

Proceeding Paper

Portable Chest X-ray Synthetic Image Generation for the COVID-19 Screening [†]

Daniel I. Morís ^{1,2,*} , Joaquim de Moura ^{1,2} , Jorge Novo ^{1,2}  and Marcos Ortega ^{1,2} 

¹ Centro de Investigación CITIC, Universidade da Coruña, 15071 A Coruña, Spain; joaquim.demoura@udc.es (J.d.M.); jnovo@udc.es (J.N.); mortega@udc.es (M.O.)

² Grupo Varpa, Instituto de Investigación Biomédica de A Coruña (INIBIC), Universidade da Coruña, 15006 A Coruña, Spain

* Correspondence: daniel.iglesias.moris@udc.es; Tel.: +34-981-167-000 (ext. 1330)

[†] Presented at the 4th XoveTIC Conference, A Coruña, Spain, 7–8 October 2021.

Abstract: The global pandemic of COVID-19 raises the importance of having fast and reliable methods to perform an early detection and to visualize the evolution of the disease in every patient, which can be assessed with chest X-ray imaging. Moreover, in order to reduce the risk of cross contamination, radiologists are asked to prioritize the use of portable chest X-ray devices that provide a lower quality and lower level of detail in comparison with the fixed machinery. In this context, computer-aided diagnosis systems are very useful. During the last years, for the case of medical imaging, they are widely developed using deep learning strategies. However, there is a lack of sufficient representative datasets of the COVID-19 affection, which are critical for supervised learning when training deep models. In this work, we propose a fully automatic method to artificially increase the size of an original portable chest X-ray imaging dataset that was specifically designed for the COVID-19 diagnosis, which can be developed in a non-supervised manner and without requiring paired data. The results demonstrate that the method is able to perform a reliable screening despite all the problems associated with images provided by portable devices, providing an overall accuracy of 92.50%.

Keywords: COVID-19; portable chest X-ray images; oversampling; CycleGAN; deep learning



Citation: Morís, D.I.; de Moura, J.; Novo, J.; Ortega, M. Portable Chest X-ray Synthetic Image Generation for the COVID-19 Screening. *Eng. Proc.* **2021**, *7*, 6. <https://doi.org/10.3390/engproc2021007006>

Academic Editors: Marco A. González, Javier Pereira and Manuel G. Penedo

Published: 28 September 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

COVID-19, declared as a global pandemic by the World Health Organization (WHO) in March 2020, mainly affects the respiratory tissues [1]. Chest X-ray imaging plays an important role in supporting the screening and early detection of the disease. In this context, radiologists are asked to prioritize the use of portable chest X-ray devices that are important to reduce the risk of cross contamination [2]. However, these devices provide a lower quality and a lower level of detail in comparison with fixed machinery [3]. In this critical scenario, computer-aided diagnosis (CAD) systems can be very useful for clinical practice. During recent years, in the scope of biomedical imaging, these diagnostic systems were usually developed using computer vision techniques as well as machine learning techniques and, specifically, deep learning strategies, which have increased their importance. However, in the context of supervised learning, deep learning models require a great amount of labeled data to be trained.

Regarding medical imaging, data scarcity is an aspect to take into account as, in many occasions, it critically affects the amount of labeled data. One of the ways to overcome data scarcity is to generate synthetic images with several network architectures, as is the case of many variants of Generative Adversarial Networks (GANs) [4]. One example of this kind of GAN model is the CycleGAN, a model that is able to translate images from a certain scenario to another different scenario.

In this work, due to the low availability of samples that show COVID-19 affection, we present novel approaches to artificially increase the size of a portable chest X-ray image dataset to diagnose COVID-19, combining three different and complementary CycleGAN architectures to perform an oversampling using a non-supervised strategy that can be performed without paired data.

2. Methodology

Thus, the presented methodology is divided in 2 different parts. The first part performs the synthetic image generation. The second part uses the novel set of generated images in order to augment the dimensionality of the original dataset, which is proven in a COVID-19 screening scenario.

2.1. Approaches for Data Augmentation

In order to increase the size of the original chest X-ray dataset, we considered 3 different complementary scenarios, which correspond to all the possible combinations given the classes of the dataset. For the first scenario, normal vs. pathological, normal samples are translated to their pathological representation and vice versa. For the second scenario, normal vs. COVID-19, normal samples are converted to their hypothetical representation showing COVID-19 affection and vice versa, and for the third scenario, pathological vs. COVID-19, we perform the same task as in the previous cases but to convert pathological samples to COVID-19 and vice versa. It is important to remark that all the images from the original dataset are used to train the CycleGAN model [5].

2.2. Approaches for Screening Tasks

For this second stage, we assess the degree of separability among the generated images and the suitability of the novel set of generated synthetic images, with the oversampled dataset. We used a Dense Convolutional Network Architecture (DenseNet) [6] model, pretrained on the ImageNet dataset, with the same training details as stated in [7,8] due to their suitability to this particular problem.

3. Results and Conclusions

The chest X-ray image dataset was provided by the Radiology Service of the Complejo Hospitalario Universitario de A Coruña (CHUAC) and is composed of 600 patients that were divided into 3 different classes [9], having 200 normal cases (i.e., from patients without evidence of pulmonary pathologies), 200 pathological cases (i.e., from patients with pulmonary pathologies other than COVID-19) and 200 COVID-19 genuine cases.

In order to demonstrate the separability and the suitability of the generated synthetic images, we conducted 4 different experiments, where the first 3 correspond to the separability among the generated images and the fourth experiment corresponds to the suitability of the novel set of generated images, evaluating the screening using the oversampled dataset. The first 3 experiments demonstrate that there is a proper separability among generated images for the 3 possible scenarios. For the fourth experiment, the model obtained a global accuracy of 0.9250 for the test. Additionally, Figure 1 shows the performance of the model for the test set for all the 4 experiments, obtaining remarkable correct classification ratios in every case.

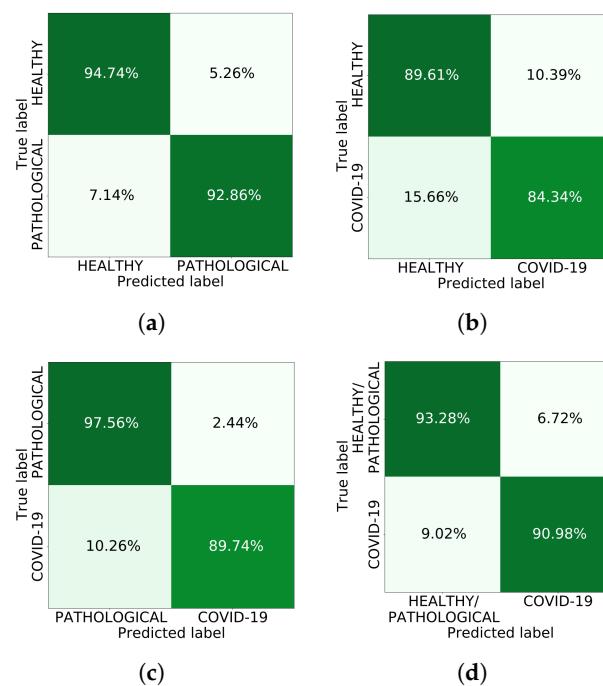


Figure 1. Confusion matrices for the four conducted experiments. (a) 1st experiment (Healthy vs. Pathological); (b) 2nd experiment (Healthy vs. COVID-19); (c) 3rd experiment (Pathological vs. COVID-19); (d) 4th experiment (Healthy and Pathological vs. COVID-19).

Author Contributions: Conceptualization, D.I.M., J.d.M., J.N. and M.O.; methodology, D.I.M., J.d.M., J.N. and M.O.; software, D.I.M. and J.d.M.; validation, J.d.M., J.N. and M.O.; investigation, J.N. and M.O.; data curation, J.N. and M.O.; writing—original draft preparation, D.I.M.; writing—review and editing, D.I.M., J.d.M., J.N. and M.O.; visualization, D.I.M.; supervision, J.d.M., J.N. and M.O.; project administration, J.N. and M.O.; funding acquisition, M.O. All authors have read and agreed to the published version of the manuscript.

Funding: 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 and postdoctoral grant contracts ref. ED481A 2021/196 and ED481B 2021/059, 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%).

Institutional Review Board Statement: The study was approved by the Ethics Review Board y Data Management Technical Commission of Galician Health Ministry for High Impact studies with protocol code 2020-007.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Pollard, C.A.; Morran, M.P.; Nestor-Kalinoski, A.L. The COVID-19 pandemic: A global health crisis. *Physiol. Genom.* **2020**, *52*, 549–557. [[CrossRef](#)] [[PubMed](#)]
- Kooraki, S.; Hosseiny, M.; Myers, L.; Gholamrezanezhad, A. Coronavirus (COVID-19) outbreak: What the department of radiology should know. *J. Am. Coll. Radiol.* **2020**, *17*, 447–451. [[CrossRef](#)] [[PubMed](#)]
- Vidal, P.L.; de Moura, J.; Novo, J.; Ortega, M. Multi-stage transfer learning for lung segmentation using portable X-ray devices for patients with COVID-19. *Expert Syst. Appl.* **2021**, *173*, 114677. [[CrossRef](#)] [[PubMed](#)]
- Creswell, A.; White, T.; Dumoulin, V.; Arulkumaran, K.; Sengupta, B.; Bharath, A.A. Generative adversarial networks: An overview. *IEEE Signal Process. Mag.* **2018**, *35*, 53–65. [[CrossRef](#)]

5. Zhu, J.Y.; Park, T.; Isola, P.; Efros, A.A. Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017. [[CrossRef](#)]
6. Huang, G.; Liu, Z.; van der Maaten, L.; Weinberger, K.Q. Densely Connected Convolutional Networks. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017. [[CrossRef](#)]
7. de Moura, J.; Novo, J.; Ortega, M. Fully automatic deep convolutional approaches for the analysis of Covid-19 using chest X-ray images. *medRxiv* **2020**. [[CrossRef](#)]
8. Morís, D.I.; de Moura, J.; Novo, J.; Ortega, M. Cycle Generative Adversarial Network Approaches to Produce Novel Portable Chest X-Rays Images for Covid-19 Diagnosis. In Proceedings of the ICASSP 2021—2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Toronto, ON, Canada, 6–11 June 2021; pp. 1060–1064. [[CrossRef](#)]
9. De Moura, J.; García, L.R.; Vidal, P.F.L.; Cruz, M.; López, L.A.; Lopez, E.C.; Novo, J.; Ortega, M. Deep convolutional approaches for the analysis of covid-19 using chest x-ray images from portable devices. *IEEE Access* **2020**, *8*, 195594–195607. [[CrossRef](#)]