

Deep Learning for Retrospective Motion Correction in MRI: A Comprehensive Review

Veronika Spieker¹, Hannah Eichhorn¹, Kerstin Hammernik¹, Daniel Rueckert¹, *Fellow, IEEE*,
Christine Preibisch¹, Dimitrios C. Karampinos¹, and Julia A. Schnabel¹, *Fellow, IEEE*

Abstract—Motion represents one of the major challenges in magnetic resonance imaging (MRI). Since the MR signal is acquired in frequency space, any motion of the imaged object leads to complex artefacts in the reconstructed image in addition to other MR imaging artefacts. Deep learning has been frequently proposed for motion correction at several stages of the reconstruction process. The wide range of MR acquisition sequences, anatomies and pathologies of interest, and motion patterns (rigid vs. deformable and random vs. regular) makes a comprehensive solution unlikely. To facilitate the transfer of ideas between different applications, this review provides a detailed overview of proposed methods for learning-based motion correction in MRI together with their common challenges and potentials. This review identifies differences and synergies in underlying data usage, architectures, training and evaluation strategies. We critically discuss general trends and outline future directions, with the aim to enhance interaction between different application areas and research fields.

Index Terms—Motion correction, motion compensation, motion artefacts, motion simulation, MRI, deep learning.

I. INTRODUCTION

MOTION remains a major challenge for fully exploiting the diagnostic potential of magnetic resonance imaging (MRI). Whereas MRI stands out as a non-invasive medical imaging modality with excellent soft tissue contrast, its intrinsically long acquisition times make it more susceptible to motion than most other modalities. However, with the fast development of deep learning in recent years, many learning-based *motion correction* (MoCo) methods have been proposed to tackle this challenge in a retrospective, data-driven manner.

Despite the existence of reviews on motion artefacts and classical MoCo [1], [2], no comprehensive overview of learning-based methods for MR motion correction exists so far. Especially an overview of the increasingly popular field of combined MoCo and image reconstruction is missing, which could foster the transfer of deep learning models between applications. Whereas differences in region of interests, acquisition schemes and motion types intrinsically affect data-driven approaches, synergies in underlying models and overall methods need to be identified. In this review, we highlight such differences and synergies at all stages of learning-based motion correction by analysing data usage, architectures, training and evaluation strategies. Furthermore, we intend to generate a general understanding of recent learning-based MoCo approaches in MRI by outlining respective obstacles and potentials and aim to enhance interaction between the fields of machine learning and MRI.

We review published articles that present methodological contributions for learning-based retrospective MoCo in MRI. We searched for articles on PubMed and GoogleScholar until August 2023, using combinations of the keywords “*Motion Correction*”, “*Motion Compensation*”, “*Deep Learning*” and “*Magnetic Resonance Imaging*”, from which we selected the most relevant ones.

The remaining review is structured as follows:

- II *Background*: MR motion artifacts and classical MoCo
- III *Data Availability & Motion Simulation*: Common brain and cardiac/abdominal data strategies
- IV *Architectures*: Image- & k-space-based MoCo methods
- V *Training Objectives*: Training strategies and losses
- VI *Evaluation Metrics*: Image Quality, Motion Detection & Estimation and Downstream Tasks

Manuscript received 23 August 2023; accepted 2 October 2023. Date of publication 13 October 2023; date of current version 2 February 2024. The work of Veronika Spieker and Hannah Eichhorn was supported in part by the Helmholtz Association under the Joint Research School “Munich School for Data Science-MUDS.” (Veronika Spieker and Hannah Eichhorn contributed equally to this work.) (Corresponding author: Julia A. Schnabel.)

Veronika Spieker and Hannah Eichhorn are with the Institute of Machine Learning for Biomedical Imaging, Helmholtz Munich, 85764, Germany, and also with the School of Computation, Information and Technology, Technical University of Munich, 80333 Munich, Germany (e-mail: v.spieker@tum.de; hannah.eichhorn@tum.de).

Kerstin Hammernik is with the Institute of Machine Learning for Biomedical Imaging, Helmholtz Munich, 85764 Munich, Germany, also with the School of Computation, Information and Technology, Technical University of Munich, 80333 Munich, Germany, and also with the Department of Computing, Imperial College London, SW7 2AZ London, U.K. (e-mail: khammernik@tum.de).

Daniel Rueckert is with the School of Computation, Information and Technology and the School of Medicine and Health, Klinikum rechts der Isar, Technical University of Munich, 80333 Munich, Germany, and also with the Department of Computing, Imperial College London, SW7 2AZ London, U.K. (e-mail: daniel.rueckert@tum.de).

Christine Preibisch is with the Department of Neuroradiology, School of Medicine and Health, Klinikum rechts der Isar, Technical University of Munich, 80333 Munich, Germany (e-mail: preibisch@tum.de).

Dimitrios C. Karampinos is with the Department of Diagnostic and Interventional Radiology, School of Medicine and Health, Klinikum rechts der Isar, Technical University of Munich, 80333 Munich, Germany (e-mail: dimitrios.karampinos@tum.de).

Julia A. Schnabel is with the Institute of Machine Learning for Biomedical Imaging, Helmholtz Munich, 85764, Germany, also with the School of Computation, Information and Technology, Technical University of Munich, 80333 Munich, Germany, and also with School of Biomedical Imaging and Imaging Sciences, King’s College London, WC2R 2LS London, U.K. (e-mail: julia.schnabel@tum.de; julia@ieee.org).

Digital Object Identifier 10.1109/TMI.2023.3323215

VII Discussion of sections III, IV, V and VI

VIII Conclusion & Outlook.

II. BACKGROUND

As a basis for understanding learning-based MoCo approaches, we summarize the fundamental principles of MR motion artefacts and classical motion correction in the following section. Please refer to [1] and [2] for more detailed overviews.

Relevant motion during MR image acquisition comprises both moving organs - e.g. due to cardiac motion or respiration - and conscious or unconscious movement of body parts - e.g. due to patient discomfort. Different motion patterns can be observed when imaging different body regions: For brain imaging, the movement of the head is usually assumed to be random and characterized by six rigid-body motion parameters (three rotational and three translational components), commonly neglecting small deformable motion, e.g. due to brain pulsation. In contrast, for abdominal and cardiac imaging the intrinsic movement of organs due to breathing and heartbeat leads to quasi-periodic patterns and deformable, non-rigid motion with significantly more degrees of freedom. For fetal body imaging, next to quasi-periodic motion from the fetus and mother, further vast and unpredictable sudden (non-)rigid motion may occur due to sudden movement of the fetus. Regardless of the exact pattern, motion of the imaged object affects the MR signal, which is acquired in frequency space (or k-space). On the one hand, changes in position disrupt the capability to encode spatial information in the acquired signal. On the other hand, physical MR signal properties are negatively influenced by second-order motion effects, e.g. due to motion-induced magnetic field inhomogeneities or spin history effects. Thus, after reconstructing motion-corrupted data from frequency to image space, complex artefacts may arise, which cannot be corrected in a straightforward process [1]. Exemplary motion-corrupted images for brain and abdominal imaging in Fig. 1A illustrate that motion may hinder successful diagnoses. The versatility of MRI protocols and motion types makes a comprehensive solution unlikely.

Several strategies have been proposed to mitigate motion artefacts. First, subject motion can be constrained physically, for instance by acquiring abdominal scans only during breath hold [4] or using sedation or general anaesthesia when imaging young children [5], [6]. Second, image acquisition schemes have been designed to be more robust towards motion, either by selectively acquiring data in certain motion states or using advanced sampling patterns [7], [8], [9]. Third, accelerated and parallel imaging methods have been introduced, which offer the advantage of shorter acquisition times corresponding to less opportunity for motion events [10], [11].

Next to these mitigation strategies, which are still susceptible to motion artefacts, another group of approaches have been proposed to directly perform *motion correction* by explicitly removing motion artefacts, and *motion compensation* by leveraging the regularity of motion patterns for a better reconstruction. These approaches include prospective methods, which are applied during image acquisition [12], [13], [14], and retrospective methods, which are applied after image acquisition at various points in the reconstruction pipeline [15],

[16]. Retrospective methods must cope with motion-induced image information loss, e.g. due to data inconsistencies in k-space [2]. For this, deep learning methods are particularly promising due to their capability to identify complex patterns in the absence of a complete analytical model.

Note that the following sections focus on motion correction and compensation of motion artefacts that originate in the acquisition process, i.e. the k-space domain, and reconstruction refers to the domain-transfer from k-space to image-space. For strategies targeting slice-to-volume reconstruction (SVR) of highly accelerated (and hence nearly artefact-free) 2D slices, such as dominantly applied for fetal motion correction, we refer the reader to [17].

III. DATA AVAILABILITY AND MOTION SIMULATION

The majority of learning-based MoCo methods rely on supervised training and thus, on the availability of paired data with and without motion artefacts. Even unsupervised or self-supervised approaches use paired data for quantitative performance evaluations (compare Section VI). Some authors acquire pairs of motion-corrupted and ground truth (GT) motion-free images for training and evaluation. However, it is costly and not always feasible to acquire large paired datasets, which is why motion simulations are commonly used.

When simulating motion artefacts, it is important to consider the typical motion patterns of the anatomy of interest, which we described in Section II. In the following, we summarize the common simulation procedures for brain as well as cardiac and abdominal imaging.

A. Brain

The simulation of rigid-body motion follows the MRI forward model in the presence of motion [18]:

$$y = \sum_{t=1}^T \mathbf{M}_t \mathcal{F} \mathbf{U}_t x, \quad (1)$$

where the Fourier transform \mathcal{F} , the sampling mask \mathbf{M}_t and the motion transform \mathbf{U}_t , are applied to the GT image x for each time point t to generate the motion-corrupted k-space y . In the case of rigid-body motion, the motion transform, $\mathbf{U}_t = \mathbf{T}_t \mathbf{R}_t$, consists of rotation and translation transforms, \mathbf{R}_t and \mathbf{T}_t . Additionally, coil sensitivities or second-order motion effects can be included in the forward model to extend the simulation to the specific application.

It is mathematically equivalent to simulate motion in image space or in k-space [19]. As visualized in Fig. 2, simulations in image space are performed by rotating and translating the image and replacing the corresponding k-space lines with the Fourier transform of the transformed image for each time step. Simulations in k-space are based on the properties of the Fourier transform: rotations of the imaged objects correspond to equivalent rotations in k-space and translations \mathbf{T} correspond to multiplications with linear phase ramps depending on the translation parameter a and k-space coordinate in readout direction k_{RO} :

$$\mathbf{T}(y) = y \cdot \exp(-i2\pi a k_{RO}). \quad (2)$$

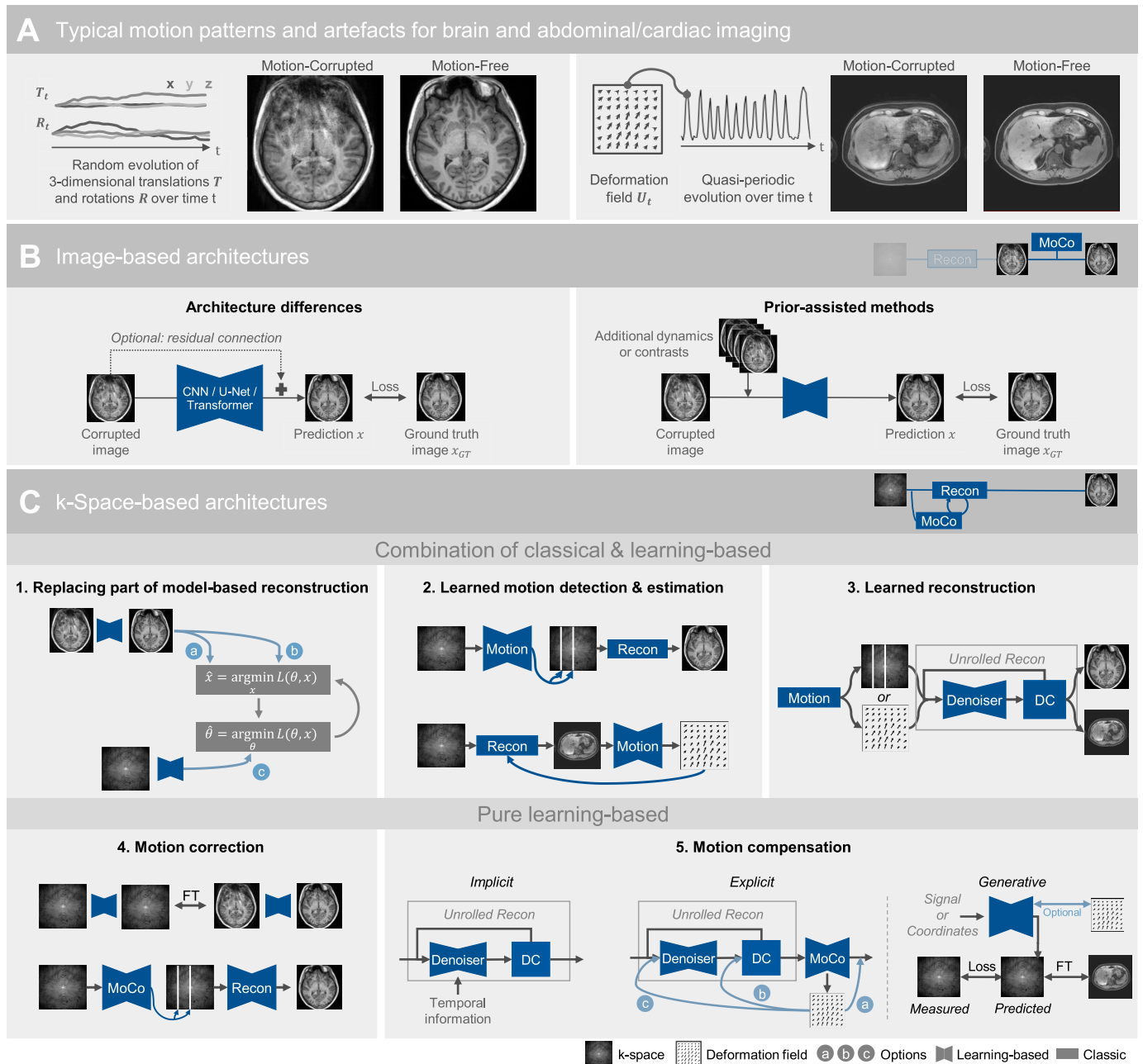


Fig. 1. (A) Typical motion patterns for brain and abdominal/cardiac imaging, together with examples of motion-corrupted and motion-free images (brain images from [3], remaining data acquired at Klinikum Rechts der Isar, Munich). For brain imaging, random rigid-body motion is typically assumed, which results in blurring and ringing artefacts, depending on the exact acquisition scheme and motion pattern. For cardiac and abdominal imaging, motion is typically deformable and quasi-periodic, leading to blurring and possibly, ghosting artefacts. (B) Visualisation of image-based MoCo with motion-corrupted images as input and motion-corrected images as output of a neural network (CNN, U-Net or transformer). Optionally, a residual connection enforces the network to learn artefact maps. Prior-assisted methods incorporate additional information, like additional dynamics or contrasts. (C) Illustration of k-space-based MoCo. Methods that combine classical and learning based modules can be categorized as follows: (1) methods that replace different components of a model-based reconstruction, which iterates between finding an image x and corresponding motion parameters θ by minimizing a loss function, $L(\theta, x)$. Learning-based modules target (a) the image initialization, (b) the loss function or (c) the motion parameters. (2) Combining a classical reconstruction with a learning-based motion detection of corrupted k-space measurements or a learning-based estimation of motion fields, U_t . (3) Combining classical motion detection or estimation with a learned unrolled reconstruction that iterates between denoising networks and data consistency (DC) blocks. Purely learning-based methods include (4) approaches that directly aim to correct motion artefacts by either performing convolutions in k-space and image-space or excluding motion-corrupted k-space measurements from the learning-based reconstruction and (5) motion compensation methods which use motion information implicitly or explicitly to achieve higher quality image reconstructions. A subclass of approaches fit a generative reconstruction model to raw data of an individual, whereas motion can optionally be modelled explicitly by including deformation fields. For all visualizations in (B) and (C) we illustrated the most prevalent anatomy, even though the anatomies are interchangeable in most cases.

Regardless of the domain, in which the simulations are performed, it is important to match the timing of the motion to the MR acquisition scheme to simulate realistic artefacts [20].

B. Cardiac and Abdominal

Following (1), image-based non-rigid motion simulation is achieved by applying a deformable vector field (DVF)

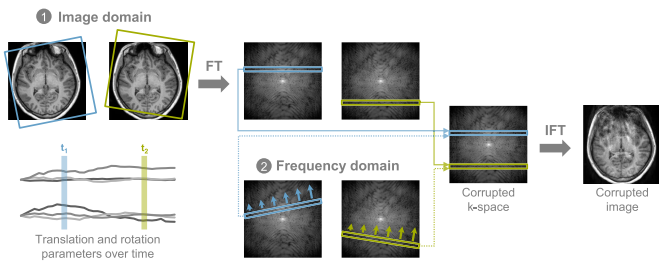


Fig. 2. Rigid-body motion simulation in image or frequency domain, based on x -, y - and z -translation and rotation parameters for each time step. When simulating in image domain (1), translation and rotation parameters are applied to the image. When simulating in frequency domain (2), corresponding k -space lines are rotated and multiplied with linear phase ramps (visualized by arrows). For both, k -space lines of different time steps are merged into one corrupted k -space.

as motion transform U_t . Realistic DVF can be obtained by registering reference images, e.g. from different motion states. Statistical modulation of the DVF allows augmentation for training purposes. If available, multiple time-resolved reconstructions x_t can be used to substitute $U_t x$ and, hence, simulate without DVF.

In contrast to rigid motion, it is not trivial to simulate non-rigid deformation in k -space directly. Therefore, cardiac and respiratory motion simulation can be approximated with varying translations in k -space. To simulate periodic motion, linear phase ramps with a periodically varying translation parameter, i.e. $a(t) \propto \sin(t)$, are applied in (2). It needs to be noted that this is a strongly simplified motion representation.

IV. ARCHITECTURES

This section covers an overview of proposed architectures for learning-based MoCo in MRI. Architectures can be categorized into (A) image-based and (B) k -space-based. Each section includes methods applied to different anatomies and highlights similarities, differences and general trends.

A. Image-Based Motion Correction

Image-based MoCo methods take motion-affected images as input and produce motion-corrected images as output, similar to image denoising or deblurring tasks, as sketched in Fig. 1B. They differ based on their (1) underlying network architecture and (2) potential use of prior information.

1) *Underlying Network Architectures*: The image-to-image translation task can be solved using convolutional neural networks (CNNs), consisting of several consecutive convolutional layers with corresponding activation functions [21], [22], [23], [24]. Some approaches process the input images in multiple resolutions, using varying patch and kernel sizes [21] or dilated convolutions [24]. Similarly, convolutional encoder-decoder structures can be utilized, which consist of a downsampling path for feature encoding, followed by an upsampling path for decoding the extracted features [23], [25], [26], [27], [28], [29], [30], [31], [32], [33], [34], [35], [36], [37], [38], [39], [40], [41], [42]. As common for U-Nets, most methods use skip connections to transfer details from the encoder to the decoder [23], [25], [26], [28], [29], [30], [32], [34], [35], [36], [37], [38], [39], [40], [42]. Additionally, some

approaches cascade multiple encoder-decoder structures consecutively [26], [32], [37], [38]. Further, to exploit temporal dependencies of dynamic data, a recurrent encoder-decoder network has been proposed, replacing convolutions with convolutional long/short-term memory models [43]. Inspired by the recent success of transformer architectures in other fields, initial works employ self-attention mechanisms to address long-distance spatial dependencies of motion artefacts [44], [45].

The above presented approaches vary with respect to whether the inputs of the networks are patches [21], [22], [24], [26], [29], [34], [39], [41] or complete images [23], [25], [27], [28], [30], [31], [32], [33], [35], [37], [38], [40], [42], [43], [44], [45]. The networks are either implemented as classical [23], [25], [27], [28], [29], [30], [31], [32], [33], [34], [35], [37], [38], [39], [40], [41], [43], [44], [45] or residual networks [21], [22], [24], [26], [36], [42], generating corrected images or motion artefact maps, respectively. To further improve the performance, the MoCo task can be combined with learning-based downstream tasks in an end-to-end manner, where both tasks can benefit from each other [30].

2) *Prior-Assisted Methods*: The presented architectures can be modified to take advantage of additional information, like different contrasts [23], [27], multi-echo or multi-parametric acquisitions [24], [33], similar slices [23], [32] or dynamic information [22], [31], [39], [46]. These prior-assisted methods process multiple inputs by either multiple or shared encoders and decoders with shared feature extraction [23], [27], [33], by concatenating the inputs on different channels [22], [23], [24], [31], [32] or by using a recurrent network structure for the additional dimension [46]. For instance, Ghodrati et al. [39] attempt to leverage temporal information by computing a loss on the features extracted by an auxiliary network pretrained on dynamic images. Moreover, dynamic information can be utilized in registration-based methods, where a CNN is used to register binned data into a common space, the combination of which results in a motion-corrected output image [47].

B. K-Space-Based Motion Correction

Contrary to image-based methods, MoCo can also leverage the additional information content of raw k -space data and thus, interact with the MR reconstruction process (see Fig. 1C). Different components of the motion-aware reconstruction pipeline can be learning-based. In the following sections we provide an overview of methods which combine classical and learning-based modules, and methods with pure learning-based modules.

1) *Combination of Classical and Learning-Based Approaches*: Multiple methods extend classical frameworks with individual learning-based MoCo or reconstruction components. A part of a model-based reconstruction, the motion analysis or the reconstruction itself can be learned.

a) *Replacing part of model-based reconstructions*: Model-based MoCo algorithms rely on the joint estimation of motion parameters and the reconstructed image. Various approaches propose to replace different parts of these optimisation procedures with learning-based components to enable faster convergence and ideally, more stable reconstructions

(Fig. 1C.1). Kuzmina et al. [48] use a CNN as part of the loss function for autofocusing, where the optimisation is based on an image quality metric. For data consistency (DC)-based optimisation procedures, CNNs or U-Nets are employed for motion parameter estimation [49], as initialisation of the motion-corrected image [50], [51] or as reconstruction networks whose weights are defined by a hypernetwork that is dependent on the motion parameters [52]. In contrast, Levac et al. [53] propose an unsupervised approach, using a score-based model, that was trained on motion-free images, in the joint estimation of image and motion parameters. All these approaches have in common that the network is pre-trained and used as plug-and-play component during test-time optimisation. Moreover, all these approaches focus on rigid-body motion, having considerably less degrees of freedom than non-rigid motion.

b) Learning-based motion analysis and classical reconstruction: Another group of methods leverage the random nature of rigid-body motion. As visualized in Fig. 1C.2, they learn detection models for motion-affected k-space measurements and inform classical reconstruction procedures with the extracted motion timing. Eichhorn et al. [54] employ a CNN for a line detection in k-space and use these line-wise classification labels as weights in the DC term of a total variation-based reconstruction procedure. Cui et al. [55] train an image-based CNN to correct motion artefacts and compare the k-space of the original and motion-corrected images to generate undersampling masks for motion-affected k-space lines. Undersampled original data are then reconstructed with a classical compressed sensing procedure.

In the case of quasi-periodic motion, the assumption of individual motion-corrupted k-space lines does not apply. Rather than correcting for single motion events, motion compensation methods leverage the periodicity of motion for higher-quality reconstructions of undersampled data. These methods learn motion estimates, which are included in a model-based classical reconstruction [56], [57], [58], [59], [60]. Motion fields are predicted using image-based registration and integrated into the forward operator of the reconstruction problem. Existing approaches vary regarding the registration network's input, i.e. complete image [58], [59], [60] vs. image patches [56], [57], and paired [56], [57] vs. grouped input [58], [59], [60]. Furthermore, the motion estimation network can be pre-trained [56], [57], [58] or optimized jointly with the reconstruction problem [59]. A hybrid approach is proposed in [60], where a motion estimation is obtained with a pre-trained multi-scale network, and consecutively optimized in an iterative reconstruction. Munoz et al. [57] leverage a diffeomorphic registration network to predict forward and backward motion fields in one run rather than individually. A different application of learned motion estimates is presented in [61], where real-time high-quality reconstructions are obtained by deforming a reference image with motion fields predicted from few k-space lines. Whereas registration is conducted in image-space, classical reconstruction layers are used to calculate the loss in k-space.

c) Classical motion analysis and learning-based reconstruction: In contrast, motion detection and estimation can also be performed classically and combined with a learned unrolled

DC-based reconstruction (Fig. 1C.3). Rotman et al. [62] detect discrete motion timings by comparing signals from two opposite coil elements and learn an unrolled reconstruction, in which the regularising network separately receives the data acquired in the dominant and remaining motion states. Miller and Johnson [63] employ a classical spatio-temporally constrained registration of dynamic images to a single motion state. The registered images are encoded and forwarded into an unrolled reconstruction, which is trained in a self-supervised manner by splitting the available data into subsets.

2) Pure Learning-Based Approaches: Compared to the previous section, several methods combine MoCo and image reconstruction in one purely learning-based framework. These can be distinguished by their aim to either correct or compensate for motion.

a) Motion correction: Proposed motion correction approaches explicitly aim to remove motion artefacts in the underlying data (Fig. 1C.4). Singh et al. [64] propose a network consisting of interleaved or alternating convolutions in image and k-space for simultaneous rigid-body MoCo and reconstruction. This approach was further developed into a data consistent method, which we already introduced in section IV-B.1.a [52]. Oksuz et al. [65] realize data consistent reconstructions of cardiac data with ECG mistriggering artefacts in line with the methods presented in section IV-B.1.b. They propose to employ a CNN to learn undersampling masks for motion-affected k-space lines and reconstruct the undersampled data with a recurrent network. In an extension, they train the detection and reconstruction networks end-to-end with a segmentation network and thus, optimize the MoCo specifically for the downstream task of interest [66].

b) Motion compensation: Presented motion compensation methods leverage occurring motion to improve reconstruction results along with accelerated acquisition times, as illustrated in Fig. 1C.5. This can be achieved *implicitly* by including the temporal dimension in the denoising process of an unrolled reconstruction [67], [68], [69], [70]. Spatial and temporal convolutions are applied to dynamic image series either with a joint [67] or separated spatio-temporal kernel, in a cascaded [68] or parallel manner [69]. To leverage further information from adjacent frames, Schlemper et al. [67] include a data sharing layer. Terpstra et al. [69] extend the implicit motion-compensated reconstruction with motion fields obtained from a pretrained model. Qin et al. [70] employ recurrent networks to exploit dependencies along temporal dimensions as well as along stages of the iterative reconstruction.

In contrast, several methods *explicitly* learn the motion model with the reconstruction problem in an end-to-end fashion [71], [72], [73], [74], [75], [76]. Huang et al. [71] append motion estimation and correction modules to a reconstruction network and train the framework with one combined loss function. Others directly feed learned motion estimates into the unrolled reconstruction process, either as input of the denoiser [72], [73] or in the DC layer [74], [75]. Additionally, these methods differ in the way motion estimates were obtained, e.g. using optical flow [72], groupwise registration [73], patch-wise registration [74] or

registration in k-space [75]. A different approach is presented by Gan et al. [76], where a motion estimation network is leveraged to train a reconstruction framework in an unsupervised manner, i.e. by deforming other dynamics for loss calculation. Whereas motion is modelled explicitly during training, it is implicitly represented in the reconstruction network at inference.

Whereas all previously presented methods have been developed for subject-independent inference, a few *generative* motion-aware reconstruction methods train a reconstruction model per subject to infer for that same individual [77], [78], [79], [80], [81], [82]. Due to their distinct training strategy, we consider these methods as a separate category. Still, motion modeling can be implicit and explicit. In particular, quasi-periodic motion can be modeled as a latent manifold and then transformed into dynamic images through a more complex representation [78]. Transformation of the resulting images into Fourier space allows for network optimization in a self-supervised manner. A different motion modelling strategy [77] learns a low-dimensional signal, which is mapped to motion estimates. These are then applied to one learned reference reconstruction. Again, predicted images are compared with the acquired data points in Fourier space. Recently, implicit neural representations (INR) have also gained attention for dynamic MR reconstruction. Based on spatial and temporal coordinates, a light-weight network predicts the corresponding intensity values in image- [80], [81] or k-space [79], [82]. By including the cardiac [79], [80], [81] or respiratory phase [82] as temporal dimension, motion is implicitly modelled and motion-resolved reconstructions can be obtained at inference. Whereas k-space-based INRs [79], [82] can directly be compared with the acquired points, image-based INRs require transformation to k-space, either by applying the non-uniform fast Fourier transform on a fully queried image [80] or taking advantage of the Fourier Slice Theorem for individual spokes [81]. To enable a better spectral representation the proposed approaches apply Fourier [71], [82], spatiotemporal Fourier [81] or Hash encoding [80].

V. TRAINING OBJECTIVES

A. Image-Based Motion Correction

As visualized in Fig. 3A (left), classical network training of image-based MoCo methods in a supervised setting is performed by calculating a voxel intensity-based cost function between the network’s prediction and a ground-truth motion-free image. Typical intensity-based cost functions are the L1 and L2 loss (which stand for the mean absolute and mean squared error, respectively) or the structural similarity index [83]. Please refer to Sec. VI-A for mathematical definitions. Next to these examples, any other image similarity metric can be used as cost function.

A different training objective, however, is employed with conditional generative adversarial networks (GANs), as illustrated in Fig. 3B. A generator network, mapping the motion-corrupted to a motion-free image, is extended with a discriminator network, which aims to distinguish the predicted image from a ground truth image. Several supervised GAN-based methods have been proposed for various anatomies [34], [35], [36], [37], [38], [39]. Next to

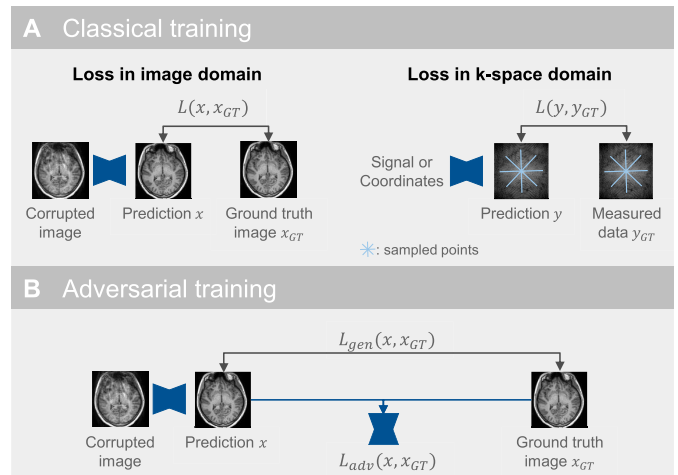


Fig. 3. Visualisation of classical and adversarial training objectives for learning-based MoCo. (A) For classical training, the loss can be calculated in image or k-space domain. In image domain, the predicted image is compared with the ground truth image, e.g. using a voxel intensity-based loss function L . In k-space domain, predicted k-space values are compared with the measured data at the sampled locations. (B) For adversarial training, next to the generative loss L_{gen} , an additional discriminator network is trained to compete with the generator network and distinguish the predicted image from the ground truth (L_{adv}).

the adversarial loss, some of these methods rely on voxel intensity-based cost functions as generator loss to compare the predicted and ground truth image [34], [35]. Others include a perceptual loss [36], [37], [38], style transfer loss [37] or structural similarity loss (SSIM) [39] to account for global changes. Bao et al. [36] propose an additional entropy loss to enhance image homogeneity. Next to the adversarial approach, Küstner et al. [37] present another supervised generative training strategy using a variational autoencoder (VAE), which attempts to learn a motion-free latent distribution directly from the image pair.

To cope with the lack of paired motion-free and corrupted data, unsupervised generative models aim to correct for motion from unpaired data [40], [41], [42], [45]. The CycleGAN architecture consisting of two GANs is adapted in [40] and [45]. Two generators, one corrupting a motion-free image and one correcting an unpaired corrupted image, are trained to invert each other (cycle-transform). Whereas the adversarial loss can be computed with an unpaired image from the other domain, the generative loss is calculated in an unsupervised manner on the cycle-transformed input from the same domain. Liu et al. [40] additionally disentangle the latent representation of the generators into artefact and content information, and train the network with images generated from content-swapped translations. Both [40] and [45] include multi-scale cost functions. In a different setting, Oh et al. [42] treat motion as a probabilistic undersampling problem and train a generator to remove undersampling artefacts. They attempt to correct motion-corrupted measurements by combining repeated randomly undersampled reconstructions. In contrast to CycleGANs, Ghodrati et al. [41] regularize the latent space of a single MoCo autoencoder by applying a discriminator with unpaired motion-free images.

B. K-Space-Based Motion Correction

k-space-based MoCo methods can be trained by comparing the final reconstruction with a ground-truth motion-free image, similarly to image-based methods. However, the loss can also be computed in the acquisition domain directly. By comparing the predicted with the available sampled k-space data, MoCo methods can be trained in a self-supervised manner, as visualised in Fig. 3A (right). The L2 loss is frequently adapted for comparison of the predicted and measured k-space values [63], [77], [78]. Due to the inherent nature of higher magnitudes towards the center of k-space, adaptations such as the L2 loss normalized by the square magnitude [80] or a high dynamic range loss [79] have been proposed, thereby allowing for a more balanced weighting of low and high-frequency components. The k-space based loss can be extended with any further image-based constraint, such as temporal total variation [80] to enforce smoothness between dynamics.

Next to the final reconstruction losses, further motion estimation and detection losses can be incorporated into the training objective. MoCo methods including explicit motion modelling can include an image similarity metric on a spatially transformed image in a self-supervised fashion, as well as spatial or temporal smoothness constraints on the predicted motion field [56], [58]. Models based on motion detection can be trained with any classification loss reflecting the correct identification of motion-affected lines, such as binary cross entropy [54], [66].

VI. EVALUATION METRICS

Since MoCo is performed as a means to an end for a high-quality image reconstruction, the performance of the presented methods is predominantly evaluated based on their final outcome, using image quality measures. However, some authors also evaluate intermediate motion estimates or make use of downstream tasks. In the following, we provide an overview of the most common evaluation strategies. Whereas we focus on evaluation metrics, most of the presented measures can also be used as loss functions, depending on the architecture and the type of training.

A. Image Quality

Quantitative or qualitative image quality evaluation can be performed by calculating image quality metrics or expert image quality rating, respectively. Quantitative image quality metrics can be either full-reference metrics, which assess the image quality by comparison to a GT reference image, or reference-free metrics, which do not rely on a separate GT image. Due to the variability of motion artefacts, no single image quality measure is sensitive to all possible artefacts.

The majority of the methods presented in section IV use two full-reference metrics that attempt to mimic human visual perception: structural similarity index (SSIM) [83], which assesses the degradation of structural information, and peak signal-to-noise ratio (PSNR), which contrasts pixel-wise errors with the maximum signal intensity. Less frequently used full-reference metrics are mean squared error (MSE), root MSE (RMSE), normalized RMSE (NRMSE), mean absolute

TABLE I
DEFINITION OF THE MOST COMMONLY USED
REAL-VALUED IMAGE QUALITY METRICS

Full-reference metrics	
SSIM	$\frac{1}{ \mathcal{M} } \sum_{m, \hat{m} \in \mathcal{M}} \frac{(2\mu_m \mu_{\hat{m}} + c_1)(2\sigma_{m\hat{m}} + c_2)}{(\mu_m^2 + \mu_{\hat{m}}^2 + c_1)(\sigma_m^2 + \sigma_{\hat{m}}^2 + c_2)}$
MSE	$\frac{1}{IJ} \sum_{i=1, j=1}^{I, J} (x_{ij} - \hat{x}_{ij})^2$
PSNR	$10 \log_{10} \frac{\max(\hat{x})^2}{\text{MSE}(x, \hat{x})}$
RMSE	$\sqrt{\text{MSE}(x, \hat{x})}$
NRMSE	$\frac{\text{RMSE}(x, \hat{x})}{c}$, with varying constant, e.g. $c = \frac{1}{ \mu_{\hat{x}} }$
MAE	$\frac{1}{IJ} \sum_{i=1, j=1}^{I, J} x_{ij} - \hat{x}_{ij} $
MAPE	$\frac{\text{MAE}(x, \hat{x})}{\max(\hat{x})}$
NMI	$\frac{H(x) + H(\hat{x})}{H(x, \hat{x})}$, with $H(Z) := -\sum_{z \in Z} z \log z$
VIF	<i>Due to the complexity, we refer the reader to the original publication [85].</i>

Reference-free metrics	
SNR	$20 \log \frac{\mu_s}{\sigma_n}$,
CNR	$20 \log \frac{ \mu_{s1} - \mu_{s2} }{\sigma_n}$,
Tenengrad	$\frac{1}{IJ} \sum_{i=1, j=1}^{I, J} (\nabla_x x_{ij}^2 + \nabla_y x_{ij}^2)$

x : image to be evaluated, \hat{x} : reference image, m/\hat{m} : patch of x/\hat{x} , μ : mean value, σ : standard deviation, $c_1/c_2 \propto L^2$: variables proportional to dynamic range L , n : noise region, s : region of interest, s_1, s_2 : two separate regions in region of interest s , n : noise region, $\nabla_{(x,y)}$: gradient in x - or y -direction

(percentage) error (MAE/MAPE), normalized mutual information (NMI) [84] and visual information fidelity (VIF) [85]. Moreover, reference-free metrics like signal-to-noise ratio (SNR) [86], contrast-to-noise ratio (CNR) and Tenengrad [87] are used to assess image quality without a reference image. Table I provides the definitions of these metrics in a consistent notation. Note that across literature there is no standard normalisation constant for NRMSE. Also, all metrics are applied to real-valued images, whereas there is no clear indication on how to handle complex features.

Several approaches also include qualitative image quality evaluation, i.e. through subjective scoring of the reconstructed images by (blinded) experts [21], [22], [25], [26], [29], [31], [37], [39], [41], [42], [43], [47], [68], [74]. However, there is no standardized way for observer scoring and it varies strongly regarding:

- evaluation categories and instructions for evaluators (e.g. overall quality, sharpness, diagnostic value),
- underlying scale (e.g. three, four, five point scale),
- level of expertise of the evaluators (e.g. radiologist, radiographer, scientist),
- number of evaluators.

B. Motion Detection and Estimation

Motion evaluation strategies can be applied if motion is *explicitly* modeled within the reconstruction framework. Models detecting motion in a line-wise manner resemble classification tasks. Therefore, they can be evaluated with

any classification evaluation metric as long as a ground truth exists. For an overview of classification metrics and their definitions we refer the reader to [88]. Observed metrics specifically applied to MR motion detection tasks include accuracy, sensitivity, precision, recall, F-score and area under ROC curve [28], [55], [89].

Motion estimates can be evaluated reference-based or reference-free. Several methods generate a reference motion field by simulating motion or obtaining the motion field through a distinct registration method. Consecutively, predicted motion parameters are compared using MAE [49] or RSME, which is also termed end-point-error (EPE) when applied to motion fields [56], [74], [75]. A further metric specifically applicable to motion fields is the end-angulation-error (EAE) [75], which computes the angle between the ground truth and predicted motion vector. When no motion estimate is available as reference, predicted motion fields can be used to warp images. Similar to registration evaluation, motion field accuracy can be evaluated by comparing the spatially transformed image with the target image. Any image-based similarity metric can be leveraged, whereas SSIM and PSNR [72], [73], Normalized Cross Correlation [56], Dice Score and Hausdorff Distance on available segmentations [61], [71] have been observed in the reviewed papers. A further image-based motion evaluation strategy compares the dynamic position of relevant organ boundaries, e.g. the hepatic dome, in the predicted motion-aware reconstruction with a motion-resolved reference [69].

C. Downstream Tasks

In some cases the MoCo framework does not solely aim to provide a high-quality reconstruction, but enables further downstream tasks. In this case the downstream findings can be evaluated independently, e.g. by calculating the Dice overlap on organ segmentations [28], [40], [66] or computing SSIM and relative error metrics on T2* maps [30]. To evaluate the added statistical power due to MoCo in longitudinal analyses, manual quality control of structural elements like cortical surface reconstructions and cortical thickness correlation analyses can be employed [29]. Especially if no reference is available, the sharpness of small anatomical features, such as coronary vessels can be analyzed [56], [74]. In cardiac imaging, cardiac function analysis [39], [41], [68] or myocardial strain measurements [43] can be evaluated. Further, as an important end goal for MR MoCo, a comparison of clinical findings in a motion corrupted and corrected scan can be conducted [26].

VII. DISCUSSION

In the previous sections, we gave an overview of available data and motion simulation for MR MoCo reconstruction (Sec. III), we outlined the state-of-the-art model architectures for learning-based MR MoCo (Sec. IV) and their common training (Sec. V) and evaluation strategies (Sec.). In the following, we critically discuss the reviewed methods presented in Secs. III-VI. We highlight common strategies and differences, pointing out their advantages, limitations and needs for improvement.

A. Data Availability and Motion Simulation

The presented *motion simulation* strategies (Sec. III) are predominantly used for training and evaluating motion correction approaches. Only a few motion compensation approaches include non-rigid motion simulation procedures and if so, only for evaluation [46], [56], [74], [77]. Non-rigid motion simulation, though, is limited both in image space [46], [56], [60], [65], [66], [74], [77] and even more, the simplified version in k-space [22], [38], [41], [43]. Simulating translation for breathing motion may broadly cover the direction of the motion but does not represent the deformable nature of real patient motion. Thus, when 3D data are available, simulation using motion fields is the more realistic and preferable approach.

Even though the motion simulation procedure based on (1) appears to be well-defined, several variations have been implemented for both rigid-body and non-rigid motion simulation. Some methods only simulate in-plane motion [21], [22], [23], [27], [38], [42], [48], [49], [52], [53], [55], [62], [64], whereas others consider full through-plane motion patterns, which better resemble real patient motion [24], [25], [29], [30], [32], [33], [34], [36], [40], [43], [45], [50], [51], [54].

Moreover, in the case of rigid-body motion, some methods limit their simulations to a small number of motion-states [23], [26], [27], [29], [30], [32], [33], [34], [35], [42], [44], [45], [51], [52], [53], [62], [64], whereas others base their simulations on time-resolved motion curves, which are either generated randomly or based on clinical measurements like functional MRI time series [21], [24], [25], [36], [43], [48], [49], [50], [54], [55], [60]. In the case of non-rigid motion, image-based simulation can only simulate discrete motion states, since multiple reconstructions are needed for merging into a corrupted image. The acquisition time for a single reconstruction limits the total number of motion states that can be simulated even for fast sequences, like cardiac cine imaging [46] and even more, for time-consuming 3D acquisitions in the abdomen.

As described in section II, second-order motion-effects influence MR signal properties in addition to the effects of positional changes. A few recent approaches consider such second-order motion effects for more realistic motion simulations in specific MR sequences, like phase shifts of stimulated echoes due to respiration [43] or motion-induced magnetic field inhomogeneity changes in T2*-weighted MRI [54].

As a simplified approach to simulate motion, a few authors explicitly exclude central k-space lines [22], [26], [29], [38], [42], [55], which severely limits the generalisability of their methods to real motion patterns.

In contrast to the above discussed simulation procedures for MoCo, there is in general no GT motion-free image for motion compensation methods, since breathing and especially heartbeat cannot be avoided. Breath hold and gated acquisitions have the potential to approximate GT motion-free images. Hence, motion-corrupted images can be simulated using deformation fields [56], [74], which can be derived from classical motion-resolved reconstructions [90], other imaging modalities or physical models, as e.g. the XCAT phantom [91]. However, such simulations might require expensive acquisitions of additional data, might not offer sufficient temporal

resolution for all applications and the XCAT phantom, specifically, simulates images based on CT images, making raw k-space data unavailable. Furthermore, given that the majority of presented motion compensation methods aims for acceleration, most methods focus on *simulating undersampling artefacts* and compare reconstructions to fully sampled acquisitions to show the acceleration potential of their approach. These simulations vary with regard to the underlying retrospective sampling trajectory (e.g. cartesian, radial or spiral). For cartesian sampling, the center of k-space is usually explicitly sampled more frequently than the periphery, which is also common for classical acceleration methods.

Regardless of whether motion or undersampling artefacts are simulated, many authors who train their models on simulated data at least show qualitative evaluations on a few unpaired real motion or prospectively undersampled datasets, respectively. However, some approaches only evaluate on simulated data [23], [28], [35], [38], [39], [44], [46], [48], [51], [51], [52], [53], [54], [55], [58], [59], [60], [61], [64], [67], [69], [70], [71], [72], [73], [74], [79], [81], which questions their generalisability to real motion or prospectively undersampled data.

In addition to these limitations of implemented simulation procedures, the majority of image-based approaches [21], [22], [23], [25], [26], [27], [28], [29], [31], [32], [32], [35], [36], [37], [44], [46] and some k-space-based approaches [51], [55], [61], [64], [65], [66], [76] rely on magnitude-only data, which is questionable especially for the latter category. Shimron et al. [92] describe the generation of k-space data from real-valued, coil-combined magnitude data as a data crime, for instance, if zero-padding was performed during the initial reconstruction process. Furthermore, phase information is relevant for several applications, e.g. quantitative susceptibility mapping or fat-water separation, and is part of postprocessing pipelines, e.g. for background field correction. Magnitude-only approaches cannot be applied to these applications in a straightforward manner.

Public raw multi-coil datasets with paired motion experiments could enable a more realistic method development and evaluation for researchers who do not have the possibilities to acquire such raw k-space data in a paired setting. For brain imaging, currently only magnitude data with and without intentional subject motion [3], [93] and motion-free k-space datasets [94] are available. For motion compensation in cardiac imaging, breath-hold cardiac gated radial dataset [95] as well as an undersampled, free-breathing k-space dataset [96] are available.

B. Architectures

A wide variety of architectures has been proposed to target MR MoCo (Sec. IV). While aiming for different applications, we outline common trends regarding the (a) data domain, (b) targeted motion types and (c) motion modeling. Furthermore, we discuss (d) model interpretability, (e) patient-specific models and (f) the interchangeability of modules.

1) *Image-Based vs. k-Space-Based Motion Correction*: The review of proposed architectures for motion-corrected MR reconstruction shows that both image and k-space methods are commonly applied. Image-based methods profit from broadly

available data, since they can be applied on existing MR image databases. Additionally, data are frequently limited to magnitude values, which reduces the complexity of the architecture. Nevertheless, image-based methods are more likely to produce hallucinations, as lack of raw k-space data restricts the ability to perform data consistency checks. Also, these approaches lack the flexibility to adapt the reconstruction strategy based on motion parameters [52].

In contrast, k-space-based methods benefit from a more comprehensive data representation that includes additional information such as phase and coil sensitivities, which can improve final image quality [97]. Besides a low availability of raw k-space data in practice, a potential disadvantage is that the reconstruction parameters and hardware may have a larger influence on the final image. Thus, it may be more difficult to compare results across different systems [97].

2) *Different Motion Types*: As pointed out in Sec. III, the types of motion observed in brain and cardiac/abdominal imaging are distinct. Motion artefacts in brain images mostly originate from rigid-body motion of different severity at random time points, generally resulting in blurring. Quasi-periodic motion, which is typical in cardiac and abdominal imaging, can additionally result in ghosting artefacts [1]. Because of these distinctive visual characteristics, architectures are frequently trained and tested on specific body regions.

For *image-based methods*, the presented approaches predominantly target motion correction in the brain, potentially due to the inherent capability of CNNs and encoder-decoder structures to sharpen edges, i.e. denoise artefacts apparent as blurring. To counteract periodic signal modulations leading to ghosting artefacts, abdominal and cardiac image-based methods reconstruct images from specific time intervals, i.e. motion states. Still, residual blurring persists due to continuous motion within this time interval, and missing data lead to undersampling artefacts. Therefore, image-based MoCo strategies applied to quasi-periodic moving organs mainly rely on information fusion from multiple dynamics. Only few methods aim to learn a latent motion-free representation of the heart or abdomen from single reconstructions.

Also, for *k-space-based methods*, distinct architecture approaches exist for different motion types. The modeling of rigid-body motion with only few parameters facilitates joint optimization of motion estimation and reconstruction with a reasonable computational overhead, making it a technique typically limited to the head region (IV-B.1.a). With the random timing of motion in the head, some motion correction strategies focus on identifying the time of occurrence, i.e. motion detection. In contrast, methods targeting periodic motion mostly rely on fusion of data from different motion states, i.e. motion compensation. Since data consistency is crucial to ensure physical plausibility, considerably more k-space domain approaches have been presented than pure image-based methods for motion compensation. Whereas quasi-periodic motion compensation is the aim of most methods for the cardiac and abdominal anatomy, few methods target explicit motion correction of irregular motion sources, e.g. due to mistrigging artefacts [65], [66].

3) *Motion Modelling*: Many methods combine MoCo and reconstruction in one process (Sec. IV-B). The majority of

these hybrid approaches furthermore includes an explicit motion model, regardless of the optimization process or type of motion. However, such an explicit motion model only offers an approximation of the actual motion, which is considered in just one published work [60]. As a result, the accuracy of the reconstruction is constrained [98]. Nevertheless, compared to implicit motion modeling methods, explicit motion models allow for additional quality control (refer to Sec. VI).

Both explicit and implicit motion compensation techniques frequently rely on data that have been temporally separated into several motion states throughout the cardiac or respiratory cycle. Explicit methods estimate motion between the binned reconstructions, whereas implicit methods do not directly model motion but exploit temporal redundancies in the dynamic data. On the one hand, this requires a reliable navigator signal representing the actual motion of the organ of interest. On the other hand, binning of data from multiple cycles is susceptible to inter-cycle variability [4]. Whereas preliminary work on motion estimation uncertainty exists [99], almost none of the presented MoCo architectures consider uncertainty in their motion modelling within the reconstruction pipeline.

As an additional drawback of methods modelling motion based on binned motion states, the residual motion within these states blurs the reconstruction. Increasing the temporal resolution by binning fewer data points to one motion state would lead to increased undersampling, and, therefore, affect the potential to generate reliable motion estimates. Although motion occurs continuously, many models are restricted to discrete representations. An initial attempt to avoid motion states is proposed in [79], where a continuous representation of the motion dimension is learned.

Lastly, current motion models integrated into learning-based MoCo architectures mainly focus on primary motion effects in k-space. Especially when handling raw data, further physics-based motion-induced secondary effects should be considered. Consequently, the motion-aware reconstruction could be extended beyond the physical motion modelling, e.g. by correcting spin history effects or B_0 - and B_1 -distributions [100].

4) Interpretability: When using any learning-based motion-correcting reconstruction framework, it is important to understand the model’s behaviour. Employing models without such knowledge may lead to undesired effects like hallucinations, directly influencing the critical process of medical diagnosis. To avoid adverse effects of “black-box” models, interpretability should already be considered in the architecture design. Including physical knowledge, e.g. by explicit motion modelling, can aid in generating interpretable reconstruction results as well as influences of intermediate steps. For implicit models, in contrast, it is important to understand the bottle-necks. Disentangling a learned low-dimensional representation, i.e. the latent space, [40] is a first step towards such informed modeling.

5) Patient-Specific Models: Recently, patient-specific generative models have been developed as novel direction for motion-compensated MR reconstruction (see Sec. IV-B.2). Since motion patterns can strongly vary between patients, such individually learned representations may perform better than generalized approaches. Nevertheless, the need to retrain

the model comes with prolonged reconstruction times and increased computational resources. Transferable concepts have not yet been proposed for patient-specific generative models.

6) Interchangeable Modules: The presented architectures for motion-corrected MR reconstruction aim at different motion patterns, anatomical regions and sequences. While a general solution is unlikely, some components can be seen as interchangeable modules to enable further development and improvement of methods. For example, varying methods for undersampled reconstruction [101] may be integrated into approaches that aim at removing motion-corrupted lines. Another exchangeable module can be motion estimation, which, in theory, could be conducted with any other learning-based registration method [102], but needs to consider strong undersampling artefacts in the input images.

While the reconstruction concept dominantly used in fetal motion correction is fundamentally different (Sec. II), individual motion estimation concepts may be transferable as well. SVR-based MoCo methods frequently model rigid intra-slice motion estimation as well, e.g. based on gated recurrent units [103] or as learnable parameter within the SVR reconstruction problem [104]. Non-rigid estimation techniques or architecture backbones could be transferred from and to other applications.

C. Training Objectives

A variety of training strategies and objective functions have been adopted for optimizing MoCo models, as described in Sec. V. Next to back-propagating errors from the motion-corrected image and from intermediate steps, like e.g. motion estimation, MoCo models can also be trained in an end-to-end fashion with a downstream task. For instance, Xu et al. [30] combine MoCo with T_2^* parameter quantification and Oksuz et al. [66] with cardiac segmentation. For such a joint optimization, the MoCo task and the downstream task of interest might benefit from each other, improving the overall performance. However, the resulting motion corrected images might not be suitable for different downstream tasks.

Additionally, due to the limited availability of GT data (see Sec. III for details), more and more self-supervised and unsupervised methods have been proposed for data-efficient training. Adversarial training is employed to cope with unpaired image data. If k-space data are available, application of data consistency allows for self-supervised training. Proposed subject-specific generative models [77], [78], [79], [80] are optimized by comparing the reconstruction result with measured data, and therefore, are inherently self-supervised. The Noise2Noise concept, originally proposed for image restoration [105], is adapted in a self-supervised generative [63] as well as unsupervised inter-subject [76] motion-corrected reconstruction strategy. This highlights the potential to transfer further computer vision training strategies to cope with limited data.

In general, many approaches combine various objective functions in order to guide the optimization, such as the combination with downstream losses or the combination of motion estimation or adversarial losses with image-based losses of the final reconstruction. While this can enforce specific properties in the result, like e.g. imposing more realistic

motion patterns by regularizing motion fields [56], [58], [74], the training process might become more complex, since the weighting of different losses is not straightforward, but rather another hyper-parameter to be tuned. Furthermore, computational effort might increase, e.g. when combining losses in image- and k-space domain for non-Cartesian sampling patterns requiring a costly domain transformation [80].

D. Evaluation Metrics

As outlined in Sec. VI, the most commonly used evaluation metrics are image quality metrics, which evaluate the main goal of MoCo: a high quality image. Among these, especially the full-reference metrics SSIM and PSNR stand out, which also seem to correlate well with radiological assessment [106]. A downside of these full-reference methods is that they rely on the availability of paired GT data (compare Sec. III). Reference-free methods, on the other hand, do not require a GT image but are not yet widely used, since they are less consistent. Another important consideration for full-reference metrics is that the evaluated and GT image might not be perfectly aligned [107]. In order to not overestimate motion-induced errors, some authors include a co-registration step before calculating full-reference metrics. However, since registration might also introduce interpolation errors, further research is needed on this topic.

In general, there is no standardized way of evaluating image quality in practice, which is not only problematic for learning-based MoCo, but extends to the entire fields of MoCo and image reconstruction. A variety of metrics is used by different authors and for some metrics, e.g. SSIM, hyperparameters can be set manually. This heterogeneity limits the comparability of different methods, even when ignoring the fact that different methods are evaluated on different datasets. Furthermore, the lack of standardized recommendations for evaluation also leaves room for “metric picking”, which might lead to overestimated performances and misguide future research. However, when aiming to develop general recommendations, investigations on the relevance of different image quality metrics on diverse datasets are urgently needed. Since no single image quality metric can be expected to be sensitive to all possible image artefacts, such recommendations may comprise a broad, generally accepted set of metrics.

Similarly, for subjective image quality scoring, the variability of strategies regarding instructions, scales and evaluators limits the comparability of different methods. A common recommendation is to reduce inter-observer variability by averaging the scores of multiple observers. However, quality assessment is a time consuming process and experts such as radiologists already have a high workload in many hospitals, which limits the practicality of qualitative image evaluation. A possible solution might be to utilize deep learning models that can be trained to perform reference-free image quality assessment [108], [109], [110]. However, further research is needed on questions like the reliability and the generalisability of trained models to distribution shifts.

Next to the relevance of consistent image quality evaluation, we would also like to emphasize the importance of “in-between quality assurance” by evaluating motion detection and estimation as intermediate results for methods that explicitly

model motion. If the extracted motion information is incorrect, these errors might propagate into the final reconstruction. Again, standardized evaluation criteria would allow for better comparisons of different methods.

Furthermore, we would like to highlight that the additional analysis of downstream tasks is application specific, which limits the potential of general recommendations. Such additional evaluations, though, might be highly relevant for the translation of developed methods into clinical practice.

VIII. CONCLUSION AND OUTLOOK

In this review, we have provided a comprehensive overview of existing learning-based methods for MR MoCo, identifying synergies and differences in underlying data usage, architectures, training and evaluation strategies. In the following, we point out key findings and highlight aspects that require further investigation.

For learning-based MoCo in MR both real and simulated data can be used for training and evaluation. Motion simulation provides the benefit of an existing ground truth and can be an effective means for initial development, however pitfalls such as sole in-plane simulation, discrete motion state modelling, exclusion of central k-space lines and erroneous processing of magnitude images need to be avoided. For non-rigid motion types, motion simulation should focus on deformable motion. Nevertheless, particularly with respect to secondary motion effects and to enable a reliable transfer to clinical applications, real data needs to be employed, at least for the evaluation process. Since such data is difficult to obtain and can strongly vary from site to site, a common database with real motion artefacts is crucial for the community. Inclusion of raw k-space data would further advance the development of data-consistent methods and ensure method comparability independent of individual hardware settings.

Next to openly accessible data, we would like to emphasize the need for systematic evaluation guidelines. Up to this point, methods have been evaluated with various metrics with various definitions. Initial work on the relevance and performance of metrics needs to be extended. A standardization of both, full-reference and reference-free evaluation metrics, should be aimed for.

Architecture development should strongly focus on DC based methods, which avoid hallucinations and, thus, might be easier to translate into clinical practice. Whereas targeted motion patterns will continue to affect the architecture design, the underlying motion modelling requires careful consideration in all cases, e.g. regarding simplified assumptions of discrete motion states or uncertainties of motion estimates. Training the MoCo model in an end-to-end fashion with a downstream task to back-propagate task-specific errors seems to be promising, but is limited to the specific application. Further, self-supervised training strategies can encounter expensive acquisitions of GT data. Self-supervised patient-specific models open up a new direction, but require expensive retraining for each individual reconstruction.

In general, most state-of-the-art architectures are developed for 2D data, requiring less expensive computation. However, many 2D approaches cannot correct through-plane motion,

which limits their performance for real motion-corrupted data. Also, due to the slice-selective excitation, secondary motion effects like spin history effects impact 2D data more strongly than 3D data and cannot be simulated in a straightforward manner. Thus, future research should focus, whenever possible, on 3D acquisitions or otherwise, consider 3D motion information. Exploration of transferable models may enable reduced computational burden.

In view of the rapid development of deep learning, we expect further advances in learning-based MoCo in MRI in the near future. An initial transfer of methods developed in the machine learning community has been presented in this review, but there are many more to be exploited: methodological developments with novel architectures, e.g. diffusion models [111], neural implicit representations [112] and transformers [113], can be advanced. Motion modelling may benefit from parallel developments of probabilistic models that include uncertainty estimates. Data-efficient strategies could reduce the need for large training data sets or long training of patient-specific models.

Modern learning-based MoCo methods should consider the multi-dimensional nature of MRI. Clinical protocols often include multiple contrasts and dynamics, providing additional information at hand. Co-development of acquisition and motion monitoring techniques should be promoted. The presented and any future learning-based MoCo models may have the potential to be integrated into clinical MRI protocols, allowing for motion detection and estimation at fast inference times.

With this review we aim to bridge the gap between machine learning and MRI. We see potential for further development of clinically relevant MoCo methods. Not only would this development aid in improving current clinical protocols, but open doors to areas where motion has been a major restricting factor, e.g. due to irregular and deformable patterns. Next to MR reconstruction, advances could be transferred to multi-modal imaging techniques and foster MRI as a non-invasive motion monitoring technique for applications such as PET-MR and MR-guided radiotherapy.

REFERENCES

- [1] F. Godenschweger et al., "Motion correction in MRI of the brain," *Phys. Med. Biol.*, vol. 61, no. 5, pp. 32–56, 2016.
- [2] M. Zaitsev, J. Maclaren, and M. Herbst, "Motion artifacts in MRI: A complex problem with many partial solutions," *J. Magn. Reson. Imag.*, vol. 42, no. 4, pp. 887–901, Oct. 2015.
- [3] M. Ganz and H. Eichhorn, "Datasets with and without deliberate head movements for evaluating the performance of markerless prospective motion correction and selective reacquisition in a general clinical protocol for brain MRI," *OpenNeuro*, 2022, doi: [10.18112/open-neuro.ds004332.v1.0.0](https://doi.org/10.18112/open-neuro.ds004332.v1.0.0).
- [4] J. R. McClelland, D. J. Hawkes, T. Schaeffter, and A. P. King, "Respiratory motion models: A review," *Med. Image Anal.*, vol. 17, no. 1, pp. 19–42, Jan. 2013.
- [5] G. B. Chavhan, L. F. Roca, and A. C. Coblentz, "Liver magnetic resonance imaging: How we do it," *Pediatric Radiol.*, vol. 52, no. 2, pp. 167–176, Feb. 2022.
- [6] J. C. Uffman, D. Tumin, V. Raman, A. Thung, B. Adler, and J. D. Tobias, "MRI utilization and the associated use of sedation and anesthesia in a pediatric ACO," *J. Amer. College Radiol.*, vol. 14, no. 7, pp. 924–930, Jul. 2017.
- [7] J. G. Pipe, "Motion correction with PROPELLER MRI: Application to head motion and free-breathing cardiac imaging," *Magn. Reson. Med.*, vol. 42, no. 5, pp. 963–969, Nov. 1999.
- [8] S. Winkelmann, T. Schaeffter, T. Koehler, H. Eggers, and O. Doessel, "An optimal radial profile order based on the golden ratio for time-resolved MRI," *IEEE Trans. Med. Imag.*, vol. 26, no. 1, pp. 68–76, Jan. 2007.
- [9] D. C. Peters et al., "Undersampled projection reconstruction applied to MR angiography," *Magn. Reson. Med.*, vol. 43, no. 1, pp. 91–101, Jan. 2000.
- [10] K. P. Pruessmann, M. Weiger, M. B. Scheidegger, and P. Boesiger, "SENSE: Sensitivity encoding for fast MRI," *Magn. Reson. Med.*, vol. 42, no. 5, pp. 952–962, Nov. 1999.
- [11] M. Lustig, D. Donoho, and J. M. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging," *Magn. Reson. Med.*, vol. 58, no. 6, pp. 1182–1195, 2007.
- [12] 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, Jul. 2006.
- [13] A. J. W. van der Kouwe, T. Benner, and A. M. Dale, "Real-time rigid body motion correction and shimming using cloverleaf navigators," *Magn. Reson. Med.*, vol. 56, no. 5, pp. 1019–1032, Nov. 2006.
- [14] M. A. Bush et al., "Prospective correction of patient-specific respiratory motion in myocardial T₁ and T₂ mapping," *Magn. Reson. Med.*, vol. 85, no. 2, pp. 855–867, 2021.
- [15] D. Atkinson, D. L. G. Hill, P. N. R. 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. Imag.*, vol. 16, no. 6, pp. 903–910, Dec. 1997.
- [16] L. Cordero-Grande, R. P. A. G. Teixeira, E. J. Hughes, J. Hutter, A. N. Price, and J. V. Hajnal, "Sensitivity encoding for aligned multishot magnetic resonance reconstruction," *IEEE Trans. Comput. Imag.*, vol. 2, no. 3, pp. 266–280, Sep. 2016.
- [17] A. U. Uus, A. E. Collado, T. A. Roberts, J. V. Hajnal, M. A. Rutherford, and M. Deprez, "Retrospective motion correction in foetal MRI for clinical applications: Existing methods, applications and integration into clinical practice," *Brit. J. Radiol.*, vol. 96, Aug. 2022, Art. no. 20220071.
- [18] D. Atkinson, "Retrospective correction of motion in MR images," in *Advances in Magnetic Resonance Technology and Applications: Motion Correction in MR*, vol. 6, A. J. van der Kouwe and J. B. Andre, Eds. Cambridge, MA, USA: Academic Press, 2022, pp. 259–267.
- [19] A. Loktyushin, H. Nickisch, R. Pohmann, and B. Schölkopf, "Blind retrospective motion correction of MR images," *Magn. Reson. Med.*, vol. 70, no. 6, pp. 1608–1618, Dec. 2013.
- [20] B. Levac, S. Kumar, S. Kardonik, and J. I. Tamir, "FSE compensated motion correction for MRI using data driven methods," in *Medical Image Computing and Computer Assisted Intervention—MICCAI*, vol. 13436. Berlin, Germany: Springer, 2022, pp. 707–716.
- [21] K. Sommer, A. Saalbach, T. Brosch, C. Hall, N. M. Cross, and J. B. Andre, "Correction of motion artifacts using a multiscale fully convolutional neural network," *Amer. J. Neuroradiol.*, vol. 41, no. 3, pp. 416–423, Mar. 2020.
- [22] D. Tamada, M.-L. Kromrey, S. Ichikawa, H. Onishi, and U. Motosugi, "Motion artifact reduction using a convolutional neural network for dynamic contrast enhanced MR imaging of the liver," *Magn. Reson. Med. Sci.*, vol. 19, no. 1, pp. 64–76, 2020.
- [23] S. Chatterjee, A. Sciarra, M. Dünwald, S. Oeltze-Jafra, A. Nürnberger, and O. Speck, "Retrospective motion correction of MR images using prior-assisted deep learning," in *Proc. 34th NIPS*, 2020, pp. 1–5.
- [24] C. M. Pirkel et al., "Learning residual motion correction for fast and robust 3D multiparametric MRI," *Med. Image Anal.*, vol. 77, Apr. 2022, Art. no. 102387.
- [25] K. Pawar, Z. Chen, J. Seah, M. Law, T. Close, and G. Egan, "Clinical utility of deep learning motion correction for T₁ weighted MPRAGE MR images," *Eur. J. Radiol.*, vol. 133, Dec. 2020, Art. no. 109384.
- [26] J. Liu, M. Kocak, M. Supanich, and J. Deng, "Motion artifacts reduction in brain MRI by means of a deep residual network with densely connected multi-resolution blocks (DRN-DCMB)," *Magn. Reson. Imag.*, vol. 71, pp. 69–79, Sep. 2020.
- [27] J. Lee, B. Kim, and H. Park, "MC²-Net: Motion correction network for multi-contrast brain MRI," *Magn. Reson. Med.*, vol. 86, no. 2, pp. 1077–1092, Aug. 2021.
- [28] I. Oksuz, "Brain MRI artefact detection and correction using convolutional neural networks," *Comput. Methods Programs Biomed.*, vol. 199, Feb. 2021, Art. no. 105909.

- [29] B. A. Duffy et al., "Retrospective motion artifact correction of structural MRI images using deep learning improves the quality of cortical surface reconstructions," *NeuroImage*, vol. 230, Apr. 2021, Art. no. 117756.
- [30] X. Xu et al., "Learning-based motion artifact removal networks for quantitative R_2^* mapping," *Magn. Reson. Med.*, vol. 88, no. 1, pp. 106–119, Jul. 2022.
- [31] R. Jafari et al., "GRASPNET: Fast Spatiotemporal deep learning reconstruction of golden-angle radial data for free-breathing dynamic contrast-enhanced MRI," *NMR Biomed.*, vol. 36, no. 3, p. e4861, 2022.
- [32] M. A. Al-masni et al., "Stacked U-Nets with self-assisted priors towards robust correction of rigid motion artifact in brain MRI," *NeuroImage*, vol. 259, Oct. 2022, Art. no. 119411.
- [33] M. A. Al-masni, S. Lee, A. K. Al-Shamiri, S.-M. Gho, Y. H. Choi, and D.-H. Kim, "A knowledge interaction learning for multi-echo MRI motion artifact correction towards better enhancement of SWI," *Comput. Biol. Med.*, vol. 153, Feb. 2023, Art. no. 106553.
- [34] P. M. Johnson and M. Drangova, "Conditional generative adversarial network for 3D rigid-body motion correction in MRI," *Magn. Reson. Med.*, vol. 82, no. 3, pp. 901–910, Sep. 2019.
- [35] M. Usman, S. Latif, M. Asim, B.-D. Lee, and J. Qadir, "Retrospective motion correction in multishot MRI using generative adversarial network," *Sci. Rep.*, vol. 10, no. 1, p. 4786, Mar. 2020.
- [36] Q. Bao et al., "Retrospective motion correction for preclinical/clinical magnetic resonance imaging based on a conditional generative adversarial network with entropy loss," *NMR Biomed.*, vol. 35, no. 12, p. e4809, Dec. 2022.
- [37] T. Küstner, K. Armanious, J. Yang, B. Yang, F. Schick, and S. Gatidis, "Retrospective correction of motion-affected MR images using deep learning frameworks," *Magn. Reson. Med.*, vol. 82, no. 4, pp. 1527–1540, Oct. 2019.
- [38] Y. Wu, X. Wang, and A. K. Katsaggelos, "Motion artifact reduction in abdominal MRIs using generative adversarial networks with perceptual similarity loss," *Proc. SPIE*, vol. 12088, pp. 142–150, Dec. 2021.
- [39] V. Ghodrati et al., "Temporally aware volumetric generative adversarial network-based MR image reconstruction with simultaneous respiratory motion compensation: Initial feasibility in 3D dynamic cine cardiac MRI," *Magn. Reson. Med.*, vol. 86, no. 5, pp. 2666–2683, Nov. 2021.
- [40] S. Liu, K.-H. Thung, L. Qu, W. Lin, D. Shen, and P.-T. Yap, "Learning MRI artefact removal with unpaired data," *Nature Mach. Intell.*, vol. 3, no. 1, pp. 60–67, Jan. 2021.
- [41] V. Ghodrati et al., "Retrospective respiratory motion correction in cardiac cine MRI reconstruction using adversarial autoencoder and unsupervised learning," *NMR Biomed.*, vol. 34, no. 2, p. e4433, Feb. 2021.
- [42] G. Oh, J. E. Lee, and J. C. Ye, "Unpaired MR motion artifact deep learning using outlier-rejecting bootstrap aggregation," *IEEE Trans. Med. Imag.*, vol. 40, no. 11, pp. 3125–3139, Nov. 2021.
- [43] M. Abdi, K. C. Bilchick, and F. H. Epstein, "Compensation for respiratory motion-induced signal loss and phase corruption in free-breathing self-navigated cine DENSE using deep learning," *Magn. Reson. Med.*, vol. 89, no. 5, pp. 1975–1989, May 2023.
- [44] L. A. van der Goten and K. Smith, "Wide range MRI artifact removal with transformers," in *Proc. 33rd Brit. Mach. Vis. Conf.*, 2022, pp. 1–13.
- [45] K. Armanious, A. Tanwar, S. Abdulatif, T. Küstner, S. Gatidis, and B. Yang, "Unsupervised adversarial correction of rigid MR motion artifacts," in *Proc. IEEE 17th ISBI*, Apr. 2020, pp. 1494–1498.
- [46] Q. Lyu et al., "Cine cardiac MRI motion artifact reduction using a recurrent neural network," *IEEE Trans. Med. Imag.*, vol. 40, no. 8, pp. 2170–2181, Aug. 2021.
- [47] J. Lv, M. Yang, J. Zhang, and X. Wang, "Respiratory motion correction for free-breathing 3D abdominal MRI using CNN-based image registration: A feasibility study," *Brit. J. Radiol.*, vol. 91, Jan. 2018, Art. no. 20170788.
- [48] E. Kuzmina, A. Razumov, O. Y. Rogov, E. Adalsteinsson, J. White, and D. V. Dylow, "Autofocusing+: Noise-resilient motion correction in magnetic resonance imaging," in *Medical Image Computing and Computer Assisted Intervention—MICCAI*, vol. 13436. Berlin, Germany: Springer, 2022, pp. 365–375.
- [49] J. Hossbach et al., "Deep learning-based motion quantification from k -space for fast model-based magnetic resonance imaging motion correction," *Med. Phys.*, vol. 50, no. 4, pp. 2148–2161, 2022.
- [50] M. W. Haskell et al., "Network accelerated motion estimation and reduction (NAMER): Convolutional neural network guided retrospective motion correction using a separable motion model," *Magn. Reson. Med.*, vol. 82, no. 4, pp. 1452–1461, Oct. 2019.
- [51] C. Wang, Y. Liang, Y. Wu, S. Zhao, and Y. P. Du, "Correction of out-of-FOV motion artifacts using convolutional neural network," *Magn. Reson. Imag.*, vol. 71, pp. 93–102, Sep. 2020.
- [52] N. M. Singh et al., "Data consistent deep rigid MRI motion correction," 2023, *arXiv:2301.10365*.
- [53] B. Levac, S. Kumar, A. Jalal, and J. I. Tamir, "Accelerated motion correction with deep generative diffusion models," 2022, *arXiv:2211.00199*.
- [54] H. Eichhorn et al., "Physics-aware motion simulation for T2*-weighted brain MRI," 2023, *arXiv:2303.10987*.
- [55] L. Cui et al., "Motion artifact reduction for magnetic resonance imaging with deep learning and k -space analysis," *PLoS ONE*, vol. 18, no. 1, Jan. 2023, Art. no. e0278668.
- [56] H. Qi et al., "Non-rigid respiratory motion estimation of whole-heart coronary MR images using unsupervised deep learning," *IEEE Trans. Med. Imag.*, vol. 40, no. 1, pp. 444–454, Jan. 2021.
- [57] C. Munoz, H. Qi, G. Cruz, T. Küstner, R. M. Botnar, and C. Prieto, "Self-supervised learning-based diffeomorphic non-rigid motion estimation for fast motion-compensated coronary MR angiography," *Magn. Reson. Imag.*, vol. 85, pp. 10–18, Jan. 2022.
- [58] K. Hammernik, J. Pan, D. Rueckert, and T. Küstner, "Motion-guided physics-based learning for cardiac MRI reconstruction," in *Proc. 55th Asilomar Conf. Signals, Syst., Comput.*, Oct. 2021, pp. 900–907.
- [59] J. Pan, D. Rueckert, T. Küstner, and K. Hammernik, "Learning-based and unrolled motion-compensated reconstruction for cardiac MR CINE imaging," in *Medical Image Computing and Computer Assisted Intervention—MICCAI*, vol. 13436. Berlin, Germany: Springer, 2022, pp. 686–696.
- [60] M. S. Feinler and B. N. Hahn, "Retrospective motion correction in gradient echo MRI by explicit motion estimation using deep CNNs," 2023, *arXiv:2303.17239*.
- [61] H.-C. Shao et al., "Real-time MRI motion estimation through an unsupervised k -space-driven deformable registration network (KS-RegNet)," *Phys. Med. Biol.*, vol. 67, no. 13, Jul. 2022, Art. no. 135012.
- [62] M. Rotman, R. Brada, I. Beniaminy, S. Ahn, C. J. Hardy, and L. Wolf, "Correcting motion artifacts in MRI scans using a deep neural network with automatic motion timing detection," *Proc. SPIE*, vol. 11595, pp. 296–305, Feb. 2021.
- [63] Z. Miller and K. M. Johnson, "Motion compensated self supervised deep learning for highly accelerated 3D ultrashort echo time pulmonary MRI," *Magn. Reson. Med.*, vol. 89, no. 6, pp. 2361–2375, Jun. 2023.
- [64] N. M. Singh, J. E. Iglesias, E. Adalsteinsson, A. V. Dalca, and P. Golland, "Joint frequency and image space learning for MRI reconstruction and analysis," *Mach. Learn. Biomed. Imag.*, vol. 1, pp. 1–28, Jun. 2022.
- [65] L. Oksuz et al., "Detection and correction of cardiac MR motion artefacts during reconstruction from K -space," 2019, *arXiv:1906.05695*.
- [66] I. Oksuz et al., "Deep learning-based detection and correction of cardiac MR motion artefacts during reconstruction for high-quality segmentation," *IEEE Trans. Med. Imag.*, vol. 39, no. 12, pp. 4001–4010, Dec. 2020.
- [67] J. Schlemper, J. Caballero, J. V. Hajnal, A. N. Price, and D. Rueckert, "A deep cascade of convolutional neural networks for dynamic MR image reconstruction," *IEEE Trans. Med. Imag.*, vol. 37, no. 2, pp. 491–503, Feb. 2018.
- [68] T. Küstner et al., "CINENet: Deep learning-based 3D cardiac CINE MRI reconstruction with multi-coil complex-valued 4D spatio-temporal convolutions," *Sci. Rep.*, vol. 10, no. 1, p. 13710, Aug. 2020.
- [69] M. Terpstra, M. Maspero, J. Verhoeff, and C. van den Berg, "Accelerated respiratory-resolved 4D-MRI with separable spatio-temporal neural networks," 2022, *arXiv:2211.05678*.
- [70] C. Qin, J. Schlemper, J. Caballero, A. N. Price, J. V. Hajnal, and D. Rueckert, "Convolutional recurrent neural networks for dynamic MR image reconstruction," *IEEE Trans. Med. Imag.*, vol. 38, no. 1, pp. 280–290, Jan. 2019.
- [71] Q. Huang et al., "Dynamic MRI reconstruction with end-to-end motion-guided network," *Med. Image Anal.*, vol. 68, Feb. 2021, Art. no. 101901.
- [72] G. Seegoolam, J. Schlemper, C. Qin, A. Price, J. Hajnal, and D. Rueckert, "Exploiting motion for deep learning reconstruction of extremely-undersampled dynamic MRI," in *Medical Image Computing and Computer Assisted Intervention—MICCAI*. Cham, Switzerland: Springer, 2019, pp. 704–712.
- [73] J. Yang, T. Küstner, P. Hu, P. Lio, and H. Qi, "End-to-end deep learning of non-rigid groupwise registration and reconstruction of dynamic MRI," *Frontiers Cardiovascular Med.*, vol. 9, Apr. 2022, Art. no. 880186.

- [74] H. Qi et al., “End-to-end deep learning nonrigid motion-corrected reconstruction for highly accelerated free-breathing coronary MRA,” *Magn. Reson. Med.*, vol. 86, no. 4, pp. 1983–1996, Oct. 2021.
- [75] T. Küstner et al., “Self-supervised motion-corrected image reconstruction network for 4D magnetic resonance imaging of the body trunk,” *APSIPA Trans. Signal Inf. Process.*, vol. 11, no. 1, p. e12, 2022.
- [76] W. Gan, Y. Sun, C. Eldeniz, J. Liu, H. An, and U. S. Kamilov, “Deformation-compensated learning for image reconstruction without ground truth,” *IEEE Trans. Med. Imag.*, vol. 41, no. 9, pp. 2371–2384, Sep. 2022.
- [77] Q. Zou, L. A. Torres, S. B. Fain, N. S. Higano, A. J. Bates, and M. Jacob, “Dynamic imaging using motion-compensated smoothness regularization on manifolds (MoCo-SToRM),” *Phys. Med. Biol.*, vol. 67, no. 14, Jul. 2022, Art. no. 144001.
- [78] J. Yoo, K. H. Jin, H. Gupta, J. Yerly, M. Stuber, and M. Unser, “Time-dependent deep image prior for dynamic MRI,” *IEEE Trans. Med. Imag.*, vol. 40, no. 12, pp. 3337–3348, Dec. 2021.
- [79] W. Huang, H. Li, J. Pan, G. Cruz, D. Rueckert, and K. Hammernik, “Neural implicit k-space for binning-free non-Cartesian cardiac MR imaging,” 2022, *arXiv:2212.08479*.
- [80] J. Feng, R. Feng, Q. Wu, Z. Zhang, Y. Zhang, and H. Wei, “Spatiotemporal implicit neural representation for unsupervised dynamic MRI reconstruction,” 2022, *arXiv:2301.00127*.
- [81] T. Catalán, M. Courdurier, A. Osses, R. Botnar, F. S. Costabal, and C. Prieto, “Unsupervised reconstruction of accelerated cardiac cine MRI using neural fields,” 2023, *arXiv:2307.14363*.
- [82] V. Spieker et al., “ICoNIK: Generating respiratory-resolved abdominal MR reconstructions using neural implicit representations in k -space,” 2023, *arXiv:2308.08830*.
- [83] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image quality assessment: From error visibility to structural similarity,” *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
- [84] C. Studholme, D. L. G. Hill, and D. J. Hawkes, “An overlap invariant entropy measure of 3D medical image alignment,” *Pattern Recognit.*, vol. 32, no. 1, pp. 71–86, Jan. 1999.
- [85] H. R. Sheikh and A. C. Bovik, “Image information and visual quality,” *IEEE Trans. Image Process.*, vol. 15, no. 2, pp. 430–444, Feb. 2006.
- [86] O. Dietrich, J. G. Raya, S. B. Reeder, M. F. Reiser, and S. O. Schoenberg, “Measurement of signal-to-noise ratios in MR images: Influence of multichannel coils, parallel imaging, and reconstruction filters,” *J. Magn. Reson. Imag.*, vol. 26, no. 2, pp. 375–385, 2007.
- [87] E. Krotkov, “Focusing,” *Int. J. Comput. Vis.*, vol. 1, no. 3, pp. 223–237, 1988.
- [88] M. Hossain and M. N. Sulaiman, “A review on evaluation metrics for data classification evaluations,” *Int. J. Data Mining Knowl. Manage. Process.*, vol. 5, no. 2, pp. 1–11, Mar. 2015.
- [89] Y. Zhao et al., “Localized motion artifact reduction on brain MRI using deep learning with effective data augmentation techniques,” in *Proc. IJCNN*, 2021, pp. 1–9.
- [90] A. Johansson, J. M. Balter, and Y. Cao, “Abdominal DCE-MRI reconstruction with deformable motion correction for liver perfusion quantification,” *Med. Phys.*, vol. 45, no. 10, pp. 4529–4540, Oct. 2018.
- [91] W. P. Segars, G. Sturgeon, S. Mendonca, J. Grimes, and B. M. W. Tsui, “4D XCAT phantom for multimodality imaging research,” *Med. Phys.*, vol. 37, no. 9, pp. 4902–4915, Sep. 2010.
- [92] E. Shimron, J. I. Tamir, K. Wang, and M. Lustig, “Implicit data crimes: Machine learning bias arising from misuse of public data,” *Proc. Nat. Acad. Sci. USA*, vol. 119, no. 13, Mar. 2022, Art. no. e2117203119.
- [93] Á. Nárai et al., “Movement-related artefacts (MR-ART) dataset of matched motion-corrupted and clean structural MRI brain scans,” *Sci. Data*, vol. 9, no. 1, p. 630, Oct. 2022.
- [94] J. Zbontar et al., “FastMRI: An open dataset and benchmarks for accelerated MRI,” 2018, *arXiv:1811.08839*.
- [95] H. El-Rewaady et al., “Multi-domain convolutional neural network (MD-CNN) for radial reconstruction of dynamic cardiac MRI,” *Magn. Reson. Med.*, vol. 85, no. 3, pp. 1195–1208, 2021.
- [96] C. Chen et al., “OCMR (v1.0)-open-access multi-coil k-space dataset for cardiovascular magnetic resonance imaging,” 2020, *arXiv:2008.03410*.
- [97] D. J. Lin, P. M. Johnson, F. Knoll, and Y. W. Lui, “Artificial intelligence for MR image reconstruction: An overview for clinicians,” *J. Magn. Reson. Imag.*, vol. 53, no. 4, pp. 1015–1028, Apr. 2021.
- [98] K. Hammernik and M. Akçakaya, “Artificial intelligence for image enhancement and reconstruction in magnetic resonance imaging,” in *Artificial Intelligence in Cardiothoracic Imaging* (Contemporary Medical Imaging), Cecco, C. N. de Cecco, M. van Assen, and T. Leiner, Eds. Berlin, Germany: Springer, 2022, pp. 125–138.
- [99] N. R. F. Huttinga, T. Bruijnen, C. A. T. van den Berg, and A. Sbrizzi, “Gaussian processes for real-time 3D motion and uncertainty estimation during MR-guided radiotherapy,” 2022, *arXiv:2204.09873*.
- [100] A. Plumley, L. Watkins, M. Treder, P. Liebig, K. Murphy, and E. Kopanoglu, “Rigid motion-resolved prediction using deep learning for real-time parallel-transmission pulse design,” *Magn. Reson. Med.*, vol. 87, no. 5, pp. 2254–2270, May 2022.
- [101] K. Hammernik et al., “Physics-driven deep learning for computational magnetic resonance imaging: Combining physics and machine learning for improved medical imaging,” *IEEE Signal Process. Mag.*, vol. 40, no. 1, pp. 98–114, Jan. 2023.
- [102] G. Haskins, U. Kruger, and P. Yan, “Deep learning in medical image registration: A survey,” *Mach. Vis. Appl.*, vol. 31, pp. 1–18, Feb. 2020.
- [103] W. Shi et al., “AFFIRM: Affinity fusion-based framework for iteratively random motion correction of multi-slice fetal brain MRI,” *IEEE Trans. Med. Imag.*, vol. 42, no. 1, pp. 209–219, Jan. 2023.
- [104] J. Xu et al., “NeSVoR: Implicit neural representation for slice-to-volume reconstruction in MRI,” *IEEE Trans. Med. Imag.*, vol. 42, no. 6, pp. 1707–1719, Jun. 2023.
- [105] J. Lehtinen et al., “Noise2Noise: Learning image restoration without clean data,” 2018, *arXiv:1803.04189*.
- [106] H. Eichhorn et al., “Evaluating the match of image quality metrics with radiological assessment in a dataset with and without motion artifacts,” in *Proc. ISMRM Annu. Meeting*, 2022, p. 2.
- [107] G. Reguig, M. Lapert, S. Lehericy, and R. Valabregue, “Global displacement induced by rigid motion simulation during MRI acquisition,” 2022, *arXiv:2204.03522*.
- [108] T. Küstner et al., “A machine-learning framework for automatic reference-free quality assessment in MRI,” *Magn. Reson. Imag.*, vol. 53, pp. 134–147, Nov. 2018.
- [109] S. J. Esses et al., “Automated image quality evaluation of T₂-weighted liver MRI utilizing deep learning architecture,” *J. Magn. Reson. Imag.*, vol. 47, no. 3, pp. 723–728, Mar. 2018.
- [110] S. J. Sujit, I. Coronado, A. Kamali, P. A. Narayana, and R. E. Gabr, “Automated image quality evaluation of structural brain MRI using an ensemble of deep learning networks,” *J. Magn. Reson. Imag.*, vol. 50, no. 4, pp. 1260–1267, Oct. 2019.
- [111] F.-A. Croitoru, V. Hondru, R. T. Ionescu, and M. Shah, “Diffusion models in vision: A survey,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 9, pp. 10850–10869, Sep. 2023.
- [112] V. Sitzmann, J. Martel, A. Bergman, D. Lindell, and G. Wetzstein, “Implicit neural representations with periodic activation functions,” in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 33, 2020, pp. 7462–7473.
- [113] F. Shamshad et al., “Transformers in medical imaging: A survey,” *Med. Image Anal.*, vol. 88, Aug. 2023, Art. no. 102802.