Fine-tuning Generative Adversarial Networks Using Metaheuristics: A Case Study on Barrett's Esophagus Identification

Luis A. Souza[1,2], Leandro A. Passos[3], Robert Mendel[2], Alanna Ebigbo[4], Andreas Probst[4], Helmut Messmann[4], Christoph Palm[2,5], João P. Papa[3]

[1]Department of Computing, São Carlos Federal University - UFSCar, Brazil
[2]Regensburg Medical Image Computing (ReMIC), Ostbayerische Technische Hochschule Regensburg (OTH Regensburg) - Germany
[3]Department of Computing, São Paulo State University - UNESP, Brazil
[4]Department of Gastroenterology, University Hospital Augsburg - Germany
[5]Regensburg Center of Health Sciences and Technology (RCHST), OTH Regensburg - Germany
`luis.souza@dc.ufscar.br`

Abstract. Barrett's esophagus denotes a disorder in the digestive system that affects the esophagus' mucosal cells, causing reflux, and showing potential convergence to esophageal adenocarcinoma if not treated in initial stages. Thus, fast and reliable computer-aided diagnosis becomes considerably welcome. Nevertheless, such approaches usually suffer from imbalanced datasets, which can be addressed through Generative Adversarial Networks (GANs). Such techniques generate realistic images based on observed samples, even though at the cost of a proper selection of its hyperparameters. Many works employed a class of nature-inspired algorithms called metaheuristics to tackle the problem considering distinct deep learning approaches. Therefore, this paper's main contribution is to introduce metaheuristic techniques to fine-tune GANs in the context of Barrett's esophagus identification, as well as to investigate the feasibility of generating high-quality synthetic images for early-cancer assisted identification.

1 Introduction

Barrett's esophagus (BE) is a dangerous condition in which the mucosal cells of the lower part of the esophagus changes due to chronic gastrointestinal reflux, and may progress into esophageal adenocarcinoma [1]. Computer-aided analysis of Barrett's esophagus and adenocarcinoma has been subjected to intensive research in the past years using both handcrafted features from endoscopic images, as well as the application of Convolutional Neural Networks (CNN) for automatic identification. Some examples can be observed in recent works of Souza Jr. et al. [1], van der Sommen [2] and Passos et al. [3].

A usual drawback, the limited amount of data, restricts the development and validation of more effective methods to detect early-stage illness in medical applications. Recently, such a bottleneck has been coped through data augmentation (DA) techniques. In this context, Generative Adversarial Networks (GAN) [4] have presented significant improvements in image generation, with highlights for medical imaging [5,6]. GAN's primary idea is to train a generator and a discriminator simultaneously, aiming to generate convincing and high-quality synthetic images.

One of the main hindrances regarding GANs and most modern deep learning approaches concerns a proper selection of their hyperparameters since they pose a significant influence in the model's final output. Several works addressed a similar problem through metaheuristic optimization techniques [3]. Metaheuristic approaches refer to stochastic nature-inspired methods that mimic some natural behavior observed in groups of animals, social conduct, among others, to solve complex problems. The paradigm obtained notorious popularity due to positive results in a wide variety of applications.

As far as we know, there is no work addressing the use of metaheuristic techniques to optimize the GAN hyperparameters itself. Therefore, the main contributions of this work are three-fold: (i) to introduce metaheuristic optimization algorithms in the context of GAN hyperparameter optimization; (ii) to investigate the feasibility of using GAN parameter optimization to generate high-quality synthetic images for further assisting the identification of Barrett's esophagus and adenocarcinoma; and (iii) to evaluate whether it makes sense to perform such parameter optimization of image generation for further data augmentation and classification purposes.

2 Materials and Methods

The GAN training procedure demands the user an appropriate selection of the network hyperparameters, which poses a far from straightforward task due to the context-dependence and the sensitivity related to the selected values. To cope with such an issue, this work proposes employing nature-inspired metaheuristic optimization techniques to fine-tune a set of five main hyperparameters $\theta = \{\eta, \beta_1, \mathit{ngf}, \mathit{ndf}, \mathit{batch\ size}\}$ considering a pre-defined range, described as follows: the learning rate $\eta \in [0.0001, 0.001]$, the Adam optimizer decay control $\beta_1 \in [0.002, 0.5]$, the $\mathit{ngf} \in [1, 128]$ and the $\mathit{ndf} \in [1, 128]$, which are related to the generator and discriminator feature map depths, respectively, and the $\mathit{batch\ size} \in [1, 128]$.

The main idea behind metaheuristic optimization techniques consists of stochastically initializing a set of random solutions, and iteratively evolving towards the solution whose decision variables best fit a target objective, i.e., minimizing the quadratic difference between generator and discriminator losses. The pipeline employed to perform GAN hyperparameter fine-tuning is depicted in Fig. 1. In a nutshell, the optimization technique selects the set of hyperparameters that minimize the loss function over the training set considering an

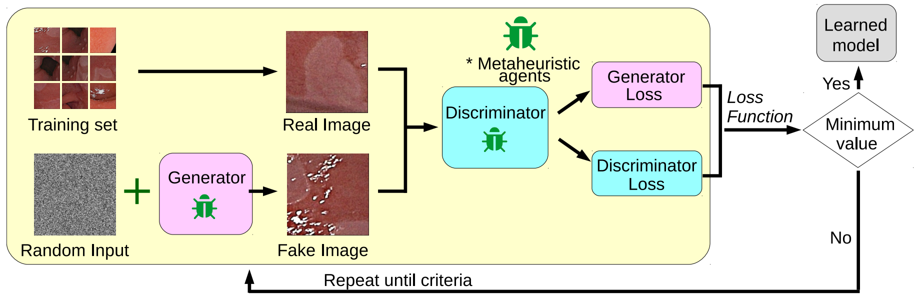


Fig. 1. Proposed approach to encode the decision variables of each optimization agent.

augmented dataset composed of endoscopy images and synthetic images generated during the process.

The following metaheuristic techniques from [3] are considered:

– BSA: Backtracking Search Optimization is an evolutionary algorithm that combines stored memories with crossover and mutation operations to generate new individuals.
– BSO: Brain Storm Optimization is a swarm-based optimization technique inspired by the creative human brainstorming process.
– FA: Firefly Algorithm tries to mimic the fireflies' behavior while searching for mating partners and preys.
– FPA: Flower Pollination Algorithm is a swarm-based optimization method that mimics the pollination process of flowering plants.
– JADE [3]: a differential evolution-based algorithm that implements the "DE/current-to-p-best" mutation strategy.

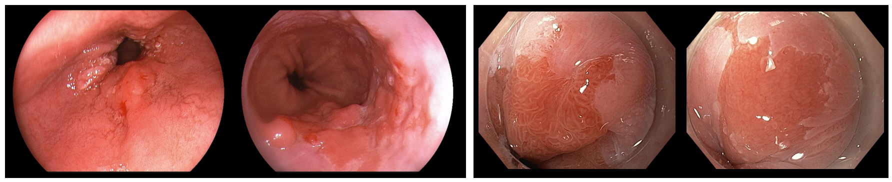
3 Results

This section briefly describes the datasets used in this work and the setup employed during the experiments.

3.1 Datasets

Two white-light endoscopic datasets were used for an in-depth analysis concerning the robustness of the proposed approach. The first one, provided at the "MICCAI 2015 EndoVis Challenge (MICCAI), comprises 100 lower esophagus endoscopic images captured from 39 individuals, 22 of them being diagnosed with Barrett's esophagus (BE), and 17 showing early-stage signs of esophageal adenocarcinoma (AD). Five different experts have individually delineated suspicious regions observed in the cancerous images.

The second dataset used for the experiments was provided by the Augsburg Hospital University and it is composed of 76 endoscopic images captured from different patients with BE (42 samples) and early AD (34 samples). The cancerous images were manually annotated by one expert. The annotations provided



(a) MICCAI (b) Augsburg

Fig. 2. MICCAI (a) and Augsburg (b) positive samples to AD and their respective delineations.

by the experts were considered for the patch label definition for both datasets. Fig. 2 (a) and (b) depicts some positive samples of the MICCAI and the Augsburg datasets.

3.2 Experimental Setup

Regarding the pre-processing step, the images were split into patches [1]. The idea is to cover the entire image with a sliding window of 200×200 pixels and overlapping of 50 pixels in horizontal and vertical directions. The label of each patch was based on the expert annotations of the full-images.

To obtain the best parameters for each dataset and class, the metaheuristic techniques were run for 40 epochs. For the data augmentation evaluation, experiments were conducted over $12,000$ epochs, generating 525 synthetic samples at every $2,000$ iterations, for each sample class (AD and BE) using the best parameters obtained in the metaheuristic experimental design. The output sample amount was related to computational limitations. We employed two different strategies to sample the synthetic images: (i) the last batch during learning, and (ii) the five last batches. The statical analysis was conducted using the Wilcoxon signed-rank test with confidence level of 5% [7].

After performing data augmentation, 80% of the new dataset was randomly selected for training purposes, and the remaining 20% was used for the testing. Such a partitioning was conducted over 20 runs for more robust evaluation. For the classification step, we employed two CNN architectures pre-trained with the ImageNet dataset: LeNet-5 and AlexNet, also running for $12,000$ epochs.

3.3 Optimization Results

The metaheuristic results for the hyperparameters fine-tuning can be observed in Table 1. The best results for MICCAI dataset (closest to 0) were obtained using BSA and FA, for AD and BE diagnosed patches, respectively, with values of 0.0033 and 0.0010. Regarding the Augsburg metaheuristic fine-tuning, best results were achieved, respectively, for AD and BE, using FA and BSA, with values of 0.0025 and 0.0011. Fine-tuned synthetic samples can be observed in Figure 3.

Table 1. Mean loss value and time consumption considering MICCAI and Augsburg datasets.

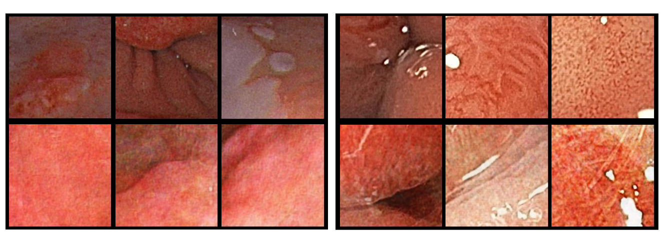
Dataset	Diagnosis	Metric	BSA	BSO	FA	FPA	JADE	RANDOM
MICCAI	AD	Loss	0.0033	0.0056	0.0046	0.0037	0.0057	0.0310
		Time(h)	2.9238	2.7872	3.7210	3.7254	3.4264	1.0706
	BE	Loss	0.0045	0.0029	0.0010	0.0011	0.0018	0.0034
		Time(h)	13.6447	15.1997	16.1333	14.6826	10.8849	4.8940
Augsburg	AD	Loss	0.0049	0.0074	0.0025	0.0140	0.0053	0.0129
		Time(h)	2.8345	2.5835	2.1745	4.0850	2.4363	0.6976
	BE	Loss	0.0011	0.0045	0.0036	0.0057	0.0105	0.0051
		Time(h)	8.6632	8.7128	8.7699	9.9796	9.3854	2.6029

3.4 Classification Results

Regarding the classification performed after the data augmentation step, one can observe the results in Table 2. Concerning MICCAI dataset, the best classification rates were obtained using BSA and FA data augmentation for AD and BE patches, respectively, with a value of 0.93 using LeNet-5 architecture and “5-last” augmentation protocol. For the Augsburg dataset classification, the best results were achieved, respectively, for AD and BE, using FA and BSA parameters, with an accuracy of 0.90 also using LeNet-5 and “5-last” augmentation protocol. The Wilcoxon test revealed statistical simmilarity for MICCAI classification results for both LeNet-5 and AlexNet architectures. The augmented dataset results using fine-tuned GAN outperformed the other experimental delineations.

4 Discussion

This paper dealt with computer-assisted Barrett’s esophagus and adenocarcinoma identification through GAN-fine-tuned data augmentation and CNNs as feature extractors. The GAN hyperparameter fine-tuning showed promising classification results after the data augmentation step, outperforming the classification rates of original and standard augmented datasets. Regarding the fine-



(a) MICCAI (b) Augsburg

Fig. 3. MICCAI (a) and Augsburg (b) dataset experiments using patches: original (top) and synthetic (bottom) images.

Table 2. Accuracy results considering MICCAI and Augsburg datasets taking different kinds of augmentation into account.

			No aug.		Standard augmentation		last GAN-augmentation		5-last GAN-augmentation	
Dataset	AD Parameters	BE Parameters	LeNet-5	AlexNet	LeNet-5	AlexNet	LeNet-5	AlexNet	LeNet-5	AlexNet
MICCAI	BSA	FA	0.81	0.82	0.83	0.83	0.89	0.88	⋆ 0.93	0.91
Augsburg	FA	BSA	0.73	0.73	0.75	0.83	0.88	0.87	0.90	0.86

tuning process, FA and BSA provided the best results for both datasets, suggesting the best performance for BE and adenocarcinoma high-quality and trustworthy sample generation for classification purposes. Such procedures provided improvements compared to previous works, suggesting the importance of enough data and CNN generalization ability to deal with BE and adenocarcinoma distinction problem. In regard of future works, we intend to evaluate fine-tuned GAN to generate full-image samples, aiming to reinforce the impact of the best hyperparameter selection in the synthetic image generation quality.

5 Acknowledgements

The authors thank Capes/Alexander von Humboldt Foundation grant number BEX 0581-16-0, CNPq grants 306166/2014-3 and 307066/2017-7, as well as FAPESP grants 2013/07375-0, 2014/12236-1, and 2016/19403-6, 2017/04847-9 and 2019/08605-5.

References

1. Souza Jr LA, Palm C, Mendel R, et al. A survey on Barrett's esophagus analysis using machine learning. Computers in Biology and Medicine. 2018;96:203 – 213.
2. van der Sommen F, Zinger S, Curvers WL, et al. Computer-aided detection of early neoplastic lesions in Barret's esophagus. Endoscopy. 2016;68:617–624.
3. Passos LA, Souza Jr LA, Mendel R, et al. Barrett's esophagus analysis using infinity Restricted Boltzmann Machines. Journal of Visual Communication and Image Representation. 2019;59:475 – 485.
4. Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative Adversarial Nets. In: Ghahramani Z, Welling M, Cortes C, et al., editors. Advances in Neural Information Processing Systems 27. Curran Associates, Inc.; 2014. p. 2672–2680.
5. Souza Jr LA, Passos LA, Mendel R, et al. Assisting Barrett's Esophagus Identification Using Endoscopic Data Augmentation Based on Generative Adversarial Networks. Computers in Biology and Medicine. 2020; p. 104029.
6. Han C, Murao K, Satoh S, et al. Learning More with Less: GAN-based Medical Image Augmentation. CoRR. 2019;abs/1904.00838.
7. Wilcoxon F. Individual Comparisons by Ranking Methods. Biometrics Bulletin. 1945;1(6):80–83.