Reference-less SSIM Regression for Detection and Quantification of Motion Artefacts in Brain MRIs

Alessandro Sciarra[*1] ALESSANDRO.SCIARRA@MED.OVGU.DE
Soumick Chatterjee[*1] SOUMICK.CHATTERJEE@OVGU.DE
Max Dünnwald[1] MAX.DUENNWALD@MED.OVGU.DE
Giuseppe Placidi[2] GIUSEPPE.PLACIDI@UNIVAQ.IT
Andreas Nürnberger[1] ANDREAS.NUERNBERGER@OVGU.DE
Oliver Speck[1] OLIVER.SPECK@OVGU.DE
Steffen Oeltze-Jafra[1] STEFFEN.OELTZE-JAFRA@MED.OVGU.DE
[1] *Otto von Guericke University Magdeburg, Germany*
[2] *University of L'Aquila, Italy*

Editors: Under Review for MIDL 2022

Abstract

Motion artefacts in magnetic resonance images can critically affect diagnosis and the quantification of image degradation due to their presence is required. Usually, image quality assessment is carried out by experts such as radiographers, radiologists and researchers. However, subjective evaluation requires time and is strongly dependent on the experience of the rater. In this work, an automated image quality assessment based on the structural similarity index regression through ResNet models is presented. The results show that the trained models are able to regress the SSIM values with high level of accuracy. When the predicted SSIM values were grouped into 10 classes and compared against the ground-truth motion classes, the best weighted accuracy of $89 \pm 2\%$ was observed with RN-18 model, trained with contrast augmentation.

Keywords: Motion artefacts, MRI, ResNet, Image quality assessment

1. Introduction

Image quality assessment (IQA) is a critical step to evaluate if the quality of the MR images can guarantee diagnostic reliability (Khosravy et al., 2019). Moreover, it is an important step for large clinical studies as typically they require high quality data. Often the evaluation process requires time and is subjectively dependent upon the observer (Ma et al., 2020). Structural similarity index measure (SSIM) is a popular way of evaluating the quality of the images objectively, but it requires reference images. If the images are artificially corrupted, then the original non-corrupted images can serve as reference to calculate the SSIM values - which is not possible during real-life acquisitions. This research proposes an automated IQA method to detect the presence of motion artefacts and quantify the level of corruption by regressing the SSIM values directly from the corrupted images using convolutional neural networks, without using any reference image.

2. Methodology

The proposed IQA method uses a ResNet model (He et al., 2016) with different depths - ResNet18 (RN18) and ResNet101 (RN-101). Given a 3D input volume during training, one random slice (2D image) is selected from one of the possible orientations - axial, sagittal, and coronal. To make the model more robust against changes in the image contrast in clinical scenarios, four different random contrast augmentation (CA) techniques were

[*] A. Sciarra and S. Chatterjee contributed equally

employed during training - gamma manipulation, logarithmic manipulation, sigmoid manipulation, and adaptive histogram adjustments. Then, artificial motion corruption was applied on these images using two different methods - the RandomMotion functionality of TorchIO (Pérez-García et al., 2021) and a more physically realistic in-house line-wise motion corruption algorithm. The SSIM values between these artificially corrupted images and the corresponding non-corrupted images were calculated, and were used as the ground-truth values to train the models. The mean squared error (MSE) between these values and the models' predictions were compared as the loss during training and optimised using the Adam optimiser, a learning rate of $1e^{-3}$ and a batch size of 100 for 2000 epochs. 300 MRI volumes with different acquisition devices and parameters were used in this research, split into 200-50-50, for training, validation, and testing, respectively. T1, T2, PD, and FLAIR images acquired at three different sites using different devices were included (114 volumes at 3T, 93 volumes at 7T, 25 volumes with different 1.5 and 3T scanners), while the remaining 68 volumes were taken from the T1, T2, and PD weighted MRIs of the publicly-available IXI dataset. All the trainings and evaluations were performed by combining all these different contrasts and other variations - to make the model generally-applicable in clinical situations. All the images were intensity-normalised by dividing by its max value and interpolated or padded to a 2D matrix size of 256x256.

3. Evaluation

In order to evaluate the performances of these models, a scatter plot for the SSIM values has been used, specifically, predicted against ground truth values, as shown in figure 1 (ii).Both models trained with contrast augmentation show less dispersion. The term dispersion refers to the distance from the ideal linear function $y = x$ with unitary coefficient, where y is the predicted SSIM value while x the ground truth value. The regression performance of the models were also evaluated using residual SSIM values (the difference between ground-truth and the predicted SSIMs), and the best performing model RN-18 with CA got -0.0009±0.0139. The predicted SSIM value can be considered to measure the distortion or corruption level of the image. However, when applying this approach to a real clinical case it is difficult to compare it with a subjective assessment. To get around this problem, the regression task was simplified as a classification by sub-dividing the SSIM range [0-1] into 10 classes: class-1:[0.00-0.10], class-2:[0.11-0.20] and so on. The SSIM values predicted by the models, as well as the ground-truth SSIMs were converted into these 10 classes, referred as the predicted classes and true classes, respectively and the results are shown in Figure 2. The best weighted accuracy has been achieved by RN-18 with CA, $89 \pm 2\%$, followed by RN-101 with CA $88 \pm 2\%$, RN-18 without CA $87 \pm 2\%$ and RN-101 without CA $86 \pm 2\%$.

4. Conclusion

This research presents an SSIM-regression based IQA technique using ResNet models, coupled with contrast augmentations to make them robust against changes in the image contrasts in clinical scenarios. The method managed to predict the SSIM values from artificially motion corrupted images without the ground-truth (motion-free) images with high accuracy (residual SSIMs as less as -0.0009±0.0139). Moreover, the motion classes obtained from the predicted SSIMs were very close to the true ones and achieved a weighted accuracy of $89 \pm 2\%$. Considering the complexity of the problem in quantifying the image degradation level due to motion artefacts and additionally the variability of the type of contrast, resolution, etc., the results obtained are promising. Further evaluations, including

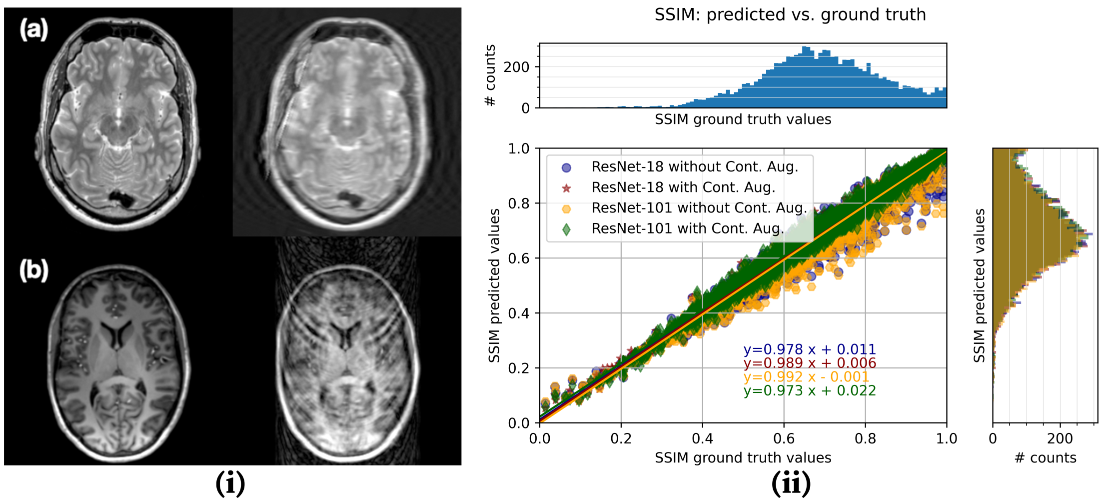

Figure 1: ***(i)***: Samples of artificially corrupted images. On the left column original images, on the right the corrupted images using **(a)** TorchIO and **(b)** in-house algorithm. ***(ii)***: (bottom-left) dispersion plot SSIM predicted vs. ground truth values; (top-left) histogram of the SSIM ground truth values; (bottom-right) histograms of the SSIM predicted values.

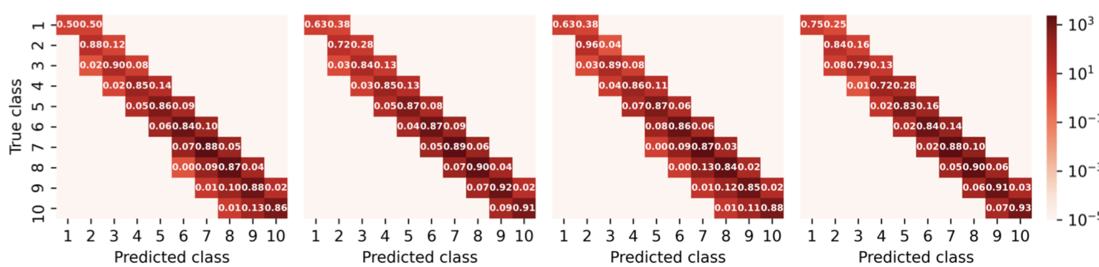

Figure 2: Confusion matrices for the classification task. From left to right: RN-18 without contrast augmentation (CA), RN-18 with CA, RN-101 without CA, RN-101 with CA

subjective evaluation, will be performed on clinical data to judge its clinical applicability and robustness against changes in real-world scenarios.

Acknowledgments

This work was supported by the state of Saxony-Anhalt ('I 88') and the ESF (ZS/2016/08/80646).

References

Kaiming He et al. Deep residual learning for image recognition. In *IEEE CVPR*, 2016.

Mahdi Khosravy et al. Image quality assessment: A review to full reference indexes. *Recent trends in communication, computing, and electronics*, 2019.

Jeffrey J Ma et al. Diagnostic image quality assessment and classification in medical imaging: Opportunities and challenges. In *IEEE ISBI*, 2020.

Fernando Pérez-García et al. Torchio: a python library for efficient loading, preprocessing, augmentation and patch-based sampling of medical images in deep learning. *Computer Methods and Programs in Biomedicine*, 208:106236, 2021.