.perez@estudiant.upc.edu, kevinivancho@hotmail.com, jose.rodellar@upc.edu

[2] Faculty of Natural Sciences, Universidad del Rosario, Bogotá, Colombia, edwin.alferez@urosario.edu.co

3 Biomedical Diagnosis Center, Hospital Clínic of Barcelona, amerino@clinic.cat

Abstract

Our multidisciplinary research group has been developing models based on Convolutional Neural Networks for the automatic morphological recognition of normal, reactive and abnormal blood cells circulating in blood, lymphoma and acute leukaemia.

The problem addressed in this study is related to the effect of different staining protocols used to prepare the blood smear from which microscope images are analyzed. Specifically, the automatic blood cell classification with models trained using an image database collected in a single hospital may be affected when images are coming from another hospital. Thus resulting in a certain number of misclassified images. Such accuracy decrease is because some color, staining and texture characteristics may suffer variations in the images depending on the staining protocol.

Generative Adversarial Networks (GANs) were leveraged to transform the images of the X domain (Hospital Germans Trias i Pujol) into the Y domain (Hospital Clínic). Characteristics of color, staining and texture of the latter were emulated. After this transformation, the classification accuracy of the images taken at Hospital Germans Trias i Pujol improved. CycleGAN and ColorizationGAN architectures were used for the transformations. ResNet 18, ResNet 34, ResNet 18 with Focal Loss and ResNet 34 with Focal Loss were used for the classifications. The classification accuracy improved from 34% to 81%.

1. Introduction

Morphological analysis of cells circulating in peripheral blood is a useful tool for the initial and quick detection of haematological and non haematological diseases. Visual inspection of blood cells on the stained smear using the microscope requires skills and expertise, being prone to subjectivity and inter-observer variability. Specialized automated microscopes are currently in use in specialized laboratories to help with such analysis by pre-classifying cells, such as CellaVision DM96 [1]. There exist a large number of pathologies, like leukemia, lymphoma and others, which are associated with the presence of specific classes of cells in peripheral blood. Morphological differences exist among those cell types but they are sometimes difficult to discern. In the last decade, researchers and practitioners have carried out a significant research effort to develop image-based machine learning algorithms to achieve automatic recognition in a more objective way and assist the clinical pathologists in their initial diagnosis [2, 3]. Deep learning has seen a recent tremendous interest in being applied to different problems involving automatic blood cell classifications [4, 5].

Generative Adversarial Networks (GAN) have been recently proved successful in allowing transformation procedures to improve classification models in biomedical problems. For instance Hu et al. [6] propose the use of these architectures as a complement to the unsupervised learning for the visual representation of histopathologies.

This work responds to the need to standardize the staining of the images obtained in different laboratories when developing and implementing machine or deep learning models. Our group has been working mainly with images obtained in the Core Laboratory of the Hospital Clínic of Barcelona (Clinic from now on). Therefore, training and validation of algorithms have been performed with data sets of images acquired with specific colorization depending on the staining used to prepare blood smears. Although smears are routinely performed by automatic stainers, they are prone to variability depending on the proportion of certain chemicals, times and local protocols. Therefore, classifiers based on convolutional neural networks (CNN) may decrease their effectiveness when tested with images from other laboratories.

This work explores whether using GAN can help to address this problem. The starting point is that CNN models have been trained with a data set of images from Clinic and we plan to test that models using images from a different center, concretely Hospital Germans Trias i Pujol (called HGTP from now on).

There are three specific objectives proposed in this work, which are directly related to three different methodology blocks:

1. Development of a preprocessing algorithm to transform the images of HGTP emulating the staining characteristics of the Clinic images with CycleGAN.
2. Assessment of the previous process through classification of the HGTP images (original and transformed ones) by CNNs pretrained with Clinic images. Moreover, qualitative evaluations of the transformation by clinical pathologists.

3. Artificial generation of abnormal peripheral lymphocytes out of normal lymphoid blood images (and vice versa) by means of CycleGAN. Those images belong to the Hospital Clínic.

2. Materials

As for Block 1, three different datasets were used to train the CycleGAN to transform the HGTP images: The Clinic training set (3,000 images), the Clinic testing set (184 images), and the HGTP dataset, used for both training and testing (due to the scarcity of the set) containing 184 images. The Clinic training and testing sets consisted of a random sampling of the Clinic database containing all reactive, abnormal lymphocytes and leukaemic cells (blasts). To train the CycleGAN, the specific number of each class was not taken into account because just changes in color and style were intended. The images were labeled with the categories mentioned before but just the general labels of HGTP and Clinic were relevant in the transformation process. Then, they were taken into consideration in the classification stage.

In the Second Block (classification) a sampling from the Clinic database was used to fine-tune the classifiers (CNN) prior to the classification of the HGTP images (both original and transformed). This data distribution is outlined in Table 1. We excluded the normal lymphocyte class as the HGTP dataset lacks this category. The HGTP dataset was split in training and testing to ensure the network does not see the same images both in training and testing, see Table 2.

Dataset	Validation	Testing	Training	Total per category
Abnormal Lymphocyte	243	69	1026	1338
Reactive Lymphocyte	112	102	439	653
Blast	227	93	866	1186
Total per set	582	264	2331	3177

Table 1. *Clínic Validation, Training and Testing Set*

Dataset	Training	Testing	Total per category
Abnormal Lymphocyte	22	27	47
Reactive Lymphocyte	27	30	57
Blast	43	37	80
Total per set	92	92	184

Table 2. *HGTP Training and Testing Set for Classification*

For Block 3, aiming to transform normal lymphocytes to abnormal lymphocytes and vice versa, a sampling dataset as shown in Table 3 was used. It was split in 3 subsets for each training and testing experiment. This was intended for performing three different experiments: confronting normal lymphocytes against lymphocytes circulating in blood in splenic marginal zone lymphoma (SMZL), then against mantle cell lymphoma cells (MCL) and finally against blasts.

	Abnormal Lymphocyte		Normal Lymphocytes	
Dataset	Training	Testing	Training	Testing
Normal Lymphocyte vs SMZL	1,000	583	799	287
Normal Lymphocyte vs MCL	800	381	800	286
Normal Lymphocyte vs Blast	1,040	200	987	100

Table 3. *Training and Testing sets for Block 3*

Regarding the network architectures, they were designed, trained, validated and tested using the *fast.ai* library (built on top of *Pytorch*) and *Python* programming language. It was also used with a DeepBox with a *GPU Nvidia Titan XP* (12 GB).

3. Methodology

3.1. Block 1 - Transformation of HGTP images by CycleGAN

Firstly, the datasets referent to the First Block were arranged to be introduced to the CycleGAN. As shown in Figure 1, both Clinic and HGTP training sets are confronted one against the other to train the CycleGAN during 200 epochs. We chose the parameter values intended for training similar dimensions datasets recommended by Jun-Yan Zhu, main developer of the architecture. These can be seen in its Github repository [7]. Mini-batches of 1 image were selected as well as a load size of 286 pixels and a crop size of 256 pixels. A number of 64 filters was chosen for both the last discriminator layer and the first generator layer.

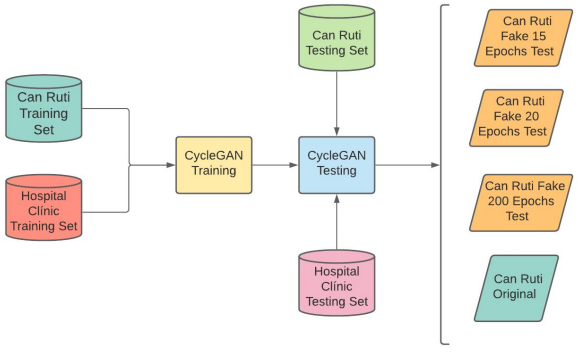

Figure 1. *Diagram of Transformation of HGTP images by CycleGAN*

After that, the testing sets were introduced to perform the testing step (transformation). The transformation was carried out bidirectionally although just the HGTP resulting datasets were kept. The CycleGAN model trained was tested at 10, 15, 20, 25 and 200 epochs. Every epoch represents one time for the model to have seen the whole training set and update its parameters. This multiple epoch testing was done because we noticed that the more epochs

the network was trained, the transformations involved more cell morphology structural changes. Therefore, it was necessary to select the epoch testing that emulates best the targeted style features. As visible in Figure 1, just 15 and 20 epoch generated datasets were selected because they complied with the Clinic standards. The transformation at 200 epochs was kept to show the resulting images when they underwent a transformation of morphology due to too many epochs in the training stage.

3.2. Block 2 - Transformed Cell Images Classification

The Block 2 procedure is outlined in Figure 2. We used ResNets with 18 and 34 layers pre-trained on the Imagenet Dataset [8] as well as a model provided by Clinic pre-trained with a bigger dataset than the used in this study (called ResNet34_Resample_NoSIND, from now on ResNet34_NoSIND for simplicity).

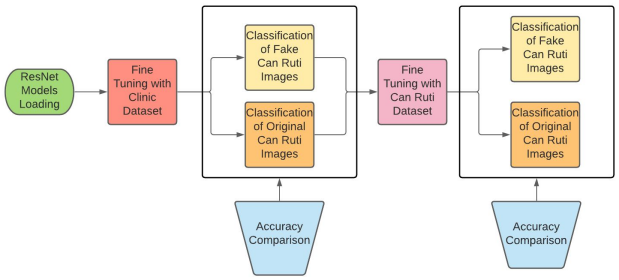

Figure 2. *Diagram of Transformed Cell Classification*

A Focal Loss function was added to the first two architectures, proven to improve the training convergence in unbalanced datasets [9]. Therefore, two new custom architectures were produced: Resnet18 Focal and Resnet34 Focal. The models (except ResNet34_NoSIND) were fine-tuned with the Clinic dataset distribution. We used the 1 cycle policy [10] to select the learning rate at the fine-tuning process (for the "Fine-tuning with Clinic Dataset" and "Fine Tuning with HGTP Dataset" stages). The HGTP Dataset was split in two (training and testing). After that, a test classification of the Fake HGTP images and the Original HGTP images was carried out (in the testing set). Then, a fine-tuning with this data distribution was performed. A posterior classification was done again with the aforementioned HGTP testing dataset after this fine-tuning. Finally, accuracy metrics were registered in both classification stages.

3.2. Block 3 - Image Cell Transformation by CycleGAN (Normal to Abnormal Lymphocyte and vice versa)

The Block 3 procedure is outlined in Figure 3. Three training experiments were done in parallel following the Block 1 same procedure for the CycleGAN training (200 epochs training).

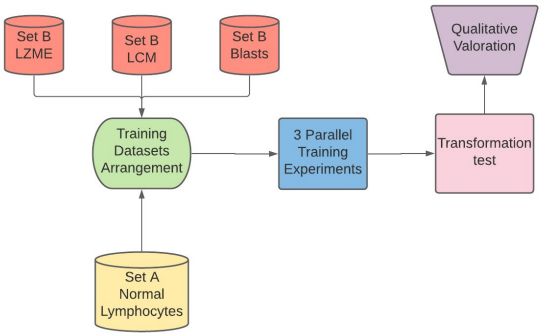

Figure 3. *Diagram of the Third Block, Transformation by CycleGAN between classes*

4. Results and Discussion

4.1. Results of Block 1 - Transformation by CycleGAN

As shown in Figure 4, after a qualitative inspection the best result was the transformation at 20 epochs of CycleGAN training. In the case of the 200 epoch transformation, it is visible that the transformations at 200 epochs of training suffered undesired morphological transformations, please see arrows in Figure 4.

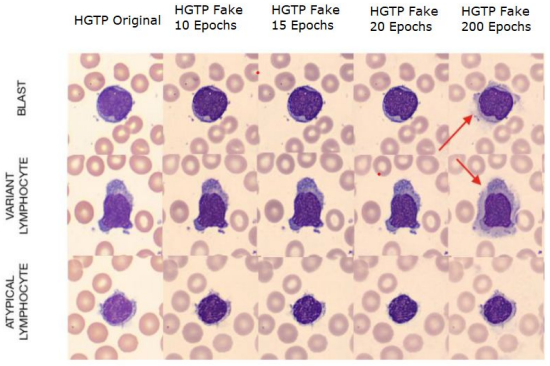

Figure 4. *HGTP Fake Dataset generation result by CycleGAN*

4.2. Results of Block 2 - Transformed Cell Classification

As shown in Table 4, there are classifications whose result did not exceed 0.5 of accuracy, representing a poor performance of the classifier model (executions in the original HGTP dataset). The accuracies in bold are those that have exceeded 0.6.

Dataset	HGTP Original	HGTP Fake 15 Epochs	HGTP Fake 20 Epochs
ResNet_18	0.3152	0.5869	0.5326
ResNet_18_Focal	0.4565	**0.6847**	0.5869
ResNet_34	0.2826	0.5543	0.5326
ResNet_34_Focal	0.3043	**0.6739**	0.5326
ResNet_34_No_Sind	0.3478	**0.8152**	**0.75**

Table 4. *Accuracy results of automatic classification of the* HGTP *Datasets (Fake ones and the Original)*

The transformation at 15 and 20 epochs showed a significant improvement in the performance of the model with respect to the classification in the Original Dataset. It is also shown in Figure 5 the normalized confusion matrix of the best test result without evident morphological changes (test with ResNet_34_No_Sind with the HGTP Fake CycleGAN dataset at 15 epochs).

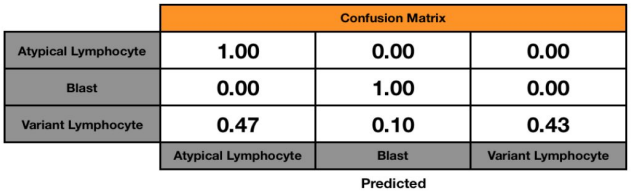

Figure 5. *Normalized Confusion Matrix of the best classification test result. Tested with the ResNet_34_No_Sind with the dataset Can Ruti Fake (15 epochs of training with CycleGAN)*

After having fine-tuned again the CNN models with a training set of HGTP dataset (for the three Fake HGTP ones and the Original) all accuracy results reached at least 0.9 for all HGTP datasets (Originals and Fake).

4.3. Results of Block 3 - Cell Transformation between Classes

In the Third Block, the image transformation from normal cells to other types of malignant cells has been possible thanks to the CycleGAN architecture. The conversion has also been performed in reverse and has been generally successful in both directions.

5. Conclusions

Regarding Blocks 1 and 2, the pre-processing algorithm to transform the HGTP into the Clinic standards has been successfully created. Classification accuracy has been used to determine the most effective image transformation technique is the CycleGAN usage with 15 epochs of training. The ResNet34_NoSIND model has achieved an accuracy of 81.52% in the dataset resulting from that transformation, increasing approximately 60 percentage points of accuracy with respect to the classification of the Original HGTP dataset. In the next stage, the high effectiveness of the fine-tuning has been proven being carried out in the training sets of the HGTP datasets, both in the Fake ones and in the Original. This is an alternative to preprocessing using CycleGAN in case the latter is not available. By means of this method, performances exceeding 90% have been achieved with all the models in the original dataset, even reaching 98.91%.

In matters of the Third Block, it has been possible to make a comparison and qualitative assessment of the cell transformations from Normal to Abnormal Lymphocytes and vice versa. The possibility of transforming blood cell images from one type to another opens up the possibility of training classifiers with images of malignant cells corresponding to rare diseases. Obtaining an adequate number of images containing this type of cells from low prevalence diseases is very difficult, so the use of CycleGAN architecture for their acquisition can be very useful.

It should be pointed out that an increase in the size of the training datasets of this study would lead to greater precision and quality of the training-testing performance of both CycleGAN and CNNs.

Acknowledgments

We are grateful to Cristian Morales-Indiano for providing the HGTP images. This work is part of a research project funded by the Ministry of Science and Innovation of Spain, with reference PID2019-104087RB-I00.

References

[1] A. Merino, R. Brugués, R. García, M. Kinder, F. Torres, and G. Escolar, "Estudio comparativo de la morfología de sangre periférica analizada mediante el microscopio y el CellaVision DM96 en enfermedades hematológicas y no hematológicas", Revista del Laboratorio Clínico, vol. 4, no. 1, pp. 3–14, Jan. 2011, doi: 10.1016/j.labcli.2010.10.002.

[2] J. Rodellar, et al. "Image processing and machine learning in the morphological analysis of blood cells", Int. Journal of Laboratory Hematology, Vol. 40, pp. 46-53, 2018.

[3] S. Alférez, et. al. "Automatic recognition of atypical lymphoid cells from peripheral blood by digital image analysis", American Journal of Clinical Pathology, Vol. 143, pp. 168-176, 2015.

[4] J. Zhao, M. Zhang, Z. Zhou, J. Chu, and F. Cao, "Automatic detection and classification of leukocytes using convolutional neural networks," Med Biol Eng Comput, vol. 55, no. 8, pp. 1287–1301, 2017, doi: 10.1007/s11517-016-1590-x. [Online]. Available: http://link.springer.com/10.1007/s11517-016-1590-x.

[5] A. Acevedo et al. "Recognition of peripheral blood cell images using convolutional neural networks", Comp Methods and Programs in Medicine, Vol. 180, 2019, https://doi.org/10.1016/j.cmpb.2019.105020.

[6] B. Hu, Y. Tang, E. I.-C. Chang, Y. Fan, M. Lai, and Y. Xu, "Unsupervised Learning for Cell-Level Visual Representation in Histopathology Images With Generative Adversarial Networks," IEEE J. Biomed. Health Inform., vol. 23, no. 3, pp. 1316–1328, May 2019, doi: 10.1109/JBHI.2018.2852639. [https://ieeexplore.ieee.org/document/8402089/.

[7] J. P. Zhu, T. Isola, and P. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," Unpaired image-to-image translation using cycle-consistent adversarial networks.

[8] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," Commun. ACM, vol. 60, no. 6, pp. 84–90, May 2017, doi: 10.1145/3065386. [Online]. Available: https://dl.acm.org/doi/10.1145/3065386.

[9] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal Loss for Dense Object Detection," arXiv: 1708.02002, Feb.2018.http://arxiv.org/abs/1708.02002.

[10] G. Sylvain, "The 1cycle policy," Github. Apr2018 https://sgugger.github.io/the-1cycle-policy.html