Synchronization in Retrospective Respiratory Gating using Deep Learning

Fernando Bayo[1], Claire J. Brooks[1], David Pérez-Benito[1], Patricio López-Expósito[1], Juan José Vaquero[1,2]

[1] Departamento de Bioingeniería, Universidad Carlos III de Madrid, Madrid, Spain
[2] Instituto de Investigación Sanitaria Gregorio Marañón, Madrid, Spain

Abstract

Breathing motion introduces artifacts during CT acquisition, what affect the quality and subsequent reconstruction of the images. This study aims to reduce artefact in CT images using deep learning techniques. Specifically, we propose the implementation of an autoencoder based on convolutional neural networks. Once the model was trained, we employed a morphing technique to generate new images with reduced respiratory motion. By analyzing the respiratory signal, we classified the different images into phases and selected those most suitable for correction. Subsequently, we applied the described method, obtaining a more homogeneous data set. The results demonstrate a significant reduction in motion when comparing intensity changes within the regions most affected by motion. Thus, we validated the efficacy of the proposed approach to mitigate breathing-induced artifacts. The application of artificial intelligence (AI) in this field represents a significant advance. This study provides promising initial results and opens up new possibilities for research and development. By complementing existing techniques, AI offers enhanced motion reduction capabilities, thus improving the quality of CT images. The potential for future advances in this direction are substantial, promising further reduction in respiratory motion artifacts and improvement to the overall quality of CT images.

1. Introduction

Tuberculosis remains a threatening global health problem, causing 1.6 million deaths in 2021 [1]. Mouse models aid in the understanding of tuberculosis due to their similarity to human infection and easy manipulation, but respiratory motion in CT images poses difficulties in the form of artifacts that hinder accurate visualization of lung lesions due to diaphragm motion and overlapping structures. Techniques like breath-holding and retrospective respiratory gating have been devised to mitigate respiratory motion artifacts in CT scans. While the latter involves selecting images from a particular phase of the respiratory cycle, the variability in a subject's breathing pattern can lead to missing images for specific phases. To overcome this limitation, our project introduces a novel AI-based solution that ensures the availability of images for each projection.

2. Materials and Methods

2.1. Algorithm

The approach proposed was based on the use of AI techniques to address the problem of motion introduced by respiration in CT images. The main objective was to replace images affected by excessive motion with images artificially corrected to remove the former. The first step consisted of extracting the respiratory signal from each image to process and identify the level of motion introduced. With this information, images with moderate motion were selected to be corrected by generating new images that resemble the target data set with a lower level of motion. Those with excessive motion were discarded. To perform the correction, an image generating model based on deep learning techniques has been used. This model allowed the creation of a new image by combining two input images, influenced by both inputs in different degrees as established. By applying this morphing process using AI, it was expected to obtain corrected images that were more similar to the desired data set, which improved the quality and usefulness of CT images by reducing motion artifacts caused by breathing. The pipeline is illustrated in Figure 1.

Figure 1: Process pipeline design.

2.2. Data set

Data acquisition was performed in cine mode, using the "step and shoot" technique, in which the X-ray tube was rotated around two subjects in 720 steps covering a 360° rotation. As a result, data sets consisted of 720 projections, each containing a total of 16 frames with a size of 570x518 pixels. While first data set was used to train the model, the second was used for test tasks.

2.3. Respiratory Signal

Since we were acquiring repeated axial CT images during a certain period of time for each position, it was possible to visualize the movement derived from the animal's breathing in the acquired images. However, if we simultaneously displayed all the frames contained in a single projection, it was practically impossible to perceive such motion with the naked eye. This is the case represented in Figure 2, in which the 16 frames contained in projection 0 are shown.

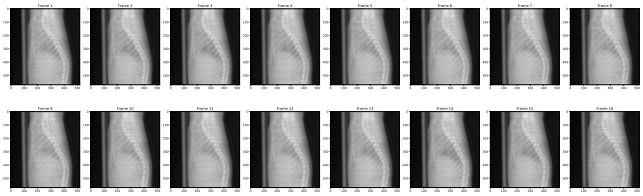

Figure 2: All 16 frames from projection '0'.

ISBN: 978-84-17853-76-1

To quantify the movement we computed the image resulting from averaging all the frames for the same projection, and we subtracted to each individually frame the average image [4].
The respiratory phase corresponding to each frame, whether inspiration or expiration, was easy to distinguish by looking at the signal intensity of the aforementioned area. This was the case of Figure 4 in which the average image of all frames and the subtraction between both was represented, resulting in another image with a darker region at the height of the diaphragm in which negative intensity values predominate. This indicated that the image corresponded to the inspiration phase. If the opposite case was observed, in which the area was highlighted in white, we would have been looking at an image related to the expiration phase.

Once the identification of respiratory phases has been defined, we proceed with the classification of all the images contained in our data set in three phases that we will call: expiration, inspiration and intermediate.

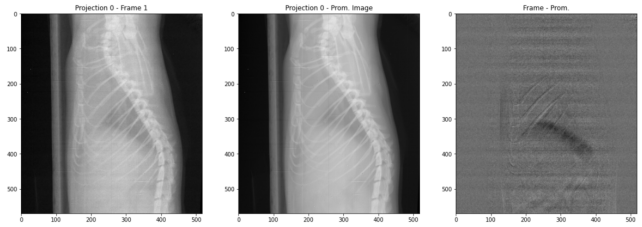

Figure 4: Identification of expiration frame.

To do this, we selected a region of interest (ROI) in the image resulting from the difference between the frame and the average image. We performed a summation of the intensities of the pixels contained in this strategically chosen ROI, which kept the highlighted part related to the movement of the diaphragm. Figure 5 shows the sum of intensity for the 16 frames of the projection we were working with. In it, it was possible to see how the lowest intensity values correspond to those frames that have a black highlighted area in the Figure 4 and the highest values correspond to the images with white highlighted areas.

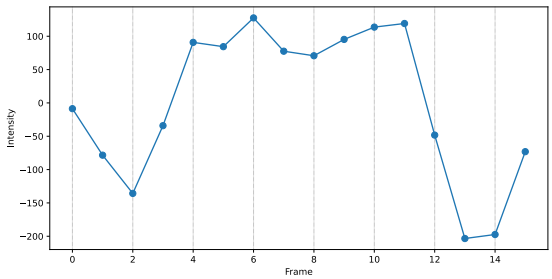

Figure 5: Respiratory signal relative to projection 0.

To classify all available images and determine which ones should be corrected to reduce the motion caused by respiration, we used the Multi-Otsu (MO) method to divide our data set. The MO is a thresholding algorithm that separates the pixels of an input image into several different classes, each obtained according to the intensity of the gray levels within the image by exhaustively searching for the threshold that minimizes the intra-class variance [5]. Applying this method, we managed to divide the data set into the three phases previously defined and *outliers*. In Figure 6, the green samples correspond to the expiration phase, the yellow ones to the inspiration phase and the red ones to the class known as 'intermediate'. Likewise, the samples that are both above threshold 3 and below threshold 4 are classified as *outliers*, those images of both expiration and inspiration with considerable movement that will introduce a lot of noise if used. In this case, the total count of images classified in phases resulted in 1088 expiration, 9032 intermediate and 1400 inspiration images.
Finally, we selected the expiration samples that were above threshold 4, as well as the inspiration samples that were below threshold 3, since these were the samples in which we corrected the movement by morphing to the sample of the same projection with the closest intensity.

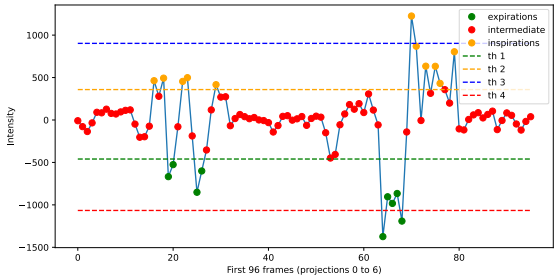

Figure 6: Classified Respiratory signal.

2.4. Model

The use of generative image models has opened up new possibilities in the field of visual content generation. These models use machine learning algorithms to learn the features and patterns present in an image set and generate new images that resemble those in the original ones. In this way, synthetic images are generated that can be used to improve the understanding of features of interest or generate additional training data sets for machine learning algorithms.
The model chosen in this case has been an autoencoder, which is a specific type of a neural network mainly designed to encode the input into a compressed and meaningful representation, and then decode it back such that the reconstructed input is similar as possible to the original one. We proposed an autoencoder architecture based on an encoder (Figure 7) which progressively reduce the dimensions and extract the essential features from the input images and a decoder (Figure 8) that aims to reconstruct the original images from the latent representation. The details of each convolutional neural network (CNN) layer used in the architecture were:
The architecture of the autoencoder, comprising encoder

ISBN: 978-84-17853-76-1

and decoder components, facilitated the compression of input images into a lower-dimensional latent representation and the subsequent reconstruction of images. By training the autoencoder to minimize the Mean Squared Error (MSE) loss between the reconstructed images and the original inputs, the model learned to capture significant features and generate precise reconstructions.

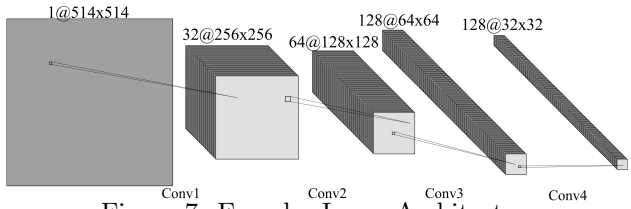

Figure 7: Encoder Layer Architecture

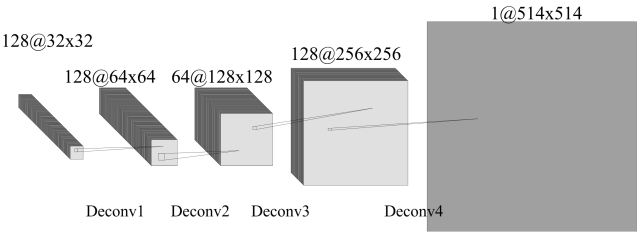

Figure 8: Decoder Layer Architecture

2.5. Morphing Technique

Morphing is a technique that had been used to generate new images lying in the feature space between two original images [6, 7, 8]. The process involved blending the latent representations of two images using a weight factor, which determined the contribution of each image to the morphed result. The weight factor ranged from 0.0 to 1.0 in increments of 0.1, indicating blending proportions. The blending was achieved by multiplying the latent representation and feature maps of the first image by 'w' and the second image by '(1-w)', with 'w' representing the weight factor. Iterating through each weight value, a morphed image was generated by passing the blended latent representation and feature maps through the decoder of the extended autoencoder [9, 10]. This morphing technique allowed for the creation of a series of images smoothly transitioning between the two originals, offering a continuous variation in the feature space, as shown in Figure 9. This enabled the exploration and generation of new artificial images with diverse characteristics.

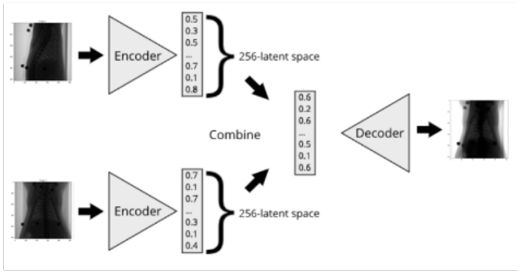

Figure 9: Fusion of the latent representation for both input images.

4. Results

The resulting reconstruction of an image is shown in Figure 10 as an example. This will be smaller due to the adjustment of dimensions that we made on the original image at the input of the autoencoder.

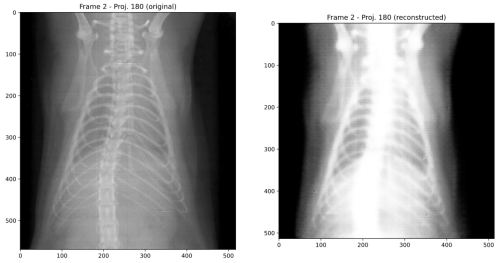

Figure 10: An example of image reconstruction using the model.

We applied the morphing technique to the dataset in order to reduce motion. Figure 6 gave an idea of how many images needed to be corrected. These images corresponded to the expiration and inspiration moments that were in the threshold ranges 1-4 and 2-3, respectively. To correct these images, we selected the frame of the same projection that has the most similar intensity in the range of *intermediate* images. In this way, we transformed the expiration and inspiration images into *intermediate* images. The evaluation method consisted of instensity profiles, drawing a line over the region of the images where we expected to find the larger motion and collecting the intensity values along that line, as shown in Figure 11.

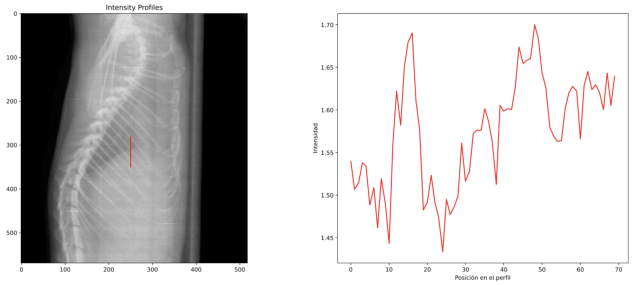

Figure 11: Method of evaluation of diaphragmatic movement on CT images.

We repeated this method for all the frames of the same projection, both for the original and the reconstructed projection. We evaluated the motion reduction after applying a Gaussian filter to remove the noise and thus better see the difference. Figure 12 showed the effect of the movement correction; the difference in intensity between the eight images depicted in the figure reflected the existing motion, as the traces do not match. In order to compare the motion between data sets (original - reconstructed) in different projections, we chose four projections at rotation degrees 1, 105, 220 and 288. The table below showed, for each angle, the number and type of images collected in the modified database: original,

ISBN: 978-84-17853-76-1

zeros and reconstructed. Therefore, we did not always have the same number of images in the same projection for both data sets. Finally, the right side of the table showed the motion quantification obtained from the profiles drawn in the projections relative to the views in both cases. The results in the table showed a tendency towards motion reduction in the modified data set. The projection at 220 degrees uses 7 original images and 8 transformed images, and it was the one that presented a greater reduction of the movement compared to the use of the original data set, with a difference of 0.183 between the intensity ranges.

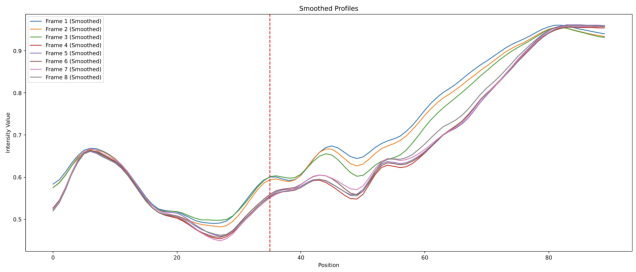

Figure 12: Motion measurement.

View	Proj.	Total Images Used			Intensity range	
		original	zeros	reconst.	original	reconst.
Right	2	11	2	3	0.082	0.017
Frontal	210	13	0	3	0.036	0.013
Left	440	3	6	7	0.161	0.121
Back	576	7	0	8	0.284	0.101

TABLE I

5. Conclusions

The fundamental objective of this work has been to address the challenge of reducing the artifact derived from the breathing of the animal under study.

Our findings conclusively support that by applying AI techniques, in particular by implementing autoencoders and neural networks, it is possible to effectively reduce artifacts and improve the quality of CT images as we use more images to average the same projection. This innovative approach has allowed us to overcome the limitation given by the use of a very small number of images associated with traditional methods.

The improvement of results using AI as a tool in this context represents a promising advance. By leveraging machine learning algorithms, we have been able to model respiratory motion more accurately and dynamically. We are excited about future research and developments in this area, with the goal of further improving medical care and clinical diagnosis. As identified limitations are addressed and more advanced approaches are explored, we are confident that even greater advances in reducing respiratory motion and optimizing CT imaging will be achieved for the benefit of patients and healthcare professionals.

Acknowledgements

This work has been supported by the Innovative Medicines Initiative 2 Joint Undertaking (JU) under grant agreement No. 853989. The JU receives support from the European Union's Horizon 2020 research and innovation programme and EFPIA and Global Alliance for TB Drug Development non-profit organisation, Bill & Melinda Gates Foundation and University of Dundee.

References

[1] Organización Mundial de la Salud. (2022). Informe mundial sobre la tuberculosis 2021. Licencia: CC BY-NC-SA 3.0 IGO

[2] PAN, Tinsu, et al. 4D-CT imaging of a volume influenced by respiratory motion on multi-slice CT. Medical physics, 2004, vol. 31, no 2, p. 333-340.

[3] NEHMEH, S. A., et al. Quantitation of respiratory motion during 4D-PET/CT acquisition. Medical physics, 2004, vol. 31, no 6, p. 1333-1338.

[4] CHAVARRÍAS, Cristina, et al. Extraction of the respiratory signal from small-animal CT projections for a retrospective gating method. Physics in Medicine Biology, 2008, vol. 53, no 17, p. 4683.

[5] LIAO, Ping-Sung, et al. A fast algorithm for multilevel thresholding. J. Inf. Sci. Eng., 2001, vol. 17, no 5, p. 713-727.

[6] LAURENT, Remy, et al. A morphing technique applied to lung motions in radiotherapy: preliminary results. Acta Polytechnica, 2010, vol. 50, no 6.

[7] JI, Dinghuang, et al. Deep view morphing. En Proceedings of the IEEE conference on computer vision and pattern recognition. 2017. p. 2155-2163..

[8] JIN, Shi, et al. Learning to dodge a bullet: Concyclic view morphing via deep learning. En Proceedings of the European Conference on Computer Vision (ECCV). 2018. p. 218-233.

[9] GONG, Maoguo, et al. A coupling translation network for change detection in heterogeneous images. International Journal of Remote Sensing, 2019, vol. 40, no 9, p. 3647-3672.

[10] DIAZ-PINTO, Andres, et al. Predicting myocardial infarction through retinal scans and minimal personal information. Nature Machine Intelligence, 2022, vol. 4, no 1, p. 55-61.