AUTOMATIC DETECTION OF MOTION ARTIFACTS IN MR IMAGES USING CNNS

Kristof Meding, Alexander Loktyushin and Michael Hirsch

Max Planck Institute for Intelligent Systems
Department of Empirical Inference
{firstname.lastname}@tuebingen.mpg.de

ABSTRACT

Considerable practical interest exists in being able to automatically determine whether a recorded magnetic resonance image is affected by motion artifacts caused by patient movements during scanning. Existing approaches usually rely on the use of navigators or external sensors to detect and track patient motion during image acquisition. In this work, we present an algorithm based on convolutional neural networks that enables fully automated detection of motion artifacts in MR scans without special hardware requirements. The approach is data driven and uses the magnitude of MR images in the spatial domain as input. We evaluate the performance of our algorithm on both synthetic and real data and observe adequate performance in terms of accuracy and generalization to different types of data. Our proposed approach could potentially be used in clinical practice to tag an MR image as motion-free or motion-corrupted immediately after a scan is finished. This process would facilitate the acquisition of high-quality MR images that are often indispensable for accurate medical diagnosis.

Index Terms— Motion Artifacts, MRI, Quality Assessment, Convolutional Neural Networks, Deep Learning

1. INTRODUCTION

In magnetic resonance imaging (MRI), patient motion during image acquisition degrades image quality, leading to ghosting and blurring artifacts [1]. Even slight motion can cause artifacts [2]. As a result, 30% of MRI scans in clinical practice cannot be used for diagnostic purposes [3], and solving the problem of unwanted motion during image acquisition is an important goal in scientific and medical MRI [4]. Although many methods to cope with motion have been proposed, no universally accepted solution currently exists (see [5] and [6] for an overview of the current state of the art).

Being able to automatically detect whether movement of a scanned subject occurred during image acquisition has considerable value. Depending on the outcome (i.e. motion was detected), another scan might be performed or the data might be discarded if it could compromise the conclusions of a scientific study. Motion artifacts can be subtle, and they are not necessarily prominent in all slices of the acquired volume. Consequently, the technician or the researcher cannot always visually identify them in situ.

During the last decade significant progress has been made in developing the means for automatic detection and estimation of motion. The existing approaches rely on navigators [7] [8], optical tracking cameras [9], analyzing the air background [10] or sequences with special motion-sensitive sampling patterns [11]. State-of-the-art methods are robust and accurate. Not only can they reveal whether motion occurred, but they can also reliably predict the exact motion trajectory. Such predictions can be used for correcting the motion prospectively or retrospectively. However, the methods rely on the use of special equipment or imaging sequence modifications. In this work we attempt to develop an automated method that can classify a scan into "motion/no motion" categories based solely on the image data output by the scanner. Importantly, we consider the intra-scan motion rather than the inter-scan motion, which is a different problem that results in misregistration of multiple MR volumes.

Recently, deep neural networks have been shown to be highly efficient in tackling many computer vision challenges, such as object detection [12] and image processing problems such as deblurring [13]. In this work, we present an algorithm based on convolutional neural networks (CNNs) that enables fully automated detection of motion artifacts in MR images. We evaluate the performance of different variations of our algorithm and show that our proposed method is both reliable and flexible, generalizing well to different types of MR images. We use a binary "motion" or "no motion" classification scheme because quantitative description of motion artifacts is ambiguous for the different types of motion, translations and rotations and the different types of images considered in our study.

2. EXPERIMENTAL SETUP

2.1. Technical Setup

For implementation, we used the Caffe [14] deep neural network toolbox. As shown in Fig. 1 the implemented network consists of two convolutional and two dense layers. To prevent the network from overfitting, a dropout layer has been

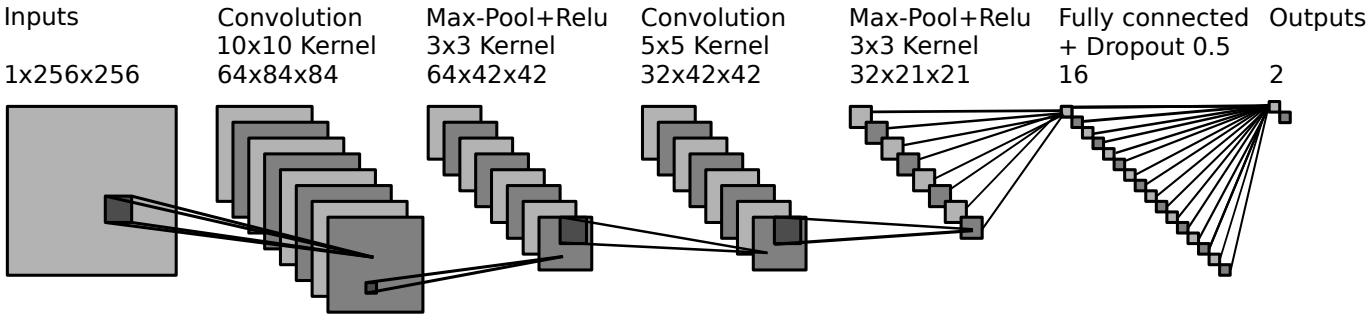

Fig. 1: Our proposed network architecture consists of two sequential convolutional layers interleaved with RELU and pooling layers. The convolutional layers are followed by two fully connected layers, a dropout layer to prevent overfitting and an output inner product layer. To measure the performance of the network, an accuracy layer was attached to the output neurons.

added [15].

All images were resized to a resolution of 256×256 pixels and zero-centred by mean-subtraction. To ensure balanced training of our binary estimator, equal numbers of images with and without motion artifacts were used. The experimenters labeled all images with "motion" or "no motion". We increased the amount of training data by chopping each 3D volume into 2D slices of transversal, coronal and sagittal sections. This procedure increased the amount of available training data by a factor of 300 for the brain scans and by a factor of 64 for the fruit scans. By training and testing our model on volume slices, we obtained slice-specific labels for use when processing entire MR volumes.

The testing and training were performed on a Nvidia Tesla K20Xm graphics card. Training a network with 6000 images until convergence took one hour. A batch size of 50 images was used, and the entire test set was passed through the network every 500 iterations. The initial learning rate was set to 0.1 with a step size of 10,000 images. As shown in Fig. 2, we compared the vanilla stochastic gradient descent solver with AdaGrad [16] and AdaDelta [17] techniques. We observed superior performance for AdaDelta, which we ultimately used in our experiments.

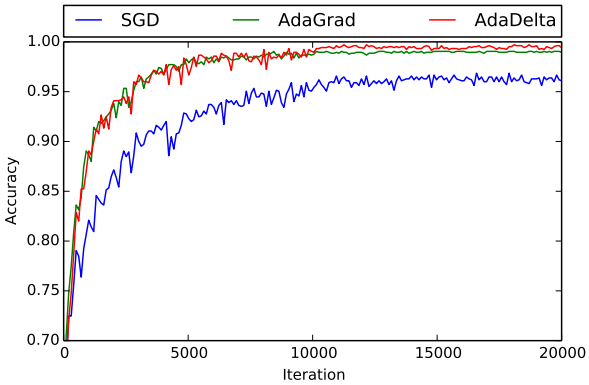

Fig. 2: Comparison of performance of different solvers on the fruit dataset. Stochastic gradient descent showed the worst performance, whereas both AdaGrad and AdaDelta achieved similarly high accuracy, with AdaDelta performing slightly better.

2.2. Dataset Setup

To evaluate the ability of our network to generalize over different object morphologies and structures, we acquired a *fruit* dataset, which consisted of tomatoes, kiwis, lemons, pineapples, mangos and kohlrabies (see Fig. 3 for representative images). Each image of the fruit dataset was acquired with the 3D FLASH sequence and two echo times, 4 ms and 8 ms, to test the generalization ability to different MR contrasts. The dataset was composed of equal numbers of motion-free and motion-corrupted images. To induce the motion the experimenter stood near the scanner bore during the acquisition and used a hand-powered external actuator to displace the fruit inside the coil while the image was being acquired. In total, 2400 fruit images were recorded. To make even more data available during training and make the algorithm agnostic against possible rotational pose changes of the scanned object, we augmented the data by rotating the acquired images by 90, 180 and 270 degrees in a post-processing step. The dataset was split into 90% training and 10% test data segments.

We also tested the sensitivity of the network to motion artifacts with different levels of intensity. The experiment was performed on a synthetic motion dataset, which was generated by taking 2700 motion-free complex-valued MR images of a human brain and using a forward model of the degradation process to simulate motion artifacts; for details on the motion model see [18]. In each experiment both translational and rotational motion was induced. Both types of motion were prescribed by generating uniformly distributed random displacement vectors. The length was chosen according to the number of phase encode steps from the range of $[-S : S]$ pixels for translational motion. For rotational motion we used the

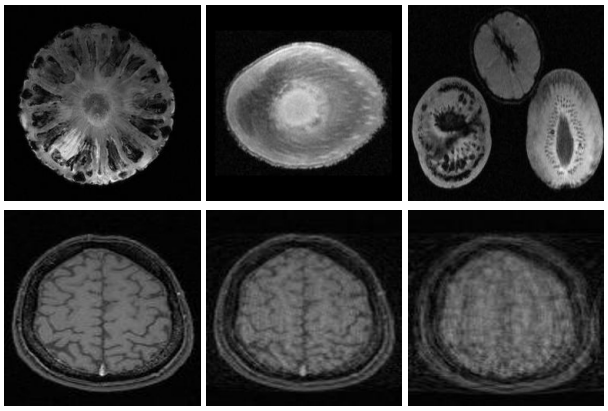

Fig. 3: First row: Fruit images used for network training. From left to right: pineapple, avocado, tomato, lemon and kiwi. Second row: motion corrupted the brain scan; motion intensity was increased from left to right.

frequency domain line segments in the range of $[-S : S] \cdot \frac{\pi}{180}$ radians as length. The "no motion" case corresponded to $S = 0$. As shown in Tbl. 1, the amplitudes used were in the range between 0.05 and 3. In Fig. 3 we show three representative images of synthetically generated motion data.

The training and evaluation were performed in two experimental phases. In the first phase, we trained the network with a fixed amplitude and tested the performance on the validation data generated with the same motion amplitude. In the second phase, we trained the network on motion-corrupted examples with a varying motion amplitude. We then tested it on data generated with pure amplitude levels to assess how well the network was able to generalize. In theory, training the network on synthetic data offsets the problem of overfitting because it is possible to generate nearly an infinite number of different training examples. It is important to emphasize that the model that we used to generate the synthetic motion did not involve approximations and was faithful to the true motion degradation process. In our last set of experiments we trained

Trained on amplitude S	**Acc. on val. set**	**Acc. on mixed Amp.**
0.05	0.568	0.648
0.1	0.702	0.592
0.15	0.610	0.722
0.2	0.713	0.843
0.25	0.923	0.907
0.3	0.972	0.948
3	0.985	0.953
mixed Ampl.	0.983	-

Table 1: Performance evaluation based on the accuracy of the network trained and tested on the artificial motion data.

the algorithm on a dataset composed of brain images affected by real subject motion. The dataset consisted of 6000 images and was again split into 90% training and 10% test data. We checked the performance and flexibility of the trained networks by conducting a cross-evaluation on the different types of data. For example, we trained the network on the brain images and tested it on fruit images and vice versa. By doing so we were able to test the out-of-domain performance of a network; that is, we tested the network on data featuring image structures that it did not "see" during training. Additionally, we conducted an experiment with mixed training data; for example, we added 10% of the fruit images to the brain dataset and 10% of the brain scans to the fruit dataset.

3. RESULTS

We first examine the results obtained on the artificial motion dataset, shown in Tbl. 1, and observe that increasing the strength of artificial motion improved the classification accuracy. The network achieved a nearly perfect classification performance at an amplitude of 0.3, which corresponds to motion artifacts barely visible to a human observer. Increasing the amplitude even further did not affect the performance, which remained close to 100% accuracy. The performance of the network trained on data with motion of mixed amplitudes was similar even though this case was more challenging to train and generalize to. Interestingly, training the network on a single amplitude level and then testing it on mixed data also led to highly accurate predictions, 0.948 for S = 0.3, which suggests that the network was able to pick up and capture certain aspects of motion artifacts that are invariant to the motion intensity. In Fig. 4 we show the learned filters extracted from the first convolutional layer. Some of them can be identified as edge detectors, while others exhibit a more intricate structure that is more difficult to interpret. For real motion data we obtained a similar performance. When we trained and tested the network either on fruit or brain images only, we got nearly 100% accuracy (see Tbl. 2). The network trained on brain scans performed slightly better than chance level on the fruit dataset, 0.56 accuracy. Even when 10% of the fruit scans were added to the training, the performance did not increase substantially. However, the network trained on fruit images initially performed with 0.60 accuracy on brain images. The performance was even better if brain scans were added, yielding an accuracy of 0.86.

The performance of the network of fruits mixed with brains suggests that the network is able, in a limited way, to detect motion features invariant to object shape. The poor performance of the network trained on brain images, even with the addition of 10% fruits, evaluated on fruit images may be explained by the heterogeneity of the fruit dataset (cf Sec. 2.2).

Further examinations of the trained network on the artificial network is presented in Tbl. 3. The results indicate that the network trained on motion with mixed amplitudes had an accuracy of 0.90 and 0.56, respectively, when tested on the

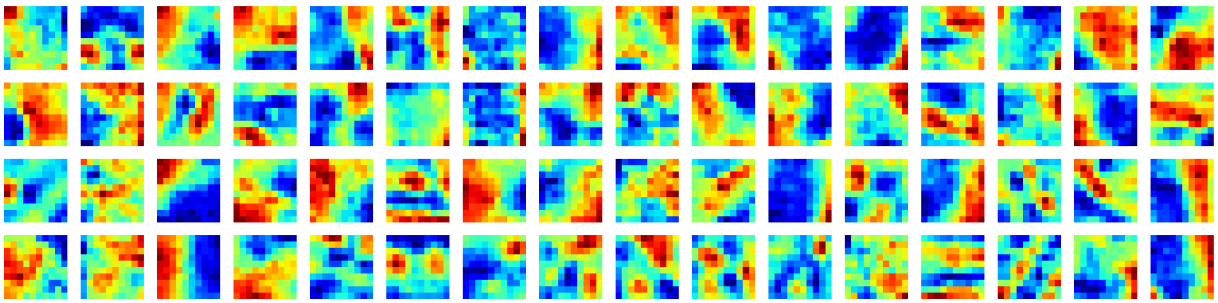

Fig. 4: Visualization of the entire set of 64 filters extracted from the first convolutional layer of our proposed network.

Trained on	Tested on	
	Brains	Fruits
Brains	0.970	0.559
Brains and fruits	0.965	0.580
Fruits	0.598	0.966
Fruits and brains	0.858	0.945

Table 2: Cross-data performance evaluation on brain vs. fruit images.

Trained on	Tested on	
	Brains	Fruits
Art. Motion	0.901	0.559

Table 3: The performance of the network trained on synthetic motion data and tested on real motion data.

real-world brain and fruit dataset. This finding is another indicator that the network is capable of learning the intrinsic features of motion artifacts. In our next experiment, we tested the generalization capability of the network when images were acquired with different contrasts. As shown in Tbl. 4 we used two different echo times. Training and testing on the same contrast resulted in good performance with high levels of accuracy. However, when we conducted the cross-contrast test, the motion detecting network had reduced accuracy. This outcome suggests that training the network on images acquired with different contrasts is important for enabling generalization.

We finish the description of our experimental results by reporting the computation times. The forward pass took approximately 1.34 ms, whereas the backward pass required 1.29 ms. Hence, the network can process an entire brain scan in less than 500 ms. This processing time allows making automated decisions on accepting or rejecting a recorded volumetric image immediately after the scan is over.

Trained on Contrast	Tested on Contrast	
	TE4	TE8
TE4	0.999	0.845
TE8	0.783	0.997

Table 4: Cross-testing the network on images featuring two different contrast levels as determined by echo times used during MR image acquisition.

4. DISCUSSION AND CONCLUSION

In this work, we trained our neural networks and made predictions based on 2D slices rather than 3D volumes. We had a twofold rationale for this approach. First, we wanted to generate as much training and test data as possible by converting a single 3D volume to possibly hundreds of 2D slices. Second, given a 3D volume as input, we could always make predictions from a subset of slices extracted from the volume, finally making a weighted decision whether an input 3D volume was affected by motion or not. In our experiments, we observed that training the network on images acquired with different sequence parameter settings is important in order to achieve good generalization to all classes of data. This would potentially require constructing a large database of scans acquired with actual sequences and instrumental parameters, which is a difficult and time-consuming task.

In its current state, our method can be used for automated decision making whenever a need exists to flag a certain MR dataset as being affected by motion. This flagging can be important, for example, when processing and analysing large databases of MR data in a fully automated way. The technique can also be used to identify a scan as being motion-corrupted right after acquisition. In a case in which high-quality artifact-free MR images are required for medical diagnosis, our system may advise reacquisition if motion artifacts are detected. Future work might involve extending our method to fine-graded predictions (i.e. motion strength) and predicting underlying motion parameters (e.g. 6 degrees of freedom for 3D).

Our results indicate that a learning-based approach to detecting motion artifacts is promising. The network is able to generalize to unseen data and detect motion with high accuracy. The approach can potentially be used in clinical practice to tag scans as motion-free or motion-corrupted immediately.

5. REFERENCES

[1] M.L. Wood and R.M. Henkelman, "MR image artifacts from periodic motion," *Medical Physics*, vol. 12, no. 2, pp. 143–151, 1985.

[2] S. Heiland, "From a as in aliasing to z as in zipper: Artifacts in MRI," *Clinical Neuroradiology*, vol. 18, no. 1, pp. 25–36, 2008.

[3] J. B. Andre, B. W. Bresnahan, M. Mossa-Basha, M. N. Hoff, C. P. Smith, Y. Anzai, and W. A. Cohen, "Toward quantifying the prevalence, severity, andcost associated with patient motion duringclinical MR examinations," *Journal of the American College of Radiology*, vol. 12, no. 7, pp. 689 – 695, 2015.

[4] T.B. Smith and K.S. Nayak, "MRI artifacts and correction strategies," *Imaging in Medicine*, vol. 2, no. 4, pp. 445–457, 2010.

[5] M. Zaitsev, J. Maclaren, and M. Herbst, "Motion artifacts in MRI: A complex problem with many partial solutions," *Journal of Magnetic Resonance Imaging*, vol. 42, no. 4, pp. 887–901, 2015.

[6] F. Godenschweger, U. Kgebein, D. Stucht, U. Yarach, A. Sciarra, R. Yakupov, F. Lsebrink, P. Schulze, and O. Speck, "Motion correction in MRI of the brain," *Physics in Medicine and Biology*, vol. 61, no. 5, pp. R32–R56, 2016.

[7] K. P.N. Forbes, J. G. Pipe, C. R. Bird, and J. E. Heiserman, "Propeller MRI: clinical testing of a novel technique for quantification and compensation of head motion," *Journal of Magnetic Resonance Imaging*, vol. 14, no. 3, pp. 215–222, 2001.

[8] T. Kober, J. P. Marques, R. Gruetter, and G. Krueger, "Head motion detection using fid navigators," *Magnetic resonance in medicine*, vol. 66, no. 1, pp. 135–143, 2011.

[9] M. Zaitsev, C. Dold, G. Sakas, J. Hennig, and O. Speck, "Magnetic Resonance imaging of freely moving objects: prospective real-time motion correction using an external optical motion tracking system," *Neuroimage*, vol. 31, no. 3, pp. 1038–1050, 2006.

[10] Bénédicte Mortamet, Matt A Bernstein, Clifford R Jack, Jeffrey L Gunter, Chadwick Ward, Paula J Britson, Reto Meuli, Jean-Philippe Thiran, and Gunnar Krueger, "Automatic quality assessment in structural brain magnetic resonance imaging," *Magnetic resonance in medicine*, vol. 62, no. 2, pp. 365–372, 2009.

[11] P. J. Bones, J. R. Maclaren, R. P. Millane, and R. Watts, "Quantifying and correcting motion artifacts in MRI," 2006, vol. 6316 of *Proc. SPIE*, pp. 631608–631608–12.

[12] A. Krizhevsky, I. Sutskever, and G.E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, vol. 2, pp. 1097–1105.

[13] C. J. Schuler, M. Hirsch, S. Harmeling, and B. Schölkopf, "Learning to deblur," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 7, pp. 1439–1451, 2016.

[14] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, "Caffe: Convolutional architecture for fast feature embedding," *arXiv preprint arXiv:1408.5093*, 2014.

[15] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *Journal of Machine Learning Research*, vol. 15, pp. 1929–1958, 2014.

[16] M. D. Zeiler, "ADADELTA: an adaptive learning rate method," *CoRR*, vol. abs/1212.5701, 2012.

[17] J. Duchi, E. Hazan, and Y. Singer, "Adaptive subgradient methods for online learning and stochastic optimization," *J. Mach. Learn. Res.*, vol. 12, pp. 2121–2159, July 2011.

[18] A. Loktyushin, H. Nickisch, R. Pohmann, and B. Schölkopf, "Blind retrospective motion correction of MR images," *Magnetic Resonance in Medicine (MRM)*, vol. 70, no. 6, pp. 16081618, 2013.