Harmonizing CT images via physics-based deep neural networks

Biel Castaño and David Faget

March 18, 2024

1 Paper summary

The paper [3] presents a novel approach to enhance the precision of CT scan quantifications for radiomics and biomarker analysis by employing a physics-based deep neural network, specifically a generative adversarial network (GAN) that incorporates the scanner's modulation transfer function (MTF). This method addresses the critical issue of CT image variability due to differences in scanning protocols, reconstruction kernels, and dose levels, which poses a significant challenge in quantitative imaging (QI). The variability can compromise the reliability and comparability of quantified imaging biomarkers, which are essential for disease diagnosis, treatment planning, and research. By harmonizing CT images to align more closely with a ground truth standard, the proposed method aims to enhance the accuracy of biomarker analysis.

Regarding its positioning with respect to the State of the Art, the paper situates its contribution against a backdrop of existing methods for image harmonization and enhancement, including traditional algorithms like parametric empirical Bayes [2] and ComBat [1] , which may compromise accuracy by altering relevant features, and physics-based methods targeting noise power spectrums (NPS) and MTF. While these approaches have their merits, they are limited by their reliance on strict assumptions or fail to comprehensively address the intrinsic variability of CT imaging. Deep neural networks (DNNs) represent a promising alternative, but they often require extensive training data and high-quality reference images. The paper's approach seeks to overcome these difficulties by training a physics-informed GAN model the model using virtual imaging trials, setting a new direction in the field.

The paper presents the successful development of a physics-informed GAN model that enhances CT image harmonization and provide three key achievements: integrating the scanner's modulation transfer function (MTF) for more accurate harmonization, utilizing a virtual CT simulator and computational patient models for generating a comprehensive training database, and significantly improving the precision of emphysema imaging biomarkers, showcasing the model's ability to reduce variability under diverse scanning conditions.

The methodology employs training a GAN using images from 40 XCAT (eXtended CArdio-Torso) computational patient models, including 10 with emphysema and the rest depicting either healthy lungs or cancer nodules. Images were generated using a CT simulator at various dose levels, kernels and scanner artifacts like noise or blur. The ground truth is based on the XCAT phantoms, excluding any scanner-induced effects.

The network architecture for the generator focuses on generating high-fidelity CT images by closely matching them to their noiseless versions without compromising spatial resolution. This is accomplished through multiple DNNs that employ nonlinear filters and work in the frequency domain using Fourier transform. An attention-focused U-Net model handles initial input adjustments, while another DNN upsamples the 2D MTF to the desired output resolution using convolutional transpose layers. A subsequent network integrates these outputs, and after applying an inverse Fourier transform, a final network generates the output. The discriminator is a fully convolutional network.

For validation, one model of emphysema and one model of a lung cancer nodule, along with their respective CT scan images, were set aside, while the remaining models of these types were used during the training phase. To evaluate the quality of the harmonized images, they were compared with the ground truth using three metrics: Structural Similarity Index Measure (SSIM), Normalized Root Mean Square Error (NRMSE), and Peak Signal-to-Noise Ratio (PSNR). The performance of the image harmonization was also assessed for various tasks, such as the quantification accuracy of biomarkers 'LAA-950' and 'Lung Mass', and the morphological radiomic features like volume or surface-to-volume ratio.

Finally, results showed harmonized images had higher structural similarity, lower error metrics, and more accurate biomarker quantifications than originals. Despite significant improvements in image quality metrics and biomarker accuracy, the harmonization process faced challenges in capturing morphological radiomics features accurately. The study underscores the harmonization process's potential to improve CT image analysis reliability in clinical and research settings, highlighting the need for refining the methodology to better capture morphological features.

2 Critical assessment

The strength of the paper summarized above lies in its comprehensive methodology, which uses virtual imaging trials and computational patient models to simulate a wide range of imaging conditions. This approach allows the model to learn from highly realistic data, improving the applicability of the results to real-world clinical scenarios. The detailed validation process, utilizing conventional quality metrics and evaluations of clinical task performance, further ensures the reliability of the proposed harmonization method in improving the precision of imaging biomarker quantifications.

However, the paper is not without its weaknesses. One notable limitation is its reduced effectiveness in harmonizing morphological radiomics features, indicating a potential gap in accurately quantifying certain biomarker types.

Additionally, the reliance on simulated data for model training may limit the adaptability of the method to real patient data variations, potentially affecting its generalizability to clinical settings because of the potential mismatch between the simulated training environments and the complex, variable conditions encountered in real clinical setting.

While the validation process employed is comprehensive, expanding testing to include real patient data from diverse clinical settings would significantly enhance the evidence supporting the model's effectiveness and its ability to generalize. Such broader testing could reveal insights into the model's performance variations across different patient populations and healthcare environments, thereby strengthening its applicability and reliability in real-world medical diagnostics.

Additionally, it is important to note that there were some orthographic errors in the text, specifically the use of "asscoated" in section 2.1, "aboce" in section 2.2 and "textrure" in section 2.3. There were also some grammatical errors, such as as the incorrect conjugation in the sentence "Ten XCAT phantoms contains variable..." in section 2.1. Addressing these errors will enhance the clarity and professionalism of the documentation.

Furthermore, it should be acknowledged that the current methodology is specifically designed for and validated only on images in the axial plane. This limitation necessitates caution when extrapolating the model's effectiveness to images captured in sagittal or coronal planes, as the model's accuracy and applicability may significantly vary across different anatomical views without additional adaptations or training tailored to those specific orientations. In addition, training images employ "forty unique computational patient models (XCAT) with variable body attributes". We think that it would be interesting to detail those attributes. Finally, we would appreciate more information about the training or inference time (this last one can be crucial for CT scans taken before a surgery, like for thrombectomy to remove the clot present in a pulmonary embolism, where each second is decisive) and about the architecture (we do not know the size of each layer and the dimensions, which should be specified in Figure 1 on the paper). We would also appreciate more information about the number of generated training and test CT scans.

To overcome these limitations, future research should focus on improving the quantification accuracy of morphological biomarkers, possibly by incorporating advanced loss functions that better capture edge information. Expanding the validation to include real patient scans from various centers could confirm the model's applicability and robustness in diverse clinical environments. Moreover, conducting a broader comparative analysis with other state-of-the-art methods, especially in real clinical settings, would provide a clearer understanding of the proposed method's strengths and limitations.

In conclusion, the paper presents a significant advancement in the field of medical imaging with its physics-informed GAN approach to CT image harmonization. By addressing its current limitations and expanding its validation, the research could further enhance its impact and applicability, offering a more reliable and accurate method for disease diagnosis and treatment monitoring in clinical practice.

References

[1] Johnson, W. E., Cheng, L., and Rabinovic, A. Adjusting batch effects in microarray expression data using empirical bayes methods. *Biostatistics 8*, 1 (2007), 118–127.

[2] Morris, C. N. Parametric empirical bayes inference: theory and applications. *Journal of the American statistical Association 78*, 381 (1983), 47–55.

[3] Zarei, M., Sotoudeh-Paima, S., McCabe, C., Abadi, E., and Samei, E. Harmonizing CT images via physics-based deep neural networks. 124631Q.