# EVALUATION OF DATA AUGMENTATION FOR GAN WITH LIMITED HISTOLOGICAL DATASETS

Ramzi HAMDI[1], Clément BOUVIER[1], Thierry DELZESCAUX[2], and Cédric CLOUCHOUX[1]

[1] WITSEE, Paris, France
[2] CEA-CNRS-UMR 9199, LMN, MIRCen, Univ. Paris-Saclay, France

**Abstract.** In digital histopathology, automatic biomarkers quantification at the whole slide level enables the precise characterization of a pathology progression or drug efficiency. Supervised Machine Learning (SML) algorithms allow the automation of such tasks but rely on large annotated learning datasets. Unfortunately, such datasets are not easily available, mainly due to the tedious task of manual segmentation. A strategy to overcome this problem is to artificially extend these sets using Data Augmentation. Although approaches relying on traditional Data Augmentation are the most developed and used, Generative Adversarial Networks (GAN) recently gained interest despite their intrinsic complexity and need of large learning databases. Indeed, GANs requirements in terms of learning data quantity and quality are often underestimated especially in cases where manually annotated data are scarce. In this exploratory work, we aimed at measuring GAN efficiency when fed with synthetic data generated with a set of traditional Data Augmentation methods. Generated models relevancy and accuracy were assessed by evaluating the quality of resulting segmentation with U-Net.

**Keywords:** Generative Adversarial Networks · Histology · Machine Learning · Data augmentation

## 1 Introduction

Histology is widely used to detect biological objects with specific staining such as protein deposits, blood vessels or cells. Staining quantification heavily relies on manual techniques, a tedious and time-consuming task, highly dependent on the bias of the experts. Therefore, automated methods, among which supervised machine learning SML [1], are increasingly used to automatically detect and quantify biological structures. One of the biggest issues facing the use of SML in Whole Slide Imaging is the lack of large, labelled and available datasets. The limited amount of annotated data can negatively impact subsequent segmentation quality of SML algorithms [2]. To overcome these issues, learning datasets can be artificially increased using data augmentation [3]. A number of methods have been proposed in the last decade, based on traditional methods. The main limitation is the degree of freedom of such methodologies [4]. In contrast GAN Data Augmentation algorithms allow a wider diversity of generated images. However GAN methods rely on large learning datasets in order to achieve robust generation with realistic images [4]. In cases where only limited learning datasets are available, traditional Data Augmentation methods can be used to increase the size of the original dataset to train the GAN Data Augmentation. However the optimal traditional Data augmentation method and the minimal amount of data to use remain not fully explored. The presented work aims at evaluating GAN generated histological learning dataset by training the GAN algorithm with different learning datasets configuration, either augmented or limited in size. The synthetic learning datasets are used to train a U-Net to classify an independent test dataset and the subsequent segmentation quality is evaluated.

## 2 Material and Methods

**Histological dataset** Dataset was composed of 114 histological sections extracted from a 13.5-months-old mouse amyloidosis model (APP/PS1dE9) brain stained with beta Amyloid Monoclonal Antibody (BAM-10) and counter-stained with Blue Reagent [4]. From the digitized sections, 100 images of 512x512 pixels were extracted and annotated by an expert. The annotations were considering two classes: 1- background and Blue Reagent stained tissue and 2- BAM-10 stained tissue. Then, 1,600 images of 128x128 pixels each were randomly extracted from the 100 images of 512x512 pixels. This dataset was split in half in an initial learning and a test dataset (800 couples each).

**Traditional Data Augmentation methods** A total of 5 common traditional Data Augmentation methods were selected to amplify the original learning dataset. The first method was based on rotation, reversal [6] and a custom-made circular translation method. The latter consisted in splitting an image into 2 parts, blending one part with the reversal of the other using the Gaussian Laplacian Pyramid Blend algorithm [6]. The other used methods were Random Affine [6], Random Gauss Blur [6], Random Elastic transformation [6] and HED Jitter [6]. The chosen methods were labeled in two groups: spatial Data Augmentation methods (sDA: Geometric, Random Affine and Random Elastic methods) and intensity Data Augmentation (iDA: Random Gaussian Blur and HED Jitter methods). The augmentation factor (i.e. the ratio between the augmented dataset and the initial one) was set to 5.

**DC-GAN induced Data Augmentation** The original method was based on the competition of two networks: a Generator network aimed at producing realistic images to fool a Discriminator network [12]. The Deep Convolutional GAN (DC-GAN) [13] consisted in convolution layers without max pooling or fully connected layers. This algorithm was chosen because of its versatility [4][14]. The DC-GAN was trained to produce a couple of generated images to simulate a learning dataset, with 360 epochs and an augmentation factor of 1.

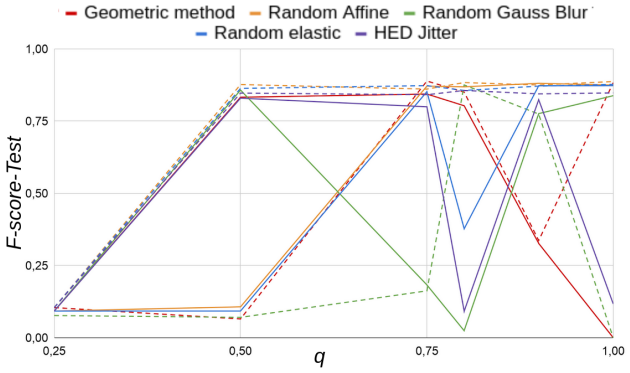
**U-Net segmentation and Data Augmentation validation** To validate the data augmentation protocol, we used U-Net, a supervised deep learning segmentation method developed for biomedical image segmentation [15]. The network was configured with default architecture parameters and trained with 35 epochs. A two-fold cross-validation (Direct Validation DV and Cross Validation CV) was performed through F-score computation against the test dataset (F-Score-Test).

**Proposed methodology** U-Net was trained with two different approaches. First the DC-GAN was trained with the learning dataset corresponding to 90% of the initial learning dataset (720 randomly selected images) without traditional Data Augmentation methods. In the second approach, the learning dataset was reduced by a factor q and amplified using each one of the 5 traditional methods previously presented. The possible values for q were 0.25, 0.5, 0.75, 0.8, 0.9 and 1. Each learning approach was performed in a two-fold-cross-validation.

## 3 Results

As a reference result, the F-Score-Test without augmentation and with the initial learning dataset was 0.877 in DV and 0.883 in CV. The first learning approach did not converge with the initial

**Fig. 1.** Comparison between traditional Data Augmentation method with the second learning approach. F-score-Test (DV continuous line and CV dashed line) against q values.




dataset and generated unusable couples of images (F-Score-Test close to 0). With a value of q<0.25, the DC-GAN was not able to generate exploitable images. The F-Score-Test for the iDA were not converging for high q values through the two validation conditions. In the sDA group, only the Geometric method has not converged for high q values (Fig. 1). The Random Affine plateaued above 0.86 for both validation conditions around q=0.75 (540 images of 128x128 pixels). The Random Elastic plateaued above 0.87 at q=0.9 (648 images of 128x128 pixels). For all synthetic datasets, no significant improvement in the resulting segmentation quality was measured.

## 4 Conclusion

In the light of these observations, the DC-GAN seemed to converge following specific conditions on the size of the intermediate learning dataset and the type of algorithm used to generate it. Random Elastic and Random Affine methods were both converging for both validation conditions at q=0.9. At this q value, the intermediate dataset sized 3,240 images of 128x128 pixels. This number of samples is consistent with the results presented in Frid-Adar et al. for a lower augmentation factor [4]. In the case of BAM-10 stained tissue, Random Elastic and Random Affine methods significantly achieved better F-Score-Test than the other tested Data Augmentation algorithms. The supposed conjecture was the high spatial distortion in the generated images. Then the convergence of DC-GAN was more sensitive to spatial diversity of the intermediate learning dataset than its colorimetric or intensity diversity [14]. To confirm this hypothesis, further work is currently conducted to measure spatial distortion resulting from the traditional Data Augmentation and correlate with the aforementioned conjecture. In this work, no synthetic learning dataset allowed a significant increase in the segmentation quality with U-Net compared with the initial learning dataset. This stagnation was due to the single augmentation factor used which was inferior to other augmentation factors found in the litterature [2]. Further work is conducted to vary the augmentation factor for each traditional Data Augmentation method to find an optimal value. With the increase of the augmentation factor and the multiplication of the generated images, new metrics - such as PSNR or SSIM - will be implemented to automatically evaluate image quality of whole datasets [16]. Other GAN implementations will be evaluated to assess the minimal size of their intermediate datasets and their optimal traditional Data Augmentation method.

## References


1. S. Vanderbeck, J. Bockhorst, et al., "Automatic classification of white regions in liver biopsies by supervised machine learning," Hum. pathol., vol. 45, no. 4, pp. 785-792, 2014
2. Tajbakhsh, Nima, et al. "Embracing imperfect datasets: A review of deep learning solutions for medical image segmentation." Medical Image Analysis 63 (2020): 101693.
3. C. Shorten, T.M. Khoshgoftaar, "A survey on Image Data Augmentation for Deep Learning,". J Big Data, vol. 6, p. 60, 2019
4. Frid-Adar, Maayan, et al. "Synthetic data augmentation using GAN for improved liver lesion classification." 2018 IEEE 15th international symposium on biomedical imaging (ISBI 2018). IEEE, 2018.
5. M.E. Vandenberghe, A.S. Herard, et al. "High-throughput 3D whole-brain quantitative histopathology in rodents", Scientific Reports, vol. 6, p.20958, 2016
6. A. Buslaev, V.I. Iglovikov, et al., "Albumentations: fast and flexible image augmentations," Information, vol.11, no.2, p. 25, 2020
7. S.T.M. Ataky, J. de Matos, et al. "Data Augmentation for Histopathological Images based on Gaussian-Laplacian Pyramid Blending," arXiv preprint, arXiv:2002.00072, 2020
8. Z. Hussain, F. Gimenez, et al., "Differential data augmentation techniques for medical imaging classification tasks," in AMIA Proc., AMIA, Washington, DC, USA, 2017, vol. 2017, p. 979
9. A. Fawzi, H. Samulowitz, et al., "Adaptive data augmentation for image classification", in International Conference on Image Processing, IEEE, Phoenix, AZ, USA, 2016 pp. 3688-3692
10. P.Y. Simard, D. Steinkraus, and J.C. Platt, "Best practices for convolutional neural networks applied to visual document analysis," in ICDAR, IEEE, Edinburgh, UK, 2003, vol. 3, no. 2003
11. D. Tellez, M. Balkenhol, et al., "Whole-Slide Mitosis Detection in HE Breast Histology Using PHH3 as a Reference to Train Distilled Stain-Invariant Convolutional Networks," IEEE Transactions on Medical Imaging, vol. 37, pp. 2126–2136, 2018
12. I. Goodfellow, J. Pouget-Abadie, et al., "Generative adversarial nets," in Proc. International Conference on Neural Information Processing Systems, MIT Press, Montreal, QC, Canada, 2014, vol. 2, pp. 2672–2680
13. A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks", arXiv preprint, arXiv:1511.06434, 2015
14. Neff, Thomas, et al. "Generative adversarial network based synthesis for supervised medical image segmentation." Proc. OAGM and ARW Joint Workshop. 2017.
15. O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," In International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, Munich, Germany, 2015, pp. 234-241
16. Hore, Alain, and Djemel Ziou. "Image quality metrics: PSNR vs. SSIM." 2010 20th international conference on pattern recognition. IEEE, 2010.