

AUTOMATIC MOTION ARTIFACT DETECTION FOR WHOLE-BODY MAGNETIC RESONANCE IMAGING

Thomas Küstner^{*†}, Marvin Jandt^{*}, Annika Liebgott^{*†}, Lukas Mauch^{*}, Petros Martirosian[†],
Fabian Bamberg[†], Konstantin Nikolaou[†], Sergios Gatidis[†], Fritz Schick[†], Bin Yang^{*}

^{*} Institute of Signal Processing and System Theory, University of Stuttgart, Stuttgart, Germany

[†] Department of Radiology, University of Tübingen, Tübingen, Germany

ABSTRACT

Magnetic resonance (MR) plays an important role in medical imaging. It can be flexibly tuned towards different applications for deriving a meaningful diagnosis. However, its long acquisition times and flexible parametrization make it on the other hand prone to artifacts which obscure the underlying image content or can be misinterpreted as anatomy. Patient-induced motion artifacts are still one of the major extrinsic factors which degrade image quality. In this work, an automatic reference-free motion artifact detection, including localization and quantification, is proposed which can be used prospectively as quality control (e.g. scan adjustment) or retrospectively as quality control (e.g. supported diagnosis). The detection is achieved via trained convolutional neural networks (CNN). This study focuses on investigating the optimal CNN architecture and required training set composition to derive a general and robust network for MR motion artifact detection. In a volunteer cohort an average accuracy of 91% was achieved.

Index Terms— machine-learning, magnetic resonance imaging, artifacts, neural networks, quality assessment

1. INTRODUCTION

Magnetic resonance imaging (MRI) is a widely used imaging modality in today's clinical diagnostic. It allows precise and non-invasive assessment of anatomical structures as well as physiological and functional processes. Besides various advantages, magnetic resonance (MR) images are susceptible to artifacts originating from hardware imperfections, applied signal processing or patient variabilities. Especially patient-induced motion artifacts are one of the major extrinsic factors which can strongly degrade image quality. The artifacts manifest in the image as shifted and aliased structures along the phase-encoding direction as well as blurring of the image content. Amongst these motion artifacts respiration in the body trunk and head movement in neuroimaging are common and severe sources. These types of motion differ due to the occurring deformations: respiration causes non-rigid displacements, i.e. structures change size and shape while displaced, and head movement induces rigid displacements, i.e. artifacts are shifted and rotated versions of original structures.

In order to guarantee sufficient data quality, it is of importance to detect and correctly classify causes of image deterioration as early as possible. This enables to seize appropriate countermeasures: automatic sequence adaptation, prospective motion correction [1, 2, 3] or motion-robust MR sequences allowing remarkable reduction of motion-induced artifacts as well as retrospective motion correction [4, 5, 6, 7] and learned correction mechanisms from the data itself

[8]. Routinely, images are inspected by a human MR specialist to ensure an acceptable level of data quality. This manual process can be very time-consuming and cost-intensive. Moreover, the quality check is often performed after the patient has already left the facility which may even demand an additional examination. In the context of large epidemiological cohort studies such as UK Biobank [9] or German National Cohort [10], the extent and complexity of acquired data make a manual or visual analysis practically impossible.

Thus, in certain situations, when a human expert is not present, a motion correction procedure is not applicable/available or in the case of large cohort studies, the potential presence of motion artifacts demands an automatic processing for a prospective quality assurance or retrospective quality assessment. In a prospective manner, a direct feedback to the scanner about the derived image quality including artifact level enables to optimize imaging protocols and provides a guided scanner operation. A retrospective quality assessment ensures correct image processing, post-processing or analysis and may thereby even provide support in the diagnostic decision.

In previous studies, several approaches for automated analysis of medical image quality have been proposed [11, 12, 13]. These methods assess image quality relative to an available reference image of optimal quality using difference metrics. These studies focused on assessing quality depending on varying acquisition or reconstruction conditions and are not explicitly considering image artifacts. First reference-free methods were investigated for MR imaging of the head. Mortamet et al. [14] described a simple approach for general quality evaluation of brain MR images by analysis of the signal distribution in the image background. Woodard et al. [15] used image features describing signal-to-noise ratio (SNR), image sharpness and image homogeneity in order to establish an automated assessment of image quality. The work of Tisdall and Atkins [16] compared an automated model observer with human observers for the task of quality evaluation in low-SNR scenarios. Atkinson et al. [8] proposed an auto-focusing metric which was further investigated by McGee et al. [17]. The idea is to minimize in a reference-free setting the self-dissimilarity of the image and by thus correcting motion-induced artifacts. However, none of the mentioned methods focused on motion quantification and localization. Recently, Iglesias et al. [18] proposed CNNs for the detection of motion artifacts in neuro-imaging cases.

Thus, our study aims to provide a localized information about the quantity of motion artifact burden. We proposed a machine-learning scheme in terms of a convolutional neural network (CNN) to learn the representation of motion artifacts [19, 20]. The initial study focused on investigations of motion artifacts in the head and abdomen from T1 weighted MR images. For each body region, separate CNNs consisting of a fairly simple and shallow network architec-

ture were trained on 2D image patches. In this work, we investigate more advanced network architectures and propose a new CNN for MR motion artifact quantification and localization in a whole-body setting. The new CNN is trained on 3D image patches to capture also through-slice motion artifacts. Moreover, the influences of body region and T1/T2 contrast weighting are examined to propose a general CNN architecture as well as the required composition of the training dataset for a reliable prediction.

2. MOTION ARTIFACT DETECTION

This study proposes an automated reference-free quantification and localization of motion artifacts in whole-body MR images. Motion artifact manifestation differs according to: body region (head, abdomen, pelvis), type of motion artifacts (rigid and non-rigid), motion origin (instructed and involuntary) and contrast weighting (T1 and T2). The training database composition defines the identifiable manifestations of motion artifacts and thus needs to be carefully chosen. After processing, a localized output with quantitative scale from 0 to 1 indicating none to high significance of motion artifact presence is desired.

2.1. MR data acquisition

Data was acquired on a 3T PET/MR (Biograph mMR, Siemens Healthineers) from 18 healthy volunteers (3 female, 25 ± 8 years). The study was approved by the local ethics committee and all volunteers gave written consent.

A T1 weighted axial 2D fast spin echo (FSE) sequence was acquired in the head, abdomen and pelvis. A T2 weighted axial FSE was acquired in the abdomen and pelvis. The head region was omitted because of scan time and specific absorption rate limitations. The acquisition parameters are depicted in Table 1. Field of view placement was fairly similar amongst volunteers with subject-specific matrix size adaptation along phase-encoding direction. Each of the five sequences was acquired twice for every volunteer, yielding in total ten measurements per volunteer. During the first acquisition, the volunteers were asked to hold their heads still for the head acquisition and to lie still in the pelvic examination. In the abdominal case, an end-expiratory breath-hold and navigator-triggering (gating window size 5 mm) was conducted for the T1w and T2w FSE, respectively. During the second acquisition, volunteers were instructed to tilt their heads side-to-side for the head acquisition and to breathe normally for the abdominal case. In the pelvic region, volunteers were instructed to move around randomly during the whole acquisition.

2.2. Data preparation

Images were normalized into an intensity range of 0 to 1 and subsequently partitioned into overlapping patches of size $40 \times 40 \times 10$ (anterior-posterior \times left-right \times superior-inferior; AP \times LR \times SI) with 50% overlap in each dimension. The resulting patches are split into training and test set by leaving out one randomly selected volunteer for testing. The labeling is simplified by deriving all labels directly from the acquisition, i.e. all patches from the motion data set are assumed to display motion. This results in 21168, 23250 and 24030 T1 weighted training patches in the head, abdomen and pelvis, respectively and 46880 and 26340 T2 weighted training patches in the abdomen and pelvis, respectively. All patches are used simultaneously for training to create a general motion artifact network. The usage of image patches allows to spatially resolve motion artifacts and enables a faster computation by the networks.

2.3. CNN architectures

The proposed method utilizes a multilayer CNN to return probability values $p \in [0, 1]$ for the presence of motion artifacts on an image patch level. Three new architectures are proposed and investigated in comparison to the previously published 2D-CNN [19]. The aim is to propose a general and robust network for identifying motion artifacts in 3D whole-body MR image patches.

The first architecture is based on the ideas of the 2D-CNN as proposed in [19]. The network is depicted in Fig. 1(a) and is named 3D-CNN. It consists of three convolutional layers with each having N filter kernels/channels of size $M \times L \times B$ and a rectified linear unit (ReLU) [21] activation function, followed by a fully-connected layer with softmax decision. The filter kernels increase dyadically to provide a multi-resolution approach from coarse to fine-grained structures.

The second architecture is inspired by the MNet [22]. However, the proposed MNetArt architecture as illustrated in Fig. 1(b) differs to the original MNet. The decoding branch (required for back-projecting the segmentation results to image space) is omitted and replaced by a dense layer. The 3D input patches are converted via 3D convolution with zero-padding to 2D patches which are managed internally. Each image slice is stored in a separate channel. This enables a fast and efficient processing in a pseudo-3D way. The network consists of four stages. Each of the first three stages has two convolutional layers with N channels of size $M \times L$ and parametric ReLU (PReLU) [23] followed by a max-pooling layer. Between the convolutions and stages a residual layer concatenates by element-wise addition the convolved and unchanged input. This improves convergence and forwards detailed structures to deeper levels. Dimensionality adjustment is achieved by zero-padding. In the final stage, two convolutions and an intermittent residual layer is completed by a dense output layer with softmax decision.

Table 1. MR data acquisition parameters.

	T1w FSE			T2w FSE	
	head	abdomen	pelvis	abdomen	pelvis
body region					
matrix size	$256 \times 196 \times 40$	$320 \times 192-256 \times 30$	$320 \times 192-240 \times 40$	$384 \times 228-288 \times 32$	$210 \times 192-240 \times 40$
voxel size [mm ³]	$1 \times 1 \times 3$	$1.25 \times 1.25 \times 5$	$1.25 \times 1.25 \times 5$	$1.04 \times 1.04 \times 5$	$1.25 \times 1.25 \times 5$
TE [ms]	8.4	10	11	100	86
TR [ms]	750	800	900	5800	4200
flip angle	140°	150°	160°	120°	150°
bandwidth [Hz/px]	260	190	190	195	200
averages	2	2	2	2	2
phase-encoding direction	left-right	anterior-posterior	anterior-posterior	anterior-posterior	anterior-posterior

The third architecture is inspired by the VNet [24]. Again the proposed VNetArt as illustrated in Fig. 1(c) differs in large parts from its original version. Similar to MNet the decoding branch is replaced by a dense decision layer. The 3D input patches are processed in four stages. In each of the first three stages two convolutional layers with N channels of size $M \times L \times B$ and PReLU activation function are followed by a concatenation layer and a final max-pooling with stride. The concatenation layer acts as a residual path between the stage's input and its convolved version. The fourth stage generates the probability output in a dense layer with softmax activation.

For all architectures, filter kernel parameter ranges were estimated by the Baum-Haussler rule [25] and coefficients were trained by optimizing the categorical cross-entropy function for a given learning rate with ℓ_2 regularization and dropout ratio. The cost function was minimized by an adaptive moment estimation (ADAM) [26] with associated ADAM parameters $\beta_1 = 0.9, \beta_2 = 0.999$ and $\epsilon = 10^{-8}$. Training was conducted for a maximal number of 200 epochs and with an early stopping criterion for a training loss update less than 10^{-3} over the last three epochs. Coefficients were initialized by a Gaussian distributed randomization. A grid search was conducted to optimize learning rate, ℓ_2 regularization and dropout ratios.

3. EXPERIMENTS

The focus lies on investigating the optimal architecture (2D-CNN, 3D-CNN, MNetArt, VNetArt) for varying body regions (head, abdomen, pelvis) with accompanied types of motion and contrast weightings (T1, T2). In a first experiment all networks are trained on the patches of the complete training database (a $40 \times 40 \times 1$ patching was used for the 2D-CNN). Afterwards the impact of contrast weighting is examined in a sense that the networks are only trained on T1 weighted images and tested for T2 weighted images and vice versa. In a final experiment the impact of body regions is investigated by training on two body parts, e.g. head and abdomen, with testing conducted on the remainder, e.g. pelvis, yielding three combinations. Evaluation is conducted via test accuracy, sensitivity (true positive rate; TPR) and specificity (true negative rate; TNR). Stated results reflect average values over all 3D patches and for a testing via a leave-one-out cross-validation over all 18 subjects.

4. RESULTS AND DISCUSSION

Fig. 2 depicts an exemplary subject slice in the abdominal region overlaid with the derived patch probabilities to illustrate the localization and quantification ability. The subject is only slightly moving with aliasing occurring at liver, spleen and background. Whilst 2D-CNN underestimates the motion, VNetArt tends to overestimate the whole volume as motion. However, this arises from the volume-wise instead of patch-wise labeling. This observation is in accordance with the previously reported patch size dependency. It was observed that larger patches (till whole volume) can better quantify motion presence, but loose localization information. Smaller patches tend to over- or underestimate motion regions as too less usable information may be contained in one patch.

As illustrated by the results in Fig. 3, motion artifact detection is feasible with a high accuracy of over 85%. The 3D motion detection (3D-CNN, MNetArt, VNetArt) is thereby superior to the 2D case (2D-CNN). The inclusion of residual pathways, i.e. forwarding feature maps from previous layers, as in MNetArt and VNetArt is even superior than only using convolutions as in 3D-CNN. The pure 3D

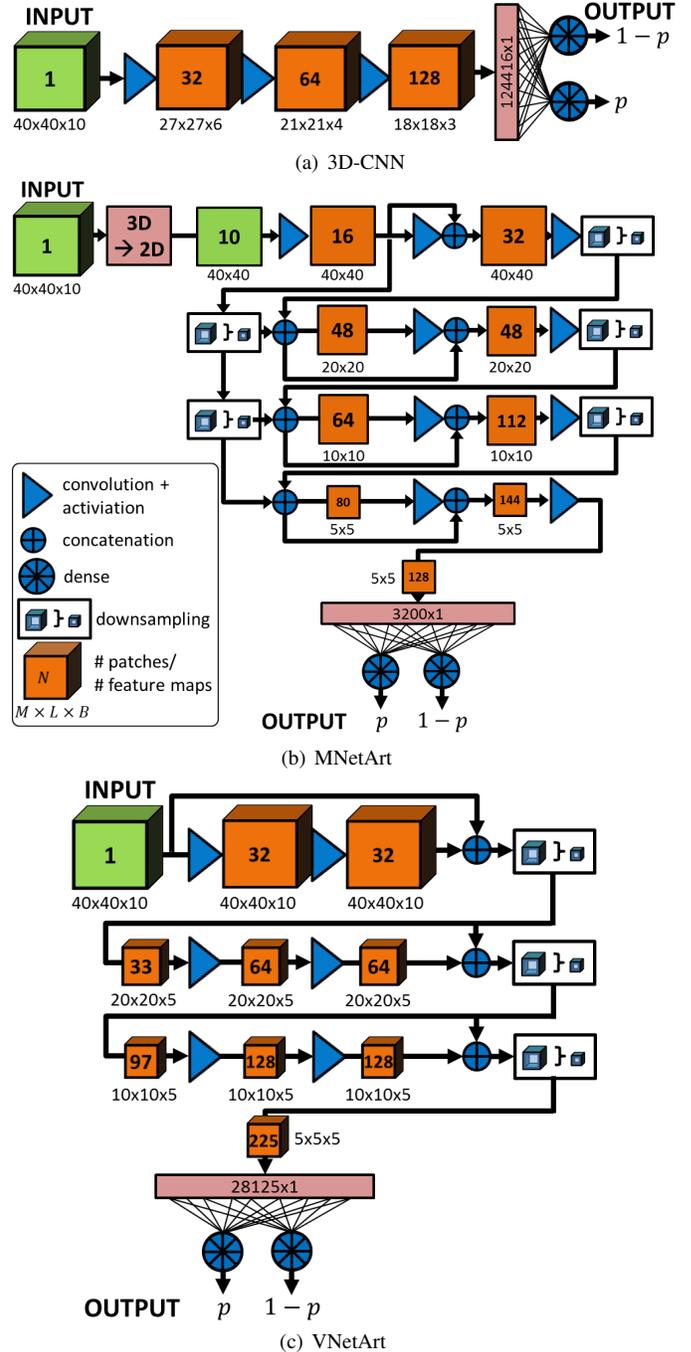


Fig. 1. Proposed CNN architectures for motion artifact detection. 3D image patches are processed to output an artifact probability p on a per-patch level. Respective feature map sizes and channels are stated for the examined input size of $40 \times 40 \times 10$.

processing of the VNetArt performs better than the pseudo-3D processing in MNetArt. Moreover, the larger dense output layer allows a better quantification ability of the VNetArt. From the histograms in Fig. 2, one can conclude the good differentiability, i.e. high specificity of the VNetArt. Sensitivity is mainly determined by the amplitude and strength of motion and is thus data-dependent. Investigations on changing contrast weightings revealed better performance than the complete training set. Comparing between the influence of T1 and T2, a slightly better performance was achieved

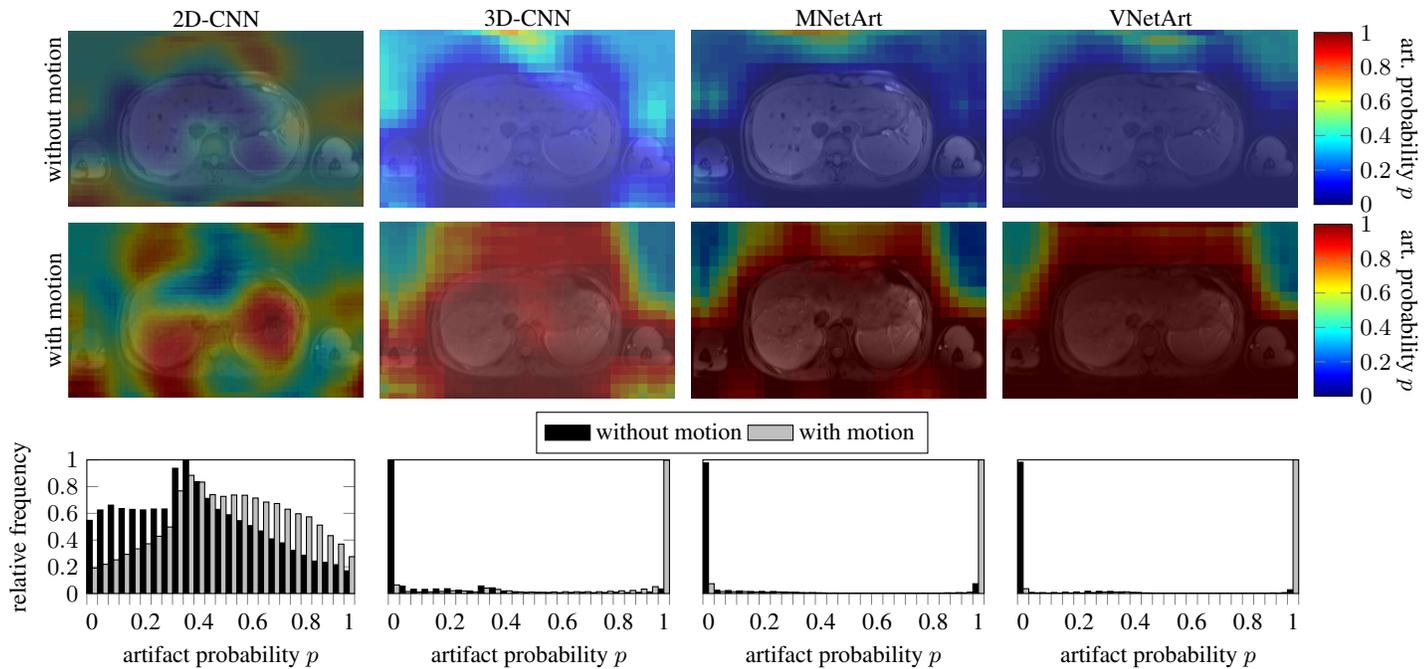


Fig. 2. T1 weighted abdominal MR image patches overlaid by color-coded artifact probability map p . The subject is breathing very shallow. The histograms depict the estimated probability distributions.

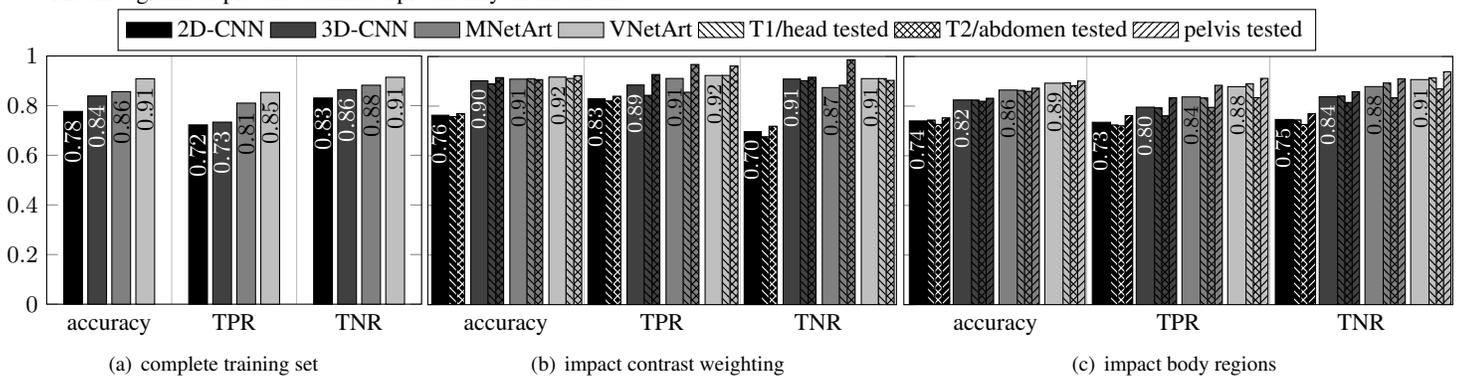


Fig. 3. Average test accuracy, sensitivity (TPR) and specificity (TNR) over all patches and over all cross-validation runs (wide bars) to illustrate impact (a) from complete training set, (b) of contrast weighting (T1, T2) and (c) of body regions (head, abdomen, pelvis). Standard deviations are neglectable small.

if networks were trained on T1 weighted images and tested on T2 because of the bright adipose tissues in which motion artifacts can be more easily identified. This allows to conclude that varying imaging contrasts can be well dealt with by the networks. The impact of body regions shows that metrics decrease if a certain kind of artifact (rigid or non-rigid) is missed during training. If e.g. trained on head (rigid, LR) and abdominal (non-rigid, AP) patches, rigid displacements in the pelvic region along AP direction can be still acceptable identified. However, the orientation of the motion artifact mainly determines the performance. Data augmentation via rotated patches may cope with this problem. Training database selection and composition mainly dominate at the moment the networks' performance, because of the limited amount of data which will be increased in future studies. Amongst all networks, VNetArt provides highest sensitivity and specificity for all experiments with the gained quantitative ability paid at the price of losing localization accuracy. This issue will be resolved in future analysis by a patch-wise labeling. Our study has limitations: The observed motion artifacts were ex-

amined in healthy volunteers which can differ in their appearance to real patients. CNNs are trained on magnitude data (neglecting any motion information in phase) because of their intended application to cohort studies which only provide magnitude images. However, further investigations considering the phase can be of interest.

5. CONCLUSION

Motion artifact detection including quantification and localization in MR images is feasible with high sensitivity (0.85) and specificity (0.91) in different body regions (head, abdomen, pelvis) and from various imaging contrasts (T1, T2). The proposed VNetArt achieves an accuracy of 91% and improves the performance of the previously published network by 17% for a whole-body setting. The proposed method is of potential interest for automated prospective image quality assurance or retrospective quality control. Once trained, these networks have learned a representation of the artifacts and may also serve as part of a correction scheme. This ensures a high data validity and by that reducing imaging costs.

6. REFERENCES

- [1] M. Zaitsev, J. Maclaren, and M. Herbst, "Motion artifacts in MRI: A complex problem with many partial solutions," *J. Magn. Reson. Imaging.*, vol. 42, no. 4, pp. 887–901, 2015.
- [2] J. Maclaren, M. Herbst, O. Speck, and M. Zaitsev, "Prospective motion correction in brain imaging: a review," *Magn. Reson. Med.*, vol. 69, no. 3, pp. 621–36, 2013.
- [3] F. Godenschweger, U. Kagebein, D. Stucht, U. Yarach, A. Sciarra, R. Yakupov, F. Lusebrink, P. Schulze, and O. Speck, "Motion correction in MRI of the brain," *Phys. Med. Biol.*, vol. 61, no. 5, pp. R32–56, 2016.
- [4] L. Feng, R. Grimm, K. T. Block, H. Chandarana, S. Kim, J. Xu, L. Axel, D. K. Sodickson, and R. Otazo, "Golden-angle radial sparse parallel MRI: combination of compressed sensing, parallel imaging, and golden-angle radial sampling for fast and flexible dynamic volumetric MRI," *Magn. Reson. Med.*, vol. 72, no. 3, pp. 707–17, 2014.
- [5] J. Y. Cheng, T. Zhang, N. Ruangwattanapaisarn, M. T. Alley, M. Uecker, J. M. Pauly, M. Lustig, and S. S. Vasanawala, "Free-breathing pediatric MRI with nonrigid motion correction and acceleration," *J. Magn. Reson. Imaging.*, vol. 42, no. 2, pp. 407–20, 2015.
- [6] G. Cruz, D. Atkinson, C. Buerger, T. Schaeffter, and C. Prieto, "Accelerated motion corrected three-dimensional abdominal MRI using total variation regularized SENSE reconstruction," *Magn. Reson. Med.*, vol. 75, no. 4, pp. 1484–98, 2016.
- [7] T. Küstner, C. Würslin, M. Schwartz, P. Martirosian, S. Gatidis, C. Brendle, F. Seith, F. Schick, N.F. Schwenzer, B. Yang, and H. Schmidt, "Self-navigated 4D cartesian imaging of periodic motion in the body trunk using partial k-space compressed sensing," *Magn. Reson. Med.*, vol. 78, no. 2, pp. 632–644, 2017.
- [8] D. Atkinson, D. L. Hill, P. N. Stoyale, P. E. Summers, and S. F. Keevil, "Automatic correction of motion artifacts in magnetic resonance images using an entropy focus criterion," *IEEE Trans. Med. Imaging.*, vol. 16, no. 6, pp. 903–10, 1997.
- [9] W. Ollier, T. Sprosen, and T. Peakman, "UK Biobank: from concept to reality," *Pharmacogenomics*, vol. 6, no. 6, pp. 639–646, 2005.
- [10] F. Bamberg, H. U. Kauczor, S. Weckbach, C. L. Schlett, M. Forsting, S. C. Ladd, K. H. Greiser, M. A. Weber, J. Schulz-Menger, T. Niendorf, T. Pischon, S. Caspers, K. Amunts, K. Berger, R. Bulow, N. Hosten, K. Hegenscheid, T. Kroncke, J. Linseisen, M. Gunther, J. G. Hirsch, A. Kohn, T. Hendel, H. E. Wichmann, B. Schmidt, K. H. Jockel, W. Hoffmann, R. Kaaks, M. F. Reiser, and H. Volzke, "Whole-Body MR Imaging in the German National Cohort: Rationale, Design, and Technical Background," *Radiology*, vol. 277, no. 1, pp. 206–20, 2015.
- [11] J. Oh, S. I. Woolley, T. N. Arvanitis, and J. N. Townend, "A multistage perceptual quality assessment for compressed digital angiogram images," *IEEE Trans. Med. Imaging.*, vol. 20, no. 12, pp. 1352–61, 2001.
- [12] J. Miao, D. Huo, and D. L. Wilson, "Quantitative image quality evaluation of MR images using perceptual difference models," *Med. Phys.*, vol. 35, no. 6, pp. 2541–53, 2008.
- [13] A. Ouled Zaid and B. B. Fradj, "Coronary angiogram video compression for remote browsing and archiving applications," *Comput. Med. Imaging. Graph.*, vol. 34, no. 8, pp. 632–41, 2010.
- [14] B. Mortamet, M. A. Bernstein, Jr. Jack, C. R., J. L. Gunter, C. Ward, P. J. Britson, R. Meuli, J. P. Thiran, and G. Krueger, "Automatic quality assessment in structural brain magnetic resonance imaging," *Magn. Reson. Med.*, vol. 62, no. 2, pp. 365–72, 2009.
- [15] J. P. Woodard and M. P. Carley-Spencer, "No-reference image quality metrics for structural MRI," *Neuroinformatics*, vol. 4, no. 3, pp. 243–62, 2006.
- [16] M. D. Tisdall and M. S. Atkins, "Using human and model performance to compare MRI reconstructions," *IEEE Trans. Med. Imaging.*, vol. 25, no. 11, pp. 1510–7, 2006.
- [17] K. P. McGee, A. Manduca, J. P. Felmlee, S. J. Riederer, and R. L. Ehman, "Image metric-based correction (autocorrection) of motion effects: analysis of image metrics," *J. Magn. Reson. Imaging.*, vol. 11, no. 2, pp. 174–181, 2000.
- [18] Juan Eugenio Iglesias, Garikoitz Lerma-Usabiaga, Luis C. Garcia-Peraza-Herrera, Sara Martinez, and Pedro M. Paz-Alonso, "Retrospective head motion estimation in structural brain mri with 3d cnns," in *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 2017, pp. 314–322.
- [19] T. Küstner, A. Liebgott, L. Mauch, P. Martirosian, F. Bamberg, K. Nikolaou, B. Yang, F. Schick, and S. Gatidis, "Automated reference-free detection of motion artifacts in magnetic resonance images," *Magn. Reson. Mater. Phys., Biol. Med.*, Sep 2017.
- [20] Thomas Küstner, Annika Liebgott, Lukas Mauch, Petros Martirosian, Konstantin Nikolaou, Fritz Schick, Bin Yang, and Sergios Gatidis, "Automatic reference-free detection and quantification of MR image artifacts in human examinations due to motion," in *Proceedings of the International Society for Magnetic Resonance in Medicine (ISMRM)*, Honolulu, Hawaii, USA, Apr. 2017.
- [21] A. Krizhevsky, I. Sutskever, and G.E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in *Advances in Neural Information Processing Systems 25*, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds., pp. 1097–1105. Curran Associates, Inc., 2012.
- [22] R. Mehta and J. Sivaswamy, "M-net: A Convolutional Neural Network for deep brain structure segmentation," in *IEEE International Symposium on Biomedical Imaging (ISBI)*, April 2017, pp. 437–440.
- [23] K. He, X. Zhang, S. Ren, and J. Sun, "Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification," *ArXiv e-prints*, Feb. 2015.
- [24] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation," *ArXiv e-prints*, June 2016.
- [25] E.B. Baum and D. Haussler, "What Size Net Gives Valid Generalization?," *Neural Comput.*, vol. 1, no. 1, pp. 151–160, Mar. 1989.
- [26] D.P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *Computing Research Repository*, vol. 1412.6980 (2014), 2014.