Adversarial Bayesian Optimization for Quantifying Motion Artifact Within MRI

Anastasia Butskova[1], Rain Juhl[1], Dženan Zukić[2], Aashish Chaudhary[2], Kilian M. Pohl[1,3], and Qingyu Zhao[1(✉)]

[1] Stanford University, Stanford, CA, USA
qingyuz@stanford.edu
[2] Kitware Inc., Carrboro, NC, USA
[3] SRI International, Menlo Park, CA, USA

Abstract. Subject motion during an MRI sequence can cause ghosting effects or diffuse image noise in the phase-encoding direction and hence is likely to bias findings in neuroimaging studies. Detecting motion artifacts often relies on experts visually inspecting MRIs, which is subjective and expensive. To improve this detection, we develop a framework to automatically quantify the severity of motion artifact within a brain MRI. We formulate this task as a regression problem and train the regressor from a data set of MRIs with various amounts of motion artifacts. To resolve the issue of missing fine-grained ground-truth labels (level of artifacts), we propose Adversarial Bayesian Optimization (ABO) to infer the distribution of motion parameters (i.e., rotation and translation) underlying the acquired MRI data and then inject synthetic motion artifacts sampled from that estimated distribution into motion-free MRIs. After training the regressor on the synthetic data, we applied the model to quantify the motion level in 990 MRIs collected by the National Consortium on Alcohol and Neurodevelopment in Adolescence. Results show that the motion level derived by our approach is more reliable than the traditional metric based on Entropy Focus Criterion and manually defined binary labels.

1 Introduction

Due to the relatively long scanning time required for MRI sequences, subject motion creates one of the most common artifacts in MR imaging that results in blurring and ghosting in the image [1]. Fast and accurate quantification of motion artifacts within MRIs, therefore, becomes critical in most clinical settings as it generates real-time feedback on whether a re-scan is needed while the subject is still in the scanner, thereby reducing the need for scheduling additional scanning sessions [2]. Moreover, in large-scale neuroimaging studies, such quantification can be used to minimize the risk of motion-related confounders biasing findings by either confining analysis to MRIs with minimum motion artifacts or controlling for the level of motion artifact in statistical models [3,4].

© Springer Nature Switzerland AG 2021
I. Rekik et al. (Eds.): PRIME 2021, LNCS 12928, pp. 83–92, 2021.
https://doi.org/10.1007/978-3-030-87602-9_8

In most existing neuroimaging studies, detection of motion artifacts in T1-weighted MRI relies on experts visually reviewing scans and labeling an MRI as either motion-free or with motion artifacts [5]. Such binary (or categorical) labeling is often subjective and inaccurate, especially when the scans are from multiple sites with various imaging protocols. An alternative to manual labeling is to use Entropy Focus Criterion (EFC) [6] to quantify the severity of motion artifact, but this metric is only suitable for detecting large-scale motions. To improve the quantification, prior works have tried to automate motion detection based on machine learning models. Lorch et al. [2] proposed a framework based on Random Forests to detect head and respiratory motion in MRIs, but their framework was only evaluated on synthetic data. Other works [7,8] designed a convolutional neural network to classify image patches and then used a special voting mechanism to determine the presence of motion artifact in the entire MRI. However, these classification models always assume availability of accurate ground-truth labels, which, as mentioned above, are difficult and expensive to produce for large multi-site studies. Moreover, these models only provide binary predictions, which cannot quantify detailed severity of the artifact. To resolve these issues, we aim to develop a regression framework to quantify the fine-grained level of motion artifact in an MRI via a convolutional neural network.

We are specifically interested in designing a training scheme to learn the regression network based on binary (presence or absence of artifact) and potentially noisy ground-truth labels derived by manual labeling. To generate training data with accurate and continuous labels (level of artifact), we injected synthetic artifacts induced by rigid motion into motion-free MRIs. To do so, we first developed a novel method called Adversarial Bayesian Optimization (ABO) to infer the distribution of motion parameters, i.e., the range of rotation and translation in 3D, underlying the motion observed in the real data set. We then generated a simulated data set with synthetic artifacts controlled by the estimated motion range, so that the regressor network could be trained in a traditional supervised fashion. We tested the proposed framework to quantify the severity of motion artifact in 990 T1-weighted MRIs from the National Consortium on Alcohol and Neurodevelopment in Adolescence (NCANDA) [9]. Results indicate that the proposed ABO could accurately estimate the range of motion parameters underlying a given data set. The regressor trained based on this estimated range more accurately quantified the level of motion in MRIs compared to traditional metrics based on EFC.

2 Methods

Let $\boldsymbol{\mathcal{X}}_c$ be a set of motion-free MRIs and $\boldsymbol{\mathcal{X}}_m$ be a set of MRIs with various level of motion, our goal is to derive a function q, such that for an MRI $\boldsymbol{x} \in \mathcal{X}_m$, $y = q(\boldsymbol{x})$ quantifies the level of motion. To train the regressor q without ground-truth y, a typical approach is to rely on data simulation. Here, we leverage the simulator function $\hat{\boldsymbol{x}} = s(\boldsymbol{x}; \boldsymbol{\tau})$ proposed in [10]. The simulator can inject "k-space" motion artifacts controlled by a set of parameters $\boldsymbol{\tau}$ (encoding rotation

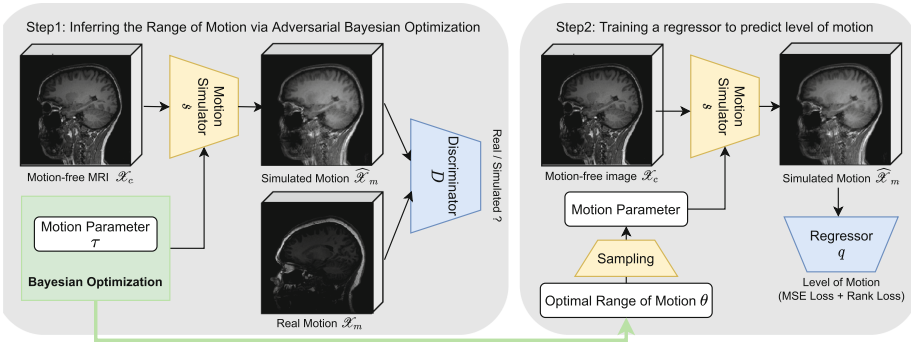


Fig. 1. Model Architecture: Step 1. Adversarial Bayesian Optimization (ABO) is used to infer the distribution of motion parameters underlying the real data set with motion artifacts $\mathcal{X}_m$; Step 2. A regressor is trained on a simulated data set, whose motion artifacts are sampled from the distribution estimated in Step 1.

and translation of rigid motion) into a motion-free image $\boldsymbol{x} \in \boldsymbol{\mathcal{X}}_c$ to generate a synthetic motion-corrupted MRI $\hat{\boldsymbol{x}} \in \hat{\boldsymbol{\mathcal{X}}}_m$. As the ground-truth motion of $\hat{\boldsymbol{x}}$ is known (related to $\boldsymbol{\tau}$), one can learn q using traditional supervised training.

However, due to the variability in imaging protocols and characteristics of the study cohorts, one typically does not know the range of motion in $\boldsymbol{\mathcal{X}}_m$, so the model trained on the simulated set $\hat{\boldsymbol{\mathcal{X}}}_m$ might not generalize well. To resolve this issue, we propose the following 2-step training strategy (Fig. 1): (1) infer the underlying distribution of motion parameters $p(\boldsymbol{\tau})$ associated with the real motion artifacts in $\boldsymbol{\mathcal{X}}_m$; and (2) train the regressor q on a simulated data set $\hat{\boldsymbol{\mathcal{X}}}_m$, where $\boldsymbol{\tau}$ (and also y) for each simulated MRI is sampled from $p(\boldsymbol{\tau})$.

2.1 Inferring Distribution of Motion Parameters via Adversarial Bayesian Optimization

Let $\boldsymbol{\tau} = \{\tau_1, ..., \tau_6\}$ be the amount of rotation and translation of the simulated motion in the 3 spatial directions, and let $\boldsymbol{\theta} = \{\theta_1, ..., \theta_6\}$ be the maximum transformation in those 6°. We then assume the transformation parameters follow independent uniform distributions $p(\boldsymbol{\tau}; \boldsymbol{\theta}) = \prod_k p(\tau_k; \theta_k) = \prod_k \mathcal{U}(-\theta_k, \theta_k)$. As such, $p(\boldsymbol{\tau}; \boldsymbol{\theta})$ can be used to characterize a data set with various level of motion artifacts.

We now aim to infer the maximum transformation $\boldsymbol{\theta}$ associated with $\boldsymbol{\mathcal{X}}_m$. We do so by adversarially training a discriminator network D with parameters $\boldsymbol{\phi}$, such that it can not distinguish the real data $\boldsymbol{\mathcal{X}}_m$ from simulated data $\hat{\boldsymbol{\mathcal{X}}}_m$, which are generated by injecting synthetic motion (sampled from $p(\boldsymbol{\tau}; \boldsymbol{\theta})$) into motion-free MRIs $\boldsymbol{\mathcal{X}}_c$. This objective function can be written as

$$\max_{\boldsymbol{\theta}} \min_{\boldsymbol{\phi}} \mathbb{E}_{\boldsymbol{x} \sim \boldsymbol{X}_m} [\log D_{\boldsymbol{\phi}}(\boldsymbol{x})] + \mathbb{E}_{\boldsymbol{\tau} \sim p(\boldsymbol{\tau}; \boldsymbol{\theta}), \boldsymbol{x} \sim \boldsymbol{\mathcal{X}}_c} [\log(1 - D_{\boldsymbol{\phi}}(s(\boldsymbol{x}; \boldsymbol{\tau})))] . \quad (1)$$

As typically performed in adversarial learning, this loss function can be alternatively optimized with respect to $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$. To update $\boldsymbol{\phi}$ given $\boldsymbol{\theta}$, we first simulate $\hat{\boldsymbol{X}}_m$ from $\boldsymbol{X}_c$ according to $p(\boldsymbol{\tau};\boldsymbol{\theta})$ and then train the discriminator D to differentiate $\hat{\boldsymbol{X}}_m$ from $\boldsymbol{X}_m$. To update $\boldsymbol{\theta}$ given a trained discriminator D, we optimize $\boldsymbol{\theta}$ so that the discriminator cannot distinguish between real and simulated data. We do so by maximizing the following function to fool the discriminator.

$$f(\boldsymbol{\theta}) = \mathbb{E}_{\boldsymbol{\tau}\sim p(\boldsymbol{\tau};\boldsymbol{\theta}),\boldsymbol{x}\sim\boldsymbol{\mathcal{X}}_c}\left[\log(1 - D_{\boldsymbol{\phi}}(s(\boldsymbol{x};\boldsymbol{\tau})))\right]. \tag{2}$$

However, a challenge in the optimization of Eq. (2) is that the gradient of $f(\boldsymbol{\theta})$ is hard to evaluate (due to the k-space transformation of $\boldsymbol{x}$ in $s(\boldsymbol{x};\boldsymbol{\tau})$), so traditional back-propagation can not be directly applied. Here, we resort to Bayesian optimization to optimize Eq. (2).

Bayesian Optimization. [11] is a machine-learning-based approach for finding the maximum of f without estimating its gradient. To do so, Bayesian optimization formulates a probability distribution over the functional $f(\cdot)$ as a Gaussian Process and repeatedly evaluates function values at various points. Let $\boldsymbol{\Theta} = [\boldsymbol{\theta}^1, ..., \boldsymbol{\theta}^t]$ be the first t evaluated points, and $\boldsymbol{f} = [f(\boldsymbol{\theta}^1), ..., f(\boldsymbol{\theta}^t)]$ be their functional values. Let $k(\cdot,\cdot)$ denote the covariance function [12] associated with the Gaussian Process. Then the functional value at an unobserved point $\boldsymbol{\theta}^*$ follows a normal distribution $f(\boldsymbol{\theta}^*) \sim \mathcal{N}(\mu[\boldsymbol{\theta}^*], \sigma^2[\boldsymbol{\theta}^*])$, where

$$\mu[\boldsymbol{\theta}^*] = K(\boldsymbol{\theta}^*, \boldsymbol{\Theta})K(\boldsymbol{\Theta}, \boldsymbol{\Theta})^{-1}\boldsymbol{f} \tag{3}$$

$$\sigma^2[\boldsymbol{\theta}^*] = K(\boldsymbol{\theta}^*, \boldsymbol{\theta}^*) - K(\boldsymbol{\theta}^*, \boldsymbol{\Theta})K(\boldsymbol{\Theta}, \boldsymbol{\Theta})^{-1}K(\boldsymbol{\Theta}, \boldsymbol{\theta}^*), \tag{4}$$

Here, $K(\boldsymbol{\Theta}, \boldsymbol{\Theta})$ is a $t \times t$ matrix with each element (i, j) recording $k(\boldsymbol{\theta}^i, \boldsymbol{\theta}^j)$, and likewise $K(\boldsymbol{\theta}^*, \boldsymbol{\Theta})$ is a $t \times 1$ vector. Given such a probability distribution, an acquisition function is used to select the next evaluation point with the highest chance to get closer to the global maximum. Here we use the Upper Confidence Bound (UCB) [12] as the acquisition function

$$\arg\max_{\boldsymbol{\theta}^*} \mathrm{UCB}(\boldsymbol{\theta}^*) = \mu[\boldsymbol{\theta}^*] + \beta\sigma[\boldsymbol{\theta}^*]; \tag{5}$$

That is, the UCB favors regions with either high expected values $\mu[\boldsymbol{\theta}^*]$ or high uncertainty (for exploring scarcely evaluated regions). As such, we repeatedly use Eqs. (3), (4), and (5) to evaluate new points until reaching a certain number of iterations. Lastly, we embed the Bayesian optimization strategy into the adversarial training scheme, resulting in the Adversarial Bayesian Optimization (see Algorithm 1).

2.2 Training a Regressor for Quantifying Level of Motion

We now simulate the data set $\hat{\boldsymbol{\mathcal{X}}}_m$ to train the regressor q. To generate a simulated $\hat{\boldsymbol{x}}$ and its corresponding ground truth y, we first randomly select $\boldsymbol{x} \in \boldsymbol{\mathcal{X}}_c$ and sample y from $\mathcal{U}(0, 1)$. Then we construct $\hat{\boldsymbol{x}} = s(\boldsymbol{x}; y\boldsymbol{\theta})$, i.e. an image with

Algorithm 1. Adversarial Bayesian Optimization (ABO)

```
Initialize θ
for n = 1 to N do
  Simulate X̂_m by sampling MRI from X_c and sampling motion from p(τ; θ)
  Train the discriminator D on X_m and X̂_m
  Bayesian Optimization to Update θ:
    for t = 1 to T do
      Evaluate f(θ^t) via Eq. (2)
      Evaluate μ[θ*] and σ²[θ*] via Eqs. (3) and (4)
      Derive θ^{t+1} by optimizing Eq. (5)
    end for
    θ ← arg max(f(θ^1)...f(θ^T))
end for=0
```

motion artifact at level y, and the maximum possible motion is guaranteed to be $\boldsymbol{\theta}$. Next, we train q on the simulated image and label pairs. As we intend to use q to rank the motion level across images, we design the training loss function as a combination of MSE loss and a rank loss [13]. Given two training images $\hat{\boldsymbol{x}}_1, \hat{\boldsymbol{x}}_2$ with $y_1 < y_2$, the objective function is

$$\arg\min_{q} \ (y_1 - q(\hat{\boldsymbol{x}}_1))^2 + (y_2 - q(\hat{\boldsymbol{x}}_2))^2 + \mathrm{ReLU}(q(\hat{\boldsymbol{x}}_1) - q(\hat{\boldsymbol{x}}_2)), \tag{6}$$

which encourages the predictions to follow the correct rank order.

3 Experimental Setup

Data Set. We used the T1-weighted MRI from the National Consortium on Alcohol and Neurodevelopment in Adolescence (NCANDA) [9,14], which acquired MRIs at 5 different sites across the US (denoted in the following as Site A through Site E). Up to year 6 of the study, all MRIs went through manual quality control, and 159 MRIs were labeled as 'unusable' due to motion artifacts, which were regarded as the real motion data set $\boldsymbol{\mathcal{X}}_m$. To construct $\boldsymbol{\mathcal{X}}_c$, we first collected the 831 MRIs from all subjects that were labeled as 'usable' by manual quality control[1]. Considering imperfect human labeling, we selected 200 MRIs with lowest Entropy Focus Criterion (EFC) as $\boldsymbol{\mathcal{X}}_c$. All MRIs analyzed in this study had a resolution of $256 \times 256 \times 146$ or $256 \times 256 \times 160$ but were downsized to the size of 128×128 in the axial plane and cropped in the z-direction to reserve 128 axial slices containing the brain region.

Validation. To test if the proposed ABO can accurately imply the range of motion parameters associated with a data set, we applied it to simulated data sets with known ground-truth $\boldsymbol{\theta}$. We did so by setting all the 6 parameters $\{\theta_1 ... \theta_6\}$ to an equal number and varied this number in the range of $\{0.1, 0.2, 0.5, 1, 2, 4\}$. We then ran the optimization using a search range of [0,

[1] Only baseline scan was used if longitudinal scans were available.

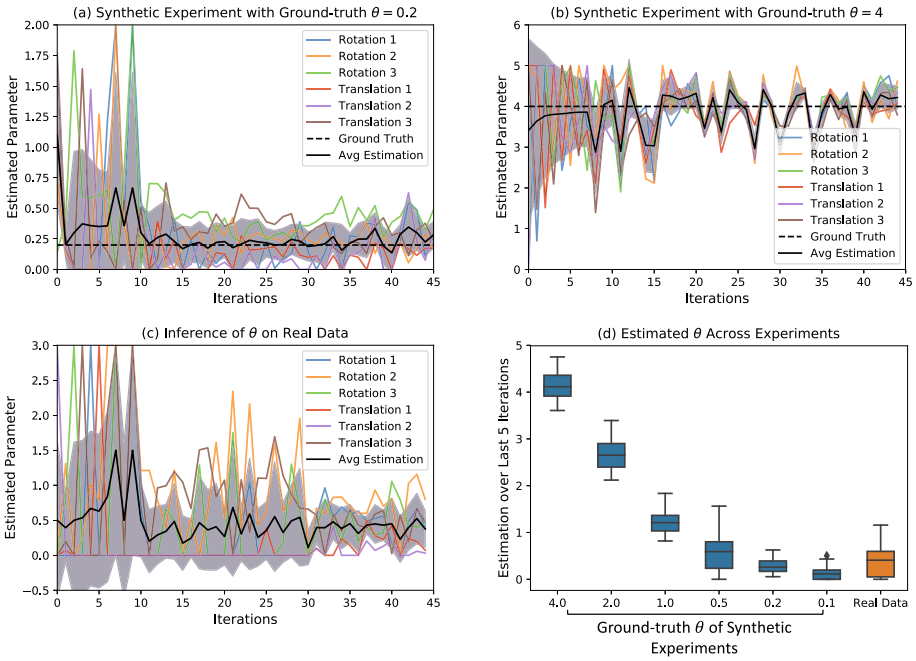


Fig. 2. (a–c). Estimated range of motion $\boldsymbol{\theta}$ along iterations of Adversarial Bayesian Optimization (ABO) for synthetic and real data sets. (d) Distribution of the 6 estimated motion parameters over the last 5 iterations of ABO for synthetic (blue) and real (orange) data sets. (Color figure online)

5] for each motion parameter, initialized $\boldsymbol{\theta}$ with 5 searches in that range, and performed additional search 45 times with $\beta = 2.58$ in the UCB, which decayed by a factor of 0.95 after the 5^{th} iteration. After this validation, we applied ABO to infer $\boldsymbol{\theta}$ associated with the real motion set $\boldsymbol{\mathcal{X}}_m$.

Next, given the estimated $\boldsymbol{\theta}$, we simulated $\hat{\boldsymbol{\mathcal{X}}}_m$ containing 200 synthetic MRIs with ground-truth y. We additionally simulated 200 MRIs as a separate validation data set. To train the regressor q, we designed a simple feed-forward convolutional neural work containing 4 double convolutional layers of feature-channel dimension increasing from 32 to 256 with ReLU activations and max-pooling layers in between. Each convolutional layer used a kernel size of 3×3, stride of 1, and zero padding. Adam optimizer was used with a learning rate of 1e-4. After training, we applied q to the validation set to measure the prediction accuracy of y based on R^2 coefficients. We also compared the proposed loss function in Eq. (6) with the vanilla regression simply based on the MSE loss. Finally, we applied q to quantify the level of motion for the real data in $\boldsymbol{\mathcal{X}}_m$ and all usable baseline scans of NCANDA. The code for ABO and regression is available at https://github.com/anastasb/MRI-Motion-Artifact-Quantification.

4 Results

Results on Adversarial Bayesian Optimization. Figure 2 shows the inference of $\boldsymbol{\theta}$ for synthetic data set $\hat{\boldsymbol{\mathcal{X}}}_m$ with known ground-truth $\boldsymbol{\theta}$ and for real data $\boldsymbol{\mathcal{X}}_m$. When $\hat{\boldsymbol{\mathcal{X}}}_m$ was constructed by ground-truth $\boldsymbol{\theta} = 0.2$, all 6 estimated parameters converged around the ground-truth value with diminishing variance (Fig. 2(a)). The results replicated when we tested the optimization on $\hat{\boldsymbol{\mathcal{X}}}_m$ with a larger ground-truth $\boldsymbol{\theta} = 4$ (Fig. 2(b)). The first 6 columns in Fig. 2(d) indicate that the proposed ABO approach consistently recovered the correct range of motion $\boldsymbol{\theta}$ for data sets with various ground-truth $\boldsymbol{\theta}$.

Next, we applied ABO to the real motion data set $\boldsymbol{\mathcal{X}}_m$ to infer its underlying motion parameters $\boldsymbol{\theta}$ (Fig. 2(c)), and the estimated parameters converged around 0.5. We also observe from Fig. 2(d) that the distribution of the 6 estimated parameters (over the last 5 iterations) for the real set $\boldsymbol{\mathcal{X}}_m$ was comparable to the distribution for the simulated set $\hat{\boldsymbol{\mathcal{X}}}_m$ with ground-truth $\boldsymbol{\theta} = 0.5$. We therefore set $\boldsymbol{\theta} = 0.5$ for all 6 motion parameters for the real motion set $\boldsymbol{\mathcal{X}}_m$.

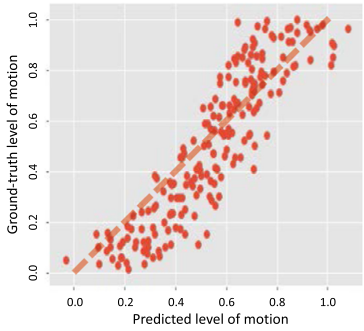


Fig. 3. Inference of motion level on the validation set.

Results on Motion Level Regression. After training q on a simulated set $\hat{\boldsymbol{\mathcal{X}}}_m$ with $\boldsymbol{\theta} = 0.5$, we tested it on the hold-out set of 200 simulated test images. Figure 3 shows that the predicted motion level generally aligned with the ground truth ($R^2 = 0.52$), although the prediction seemed to be overly sensitive for small motion artifacts (e.g., < 0.2). However, this result was still more accurate than the vanilla regression model that only incorporated the MSE loss but without using the ranking loss ($R^2 = 0.43$).

Finally, we applied the trained regressor to estimate the level of motion for all 990 MRIs. Figure 4 displays MRIs from each site with the highest and lowest

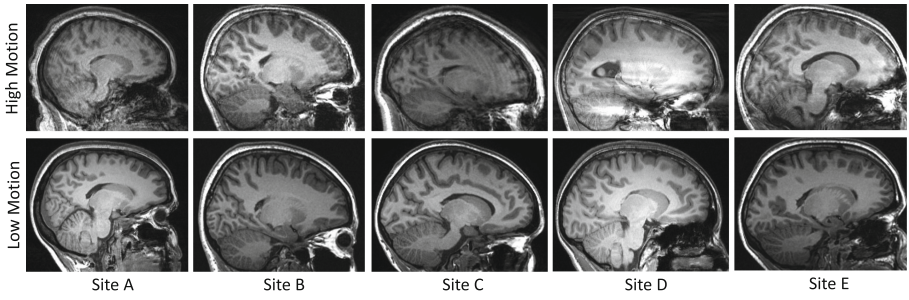
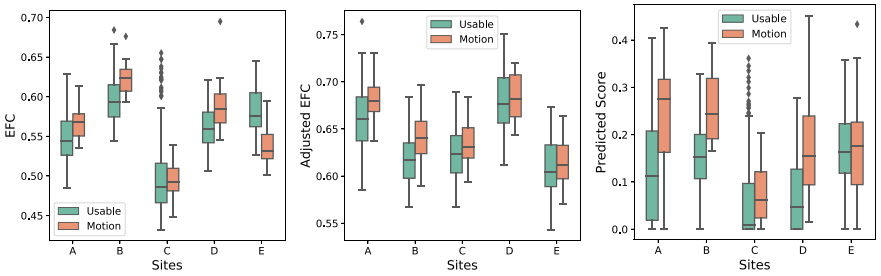


Fig. 4. MRIs from each site with the highest and lowest motion score derived by our regression model.



(a) EFC (b) Adjusted EFC (c) Our model

Fig. 5. Motion scores for the 990 MRIs of NCANDA defined by (a) EFC, (b) adjusted EFC (EFC computed over voxels with positive intensities), and (c) our model prediction. Our model results in the largest group separation in the scores for Sites A, B, and D, and has less variance across sites.

motion score. We then show the estimated scores for both usable MRIs and MRIs with motion across 5 sites in Fig. 5(c), and compared them with the scores defined by traditional metrics for motion quantification based on EFC (Fig. 5(a)+(b)). Note that the scale of motion level was incommensurate across approaches. We observe that the EFC-based scores exhibited a large variance across sites, which was significantly larger than the group difference between the usable and motion groups. This finding rendered the EFC metrics unsuitable for multi-site studies. On the other hand, the cross-site variance was largely reduced by our approach.

Compared to the EFC-based scores, the scores derived by the regressor network had a significantly larger separation between the usable and motion groups for sites A, B, and D ($p < 0.01$ two-way ANOVA). This indicates our approach could better quantify motion artifacts within MRIs. Note, although we expected a significant separation between the two groups, we did not regard the binary group assignment as ground-truth due to errors in manual quality control. This was supported by Fig. 6 showing two example MRIs defined as 'usable' in quality control but received high motion scores by our model. A follow-up visual inspection concluded that there were substantial motion artifacts (e.g., ringing effects) within these MRIs. Therefore, the overlap in the motion scores between the usable and motion groups across all sites potentially reflected the ambiguity and imperfection of manual labeling. Such noisy manual labels might prohibit effective training of supervised binary classification and supported that our regression-based framework might be more effective in quantifying motion artifacts.

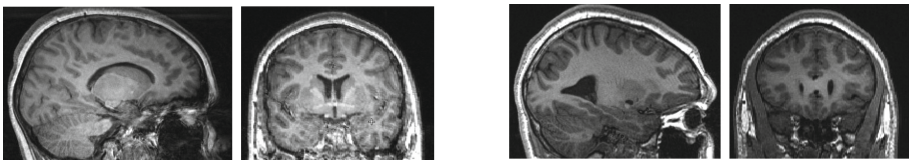


Fig. 6. Examples of MRIs which manual quality control labeled as usable but our estimator gave a high motion score.

5 Conclusion

We presented a framework for quantifying the severity of artifact induced by subject motion in T1-weighted MRIs. Being a part of the framework, we proposed Adversarial Bayesian Optimization, a novel approach for performing adversarial training when the generator component (i.e., the motion simulator in our context) is non-differentiable. The proposed motion-quantification framework can be used either as a first or second rater in quality control workflows. Here we used it as a second rater and found some instances of motion artifact which escaped manual quality control (Fig. 6). Our future research direction is to develop a method to improve the quality of MRIs with motion artifacts. In other words, given a motion-corrupted MRI, we aim to reconstruct the corresponding motion-free MRI. For that, the method presented herein was a prerequisite.

Acknowledgment. This research was supported in part by NIH U24 AA021697, MH119022, and Stanford HAI AWS Cloud Credit. The data were part of the public NCANDA data release NCANDA_PUBLIC_BASE_STRUCTURAL_V01 [15], whose collection and distribution were supported by NIH funding AA021697, AA021695, AA021692, AA021696, AA021681, AA021690, and AA02169.

References

1. Zaitsev, M., Maclaren, J., Herbst, M.: Motion artefacts in MRI: a complex problem with many partial solutions. J. Magn. Reson. Imaging: JMRI **42**, 887–901 (2015)
2. Lorch, B., Vaillant, G., Baumgartner, C., Bai, W., Rueckert, D., Maier, A.: Automated detection of motion artefacts in MR imaging using decision forests. J. Med. Eng. 1–9 (2017)
3. Reuter, M., Tisdall, M., Qureshi, A., Buckner, R., Kouwe, A., Fischl, B.: Head motion during MRI acquisition reduces gray matter volume and thickness estimates. NeuroImage **107**, 107–115 (2015)
4. Johnstone, T.: Motion correction and the use of motion covariates in multiple-subject fMRI analysis. Human Brain Mapp. **27**, 779–88 (2006)
5. Backhausen, L., Herting, M., Buse, J., Roessner, V., Smolka, M., Vetter, N.: Quality control of structural MRI images applied using freesurfer–a hands-on workflow to rate motion artifacts. Front. Neurosci. **10**, 1–10 (2016)

6. Atkinson, D., Hill, D., Stoyle, P., Summers, P., Keevil, S.: Automatic correction of motion artifacts in magnetic resonance images using an entropy focus criterion. IEEE Trans. Med. Imaging **16**, 903–910 (1997)
7. Küstner, T., et al.: Automated reference-free detection of motion artifacts in magnetic resonance images. Magn. Reson. Mater. Phys. Biol. Med. **31**, 243–256 (2017). https://doi.org/10.1007/s10334-017-0650-z
8. Fantini, I., Rittner, L., Yasuda, C., Lotufo, R.: Automatic detection of motion artifacts on MRI using deep CNN. In: 2018 International Workshop on Pattern Recognition in Neuroimaging (PRNI), pp. 1–4 (2018)
9. Brown, S.A., et al.: The national consortium on alcohol and neurodevelopment in adolescence (NCANDA): a multisite study of adolescent development and substance use. J. Stud. Alcohol Drugs **76**(6), 895–908 (2015)
10. Pérez-García, F., Sparks, R., Ourselin, S.: Torchio: a python library for efficient loading, preprocessing, augmentation and patch-based sampling of medical images in deep learning. Comput. Methods Programs Biomed. **208**, 106236 (2021)
11. Frazier, P.I.: A tutorial on Bayesian optimization arxiv:1807.02811 (2018)
12. Snoek, J., Larochelle, H., Adams, R.P.: Practical Bayesian optimization of machine learning algorithms (2012) arXiv:1206.2944
13. Chen, W., Liu, T.Y., Lan, Y., Ma, Z.M., Li, H.: Ranking measures and loss functions in learning to rank. In: Bengio, Y., Schuurmans, D., Lafferty, J., Williams, C., Culotta, A. (eds.) Advances in Neural Information Processing Systems, vol. 22, Curran Associates, Inc. (2009)
14. Pfefferbaum, A., et al.: Adolescent development of cortical and white matter structure in the NCANDA sample: role of sex, ethnicity, puberty, and alcohol drinking. Cereb. Cortex **26**, 4101–4121 (2015)
15. Pohl, K.M., Sullivan, E.V., Pfefferbaum, A.: The NCANDA_PUBLIC_BASE_STRUCTURAL_V01 data release of the national consortium on alcohol and neurodevelopment in adolescence (NCANDA). In: Sage Bionetworks Synapse (2017)