Performance analysis of GAN approaches in the portable chest X-ray synthetic image generation for COVID-19 screening

Daniel I. Morís[1,2][0000−0002−5239−0421], Mateo Gende[1,2][0000−0003−1686−7189], Joaquim de Moura[1,2,✉][0000−0002−2050−3786], Jorge Novo[1,2][0000−0002−0125−3064], and Marcos Ortega[1,2][0000−0002−2798−0788]
{daniel.iglesias.moris, m.gende, joaquim.demoura, jnovo, mortega}@udc.es

[1] Centro de Investigación CITIC, Universidade da Coruña, Campus de Elviña, s/n, 15071 A Coruña, Spain
[2] Grupo VARPA, Instituto de Investigación Biomédica de A Coruña (INIBIC), Universidade da Coruña, Xubias de Arriba, 84, 15006 A Coruña, Spain

Abstract. COVID-19 mainly affects lung tissues, aspect that makes chest X-ray imaging useful to visualize this damage. In the context of the global pandemic, portable devices are advantageous for the daily practice. Furthermore, Computer-aided Diagnosis systems developed with Deep Learning algorithms can support the clinicians while making decisions. However, data scarcity is an issue that hinders this process. Thus, in this work, we propose the performance analysis of 3 different state-of-the-art Generative Adversarial Networks (GAN) approaches that are used for synthetic image generation to improve the task of automatic COVID-19 screening using chest X-ray images provided by portable devices. Particularly, the results demonstrate a significant improvement in terms of accuracy, that raises 5.28% using the images generated by the best image translation model.

Keywords: Computer-aided diagnosis · Portable Chest X-ray · COVID-19 · Deep Learning · Synthetic image generation.

1 Introduction

The COVID-19 represents a challenge for the healthcare services since its emergence at the end of the year 2019 in Wuhan, Hubei, China [1]. This disease is caused by the coronavirus SARS-CoV-2, an extremely contagious pathogen, that was rapidly spread worldwide, aspect that forced the World Health Organization to declare this pathology as a global pandemic in March 2020. The main diagnostic tool to confirm the infection of COVID-19 is the RT-PCR test, which is considered as the gold-standard [2]. However, this tool is limited to provide a binary diagnosis and, therefore, it is unable to quantify some important aspects as the disease severity on each patient. As the COVID-19 mainly affects the

respiratory tissues, the chest X-ray imaging modality is useful to visualize the extent of the pathology [3]. To perform the chest X-ray captures, there are 2 main types of radiological devices: fixed and portable. In the current situation of global pandemic, the American College of Radiology (ACR) recommends the use of portable devices as they are easier to decontaminate and more versatile [4]. However, despite this recommendation, they provide a lower quality and level of detail in their captures. As chest X-ray visualization is a complex, tedious and subjective task that must be performed by expert clinicians, the lower quality of the provided images makes this process even more challenging. Additionally, given the situation that the healthcare services are suffering due to the pandemic, health workers are experiencing a great amount of workload, context where the Computer-Aided Diagnosis (CAD) systems can be very useful to help them to take decisions [5]. In the last years, the medical imaging analysis domain has been supported by the computer vision and machine learning techniques. Particularly, the deep learning, a subarea of machine learning, has demonstrated its great performance dealing with this kind of problems [6].

In the context of COVID-19 diagnosis using chest X-ray imaging, some works have addressed the problem of the automatic COVID-19 screening with the support of deep learning strategies. As this is a relevant issue, given the current critical health situation, many efforts have been done to tackle it, both using datasets composed of captures provided by fixed devices (as reference, [7], [8], [9]) as well as by portable devices (as reference, [10], [11], [12], [13], [14]). However, despite these works provide robust and relevant results, deep models need a great amount of labelled data to be trained even though it exists a data scarcity problem in domains such as biomedical imaging. More precisely, in the scope of COVID-19, its recent emergence makes the information gathering even more challenging, aspect that implies an accused data scarcity.

To mitigate data scarcity, synthetic image generation has emerged during the last years as a powerful data augmentation strategy [15], which is often performed using a Generative Adversarial Network (GAN) architecture [16]. Moreover, there are several implementations of GANs that aim at performing the task of image translation, such as the Conditional GANs (Pix2Pix), the Cycle-Consistent Adversarial Networks (CycleGAN) or the Contrastive Unpaired Translation (CUT) models. This kind of models are able to convert images from a certain scenario to another different scenario and vice versa. Particularly, in the field of the COVID-19 chest X-ray image generation, the work from Morís *et al.* [13] addressed the problem of generating synthetic images using a CycleGAN and a dataset of portable chest X-ray images in the field of the COVID-19 screening. Lately, a work from the same authors [14] demonstrated that adding the novel set of generated images to the original dataset to increase its dimensionality can improve the performance of the automatic COVID-19 screening model. Nevertheless, even though the obtained satisfactory results, the reference works only addressed part of the problem, as they only use the CycleGAN, without analyzing the performance of other alternative image translation architectures.

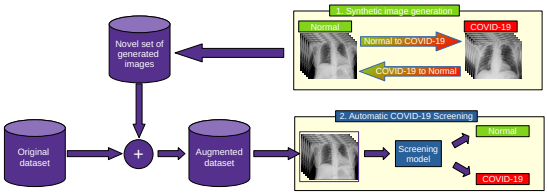

Fig. 1. Schematic description of the proposed methodology, that can be divided in 2 parts: synthetic image generation and automatic COVID-19 screening.

In this work, we propose the performance analysis of 3 different state-of-the-art GAN approaches for portable chest X-ray synthetic image generation, one of them specifically tailored to work with paired datasets (Pix2Pix) and 2 designed to work with unpaired datasets (CycleGAN and CUT). On the second part of the work, the novel set of generated image is used to improve the performance of an automatic COVID-19 screening task, evaluating the synthetic image generation with a finalist scenario. For these purposes, we use a dataset divided in 2 different classes, Normal and COVID-19. Overall, this proposal can be seen as a relevant contribution to the clinical community, as it solves a critical problem given the current situation of world health crisis.

2 Materials and Methods

An schematic representation of the methodological process is depicted in Fig. 1. There, it can be seen that, first of all, the synthetic image generation process is performed in order to obtain a novel set of images. Then, this set of images is added to the original dataset in order to increase its dimensionality. Finally, the automatic COVID-19 screening model is trained using this augmented dataset. However, it is important to note that the training set is composed of both original and generated images but the test set is only composed of original images, in order to perform a comparison with the baseline on equal conditions.

2.1 Dataset

The portable chest X-ray imaging dataset was provided by the Complexo Hospitalario Universitario de A Coruña (CHUAC) and specifically designed for the purposes of this work, having 797 Normal control cases (without evidences of pulmonary affectation that could present abnormalities in other parts of the visualized region) and 2,071 genuine COVID-19 cases. The captures were performed using 2 portable chest X-ray devices: Agfa dr100E and Optima Rx200.

2.2 Synthetic image generation

For the synthetic image generation, 2 different pathways are followed: the one that converts from Normal to COVID-19 and the one that converts from COVID-19 to Normal. In the same way, both pathways are followed by the 3 considered

image translation models, Pix2Pix [17], CycleGAN [18] and CUT [19], as this kind of image translation architectures have demonstrated satisfactory results for similar tasks of image generation [13]. For these 3 architectures, the same training details apply. First of all, for the training process, all the images of the dataset are used. With regards to the generator models, the selected configuration uses a ResNet with 6 residual blocks as it demonstrated to obtain robust and relevant results in similar problems [13], [14]. On the other hand, all the input images are resized to 512×512 pixels. Moreover, all the models are trained during 200 epochs using the Adam optimizer algorithm with a constant learning rate of $\alpha = 0.0002$ and a mini-batch size of 1.

2.3 Automatic COVID-19 screening

To understand the impact of the novel set of synthetic images on the model effectiveness, we performed an automatic COVID-19 screening. For these purposes, we used a DenseNet-161 as the deep network architecture due to its demonstrated capability dealing with a similar context [7]. The training process is set to 200 epochs and all the images are resized to a resolution of 512×512 pixels. In terms of dataset splitting, the 60% of the samples were used for training, the 20% for validation and the remaining 20% for test. Furthermore, the cross-entropy was used as loss function. In this case, the Stochastic Gradient Descent (SGD) [20] was selected to optimize the weights of the network, using a constant learning rate of $\alpha = 0.01$, a first-order momentum of 0.9 and a mini-batch size of 4. Finally, other relevant detail is that, to have a better understanding of the network behaviour, the training process is repeated 5 times, performing a different random dataset splitting for each repetition.

3 Results and discussion

The results of the test set for the automatic COVID-19 screening are depicted in Table 1 where the baseline approach refers to the case of training with only original images, the approach 1 refers to the data augmentation approach that uses the images generated by the Pix2Pix model, the approach 2 refers to the case of using the images generated by the CycleGAN model and the approach 3 refers to the case of using the images generated by the CUT model. Overall, as expected, the approach 1 has a performance drop in comparison with the

Table 1. Results from the test set for the baseline approach and the 3 approaches of data augmentation.

	Baseline	Approach 1 (Pix2Pix)	Approach 2 (CycleGAN)	Approach 3 (CUT)
Accuracy	93.33 % ± 4.68 %	90.69 % ± 5.00 %	97.43 % ± 2.41 %	98.61 % ± 1.24
Precision	92.64 % ± 5.30 %	94.52 % ± 4.67 %	98.55 % ± 1.18 %	98.47 % ± 1.03 %
Recall	94.30 % ± 4.76 %	86.94 % ± 11.47 %	96.25 % ± 3.79 %	98.75 % ± 1.48 %
Specificity	92.36 % ± 5.74 %	94.44 % ± 4.99 %	98.61 % ± 1.08 %	98.47 % ± 1.02 %
F1-Score	93.41 % ± 4.60 %	89.96 % ± 5.96 %	97.63 % ± 2.52 %	98.61 % ± 1.24 %
AUC	0.9748 ± 0.0230	0.9626 ± 0.0247	0.9891 ± 0.0109	0.9962 ± 0.0040

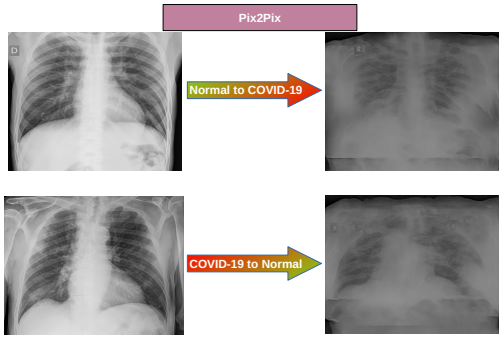

Fig. 2. Examples of synthetic images generated by Pix2Pix for both pathways.

baseline due to the bad quality of the images generated by Pix2Pix. However, it is remarkable that this drop is smaller that could be expected. In fact, the mean accuracy drops from 93.33% to 90.69% and the standard deviation remains very similar and both the recall and F1-Score experience an important drop, being the mean value lower than 90%. On its way, the mean AUC value drops from 0.9748 to 0.9626. The most remarkable performance drop can be seen in the case of recall, where the standard deviation rises from 4.76% to 11.47%. However, it is also remarkable that there is an improvement in terms of precision and specificity. With regards to the other 2 approaches of data augmentation, there is an improvement for all the metrics in comparison with the baseline. Furthermore, the approach 3 obtains a greater performance in comparison with the approach 2, thus concluding that the CUT model generates more useful synthetic images than CycleGAN. Particularly, in the case of the CycleGAN, the mean accuracy improves from 93.33% to 97.43% and to 98.61% in the case of the CUT. Something similar happens with the value of AUC, with an improvement from 0.9748 to 0.9891 in the case of the CycleGAN and to 0.9962 in the case of the CUT. Other point that must be noticed is that not only the mean values improve globally, but also the standard deviation values. In fact, the standard deviation of the accuracy lowers from 4.68% to 2.41% in the case of the CycleGAN and to 1.24% in the case of the CUT. Moreover, for the AUC, this improvement means a standard deviation drop from 0.0230 to 0.0109 with the CycleGAN and to 0.0040 with the CUT.

On the other hand, some examples of synthetic images generated by the Pix2Pix model are depicted in Fig. 2. There, it can be seen that the model generates images with a low quality whose appearance is far from a real chest X-ray image. This can be explained because Pix2Pix is designed to work with paired datasets, while in this work it is necessary to use a method able to deal with unpaired datasets. On the other hand, both the CycleGAN and the CUT are designed to work with this kind of datasets, aspect that is reflected on the quality

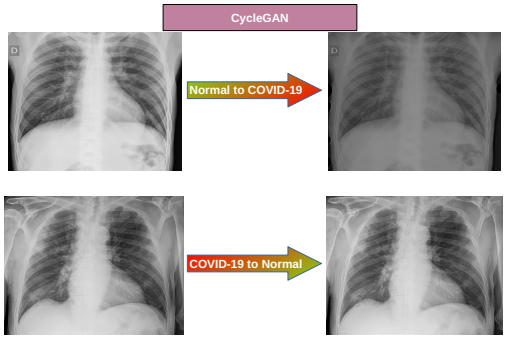

Fig. 3. Examples of images generated by the CycleGAN model for both scenarios.

of the generated images. Some representative examples of both approaches of image generation can be seen in Fig. 3 and Fig. 4 in the case of the CycleGAN and the CUT, respectively. There, it can be clearly seen a notable difference in contrast with those images generated by Pix2Pix. In this case, the obtained images have a realistic appearance, coherent with a real chest X-ray capture, that show well-synthesized differences in pulmonary regions.

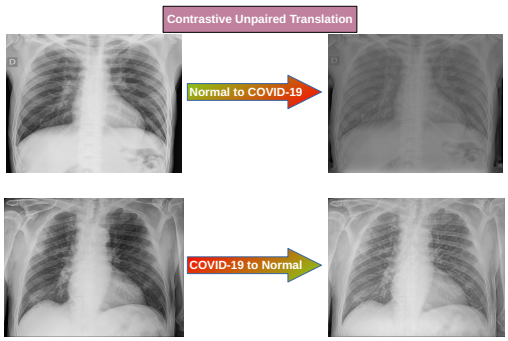

Fig. 4. Examples of synthetic images generated by the CUT models in both pathways.

4 Conclusion

In this work, we propose the analysis of 3 different state-of-the-art GAN architectures for synthetic image generation (Pix2Pix, CycleGAN and CUT) in the

context of an automatic COVID-19 screening using portable chest X-ray images. Results demonstrate that the Pix2Pix is inadequate for the problem, as it is only capable of dealing with paired datasets, while the dataset used for the purposes of this work contains unpaired data. This leads to a poor performance of the model, generating considerably different images in comparison with a real chest X-ray image. The opposite results can be seen in the case of the CycleGAN and the CUT. In fact, the generated synthetic images have a good quality and a realistic appearance. In the same way, the models are able to generate well-synthesized differences in the pulmonary regions, aspect that is clinically relevant. These conclusions apply in the same way for the automatic COVID-19 screening performance. As expected, the data augmentation using the images generated by Pix2Pix implies a performance drop, while both the images generated by the CycleGAN and the CUT are able to improve the evaluation metrics values. Overall, the best performance is achieved using the images generated by the CUT model, obtaining a $98.61\% \pm 1.24\%$ of accuracy and a 0.9962 ± 0.0040 of AUC.

Acknowledgements

This research was funded by Instituto de Salud Carlos III, Government of Spain, DTS18/00136 research project; Ministerio de Ciencia e Innovación y Universidades, Government of Spain, RTI2018-095894-B-I00 research project; Ministerio de Ciencia e Innovación, Government of Spain through the research project with reference PID2019-108435RB-I00; Consellería de Cultura, Educación e Universidade, Xunta de Galicia through the predoctoral grant contracts ref. ED481A 2021/196 and ED481A 2021/161, respectively; and Grupos de Referencia Competitiva, grant ref. ED431C 2020/24; Axencia Galega de Innovación (GAIN), Xunta de Galicia, grant ref. IN845D 2020/38; CITIC, Centro de Investigación de Galicia ref. ED431G 2019/01, receives financial support from Consellería de Educación, Universidade e Formación Profesional, Xunta de Galicia, through the ERDF (80%) and Secretaría Xeral de Universidades (20%).

References

1. M. Ciotti, M. Ciccozzi, A. Terrinoni, W.-C. Jiang, C.-B. Wang, and S. Bernardini, "The COVID-19 pandemic," *Critical reviews in clinical laboratory sciences*, vol. 57, no. 6, pp. 365–388, 2020.
2. J. V. Waller, P. Kaur, A. Tucker, K. K. Lin, M. J. Diaz, T. S. Henry, and M. Hope, "Diagnostic tools for coronavirus disease (COVID-19): comparing CT and RT-PCR viral nucleic acid testing," *American Journal of Roentgenology*, vol. 215, no. 4, pp. 834–838, 2020.
3. A. Jacobi, M. Chung, A. Bernheim, and C. Eber, "Portable chest X-ray in coronavirus disease-19 (COVID-19): A pictorial review," *Clinical imaging*, vol. 64, pp. 35–42, 2020.

4. S. Kooraki, M. Hosseiny, L. Myers, and A. Gholamrezanezhad, "Coronavirus (COVID-19) outbreak: what the department of radiology should know," *Journal of the American college of radiology*, vol. 17, no. 4, pp. 447–451, 2020.
5. B. Halalli and A. Makandar, "Computer aided diagnosis-medical image analysis techniques," *Breast imaging*, vol. 85, 2018.
6. K. Suzuki, "Overview of deep learning in medical imaging," *Radiological physics and technology*, vol. 10, no. 3, pp. 257–273, 2017.
7. J. de Moura, J. Novo, and M. Ortega, "Fully automatic deep convolutional approaches for the analysis of COVID-19 using chest X-ray images," *Applied Soft Computing*, vol. 115, p. 108190, 2022.
8. A. Narin, C. Kaya, and Z. Pamuk, "Automatic detection of coronavirus disease (covid-19) using x-ray images and deep convolutional neural networks," *Pattern Analysis and Applications*, vol. 24, no. 3, pp. 1207–1220, 2021.
9. L. Alvarez-Rodríguez, J. de Moura, J. Novo, and M. Ortega, "Does imbalance in chest X-ray datasets produce biased deep learning approaches for COVID-19 screening?," 2021.
10. P. Vidal, J. d. Moura, J. Novo, and M. Ortega, "Pulmonary-Restricted COVID-19 Informative Visual Screening Using Chest X-ray Images from Portable Devices," in *International Conference on Image Analysis and Processing*, pp. 1–8, Springer, 2022.
11. J. De Moura, L. R. García, P. F. L. Vidal, M. Cruz, L. A. López, E. C. Lopez, J. Novo, and M. Ortega, "Deep convolutional approaches for the analysis of covid-19 using chest x-ray images from portable devices," *IEEE Access*, vol. 8, pp. 195594–195607, 2020.
12. D. I. Morıs, J. de Moura, J. Novo, and M. Ortega, "Comprehensive Analysis of the Screening of COVID-19 Approaches in Chest X-ray Images from Portable Devices," in *European Symposium on Artificial Neural Networks*, pp. 1–6, ESANN 2021, 2021.
13. D. I. Morís, J. de Moura, J. Novo, and M. Ortega, "Cycle Generative Adversarial Network Approaches to Produce Novel Portable Chest X-Rays Images for Covid-19 Diagnosis," in *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1060–1064, 2021.
14. D. I. Morís, J. J. de Moura Ramos, J. N. Buján, and M. O. Hortas, "Data augmentation approaches using cycle-consistent adversarial networks for improving COVID-19 screening in portable chest X-ray images," *Expert Systems with Applications*, vol. 185, p. 115681, 2021.
15. F. H. K. d. S. Tanaka and C. Aranha, "Data augmentation using GANs," *arXiv preprint arXiv:1904.09135*, 2019.
16. A. Creswell, T. White, V. Dumoulin, K. Arulkumaran, B. Sengupta, and A. A. Bharath, "Generative adversarial networks: An overview," *IEEE Signal Processing Magazine*, vol. 35, no. 1, pp. 53–65, 2018.
17. P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1125–1134, 2017.
18. J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proceedings of the IEEE international conference on computer vision*, pp. 2223–2232, 2017.
19. T. Park, A. A. Efros, R. Zhang, and J.-Y. Zhu, "Contrastive learning for unpaired image-to-image translation," in *European Conference on Computer Vision*, pp. 319–345, Springer, 2020.
20. N. Ketkar, "Stochastic gradient descent," pp. 113–132, 2017.