

# Deep learning based correction of RF field induced inhomogeneities for T2w prostate imaging at 7 T

**Citation for published version (APA):**

Harreveld, S. D., Meliado, E. F. M., van Lier, A. L. H. M. W., Reesink, D., Meijer, R. P., Pluim, J. P. W., & Raaijmakers, A. J. E. (2023). Deep learning based correction of RF field induced inhomogeneities for T2w prostate imaging at 7 T. *NMR in Biomedicine*, 36(12), Article e5019. <https://doi.org/10.1002/nbm.5019>

**Document license:**

CC BY-NC-ND

**DOI:**

[10.1002/nbm.5019](https://doi.org/10.1002/nbm.5019)

**Document status and date:**

Published: 01/12/2023

**Document Version:**

Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

**Please check the document version of this publication:**

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

**General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

[www.tue.nl/taverne](http://www.tue.nl/taverne)



**Take down policy**

If you believe that this document breaches copyright please contact us at:

[openaccess@tue.nl](mailto:openaccess@tue.nl)

providing details and we will investigate your claim.

# Deep learning based correction of RF field induced inhomogeneities for T2w prostate imaging at 7 T

Seb D. Harrevelt<sup>1</sup>  | Ettore F. M. Meliado<sup>2</sup> | Astrid L. H. M. W. van Lier<sup>2</sup> |  
Daan Reesink<sup>3</sup>  | Richard P. Meijer<sup>3</sup> | Josien P. W. Pluim<sup>1</sup> |  
Alexander J. E. Raaijmakers<sup>1,2</sup>

<sup>1</sup>Department of Biomedical Engineering, Technische Universiteit Eindhoven, Eindhoven, The Netherlands

<sup>2</sup>Department of Radiotherapy, UMC Utrecht, Utrecht, The Netherlands

<sup>3</sup>Department of Oncological Urology, UMC Utrecht, Utrecht, The Netherlands

## Correspondence

Seb D. Harrevelt, Department of Biomedical Engineering, Technische Universiteit Eindhoven, Eindhoven, Harrevelt, Postbus 513, 5600 MB, Eindhoven, The Netherlands.  
Email: [s.d.harrevelt@tue.nl](mailto:s.d.harrevelt@tue.nl)

## Abstract

At ultrahigh field strengths images of the body are hampered by  $B_1$ -field inhomogeneities. These present themselves as inhomogeneous signal intensity and contrast, which is regarded as a “bias field” to the ideal image. Current bias field correction methods, such as the N4 algorithm, assume a low frequency bias field, which is not sufficiently valid for T2w images at 7 T. In this work we propose a deep learning based bias field correction method to address this issue for T2w prostate images at 7 T. By combining simulated  $B_1$ -field distributions of a multi-transmit setup at 7 T with T2w prostate images at 1.5 T, we generated artificial 7 T images for which the homogeneous counterpart was available. Using these paired data, we trained a neural network to correct the bias field. We predicted either a homogeneous image (t-Image neural network) or the bias field (t-Biasf neural network). In addition, we experimented with the single-channel images of the receive array and the corresponding sum of magnitudes of this array as the input image. Testing was carried out on four datasets: the test split of the synthetic training dataset, volunteer and patient images at 7 T, and patient images at 3 T. For the test split, the performance was evaluated using the structural similarity index measure, Wasserstein distance, and root mean squared error. For all other test data, the features Homogeneity and Energy derived from the gray level co-occurrence matrix (GLCM) were used to quantify the improvement. For each test dataset, the proposed method was compared with the current gold standard: the N4 algorithm. Additionally, a questionnaire was filled out by two clinical experts to assess the homogeneity and contrast preservation of the 7 T datasets. All four proposed neural networks were able to substantially reduce the  $B_1$ -field induced inhomogeneities in T2w 7 T prostate images. By visual inspection, the images clearly look more homogeneous, which is

**Abbreviations:** BOLD, blood oxygen level diffusion; CLEAR, constant level appearance; DICOM, Digital Imaging and Communications in Medicine; GLCM, gray level co-occurrence matrix; GPU, graphical processing unit; MRA, magnetic resonance angiography; MRI, magnetic resonance imaging; MRS, magnetic resonance spectroscopy; PURE, phased array uniformity enhancement; RF, radio frequency; RMSE, root mean squared error; ROI, region of interest; SENSE, sensitivity encoding; SNR, signal-to-noise ratio; SSIM, structural similarity measure; SWI, susceptibility weighted imaging; t-Biasf, network with a bias field as target image; t-Image, network with a homogeneous image as target image; T1w, T1 weighted; T2w, T2 weighted; TIAMO, time interleaved acquisition of modes; TSE, turbo spine echo; UHF, ultrahigh field; WD, Wasserstein distance.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2023 The Authors. *NMR in Biomedicine* published by John Wiley & Sons Ltd.

confirmed by the increase in Homogeneity and Energy in the GLCM, and the questionnaire scores from two clinical experts. Occasionally, changes in contrast within the prostate were observed, although much less for the t-Biasf network than for the t-Image network. Further, results on the 3 T dataset demonstrate that the proposed learning based approach is on par with the N4 algorithm. The results demonstrate that the trained networks were capable of reducing the  $B_1$ -field induced inhomogeneities for prostate imaging at 7 T. The quantitative evaluation showed that all proposed learning based correction techniques outperformed the N4 algorithm. Of the investigated methods, the single-channel t-Biasf neural network proves most reliable for bias field correction.

#### KEYWORDS

7 T, bias field removal, deep learning, neural networks

## 1 | INTRODUCTION

Imaging at ultrahigh field strength (UHF) provides higher SNR and is known to improve functional mechanisms such as BOLD, SWI, MRA, and MRS.<sup>1-3</sup>

Although UHF MRI is frequently applied for brain imaging applications, body imaging at UHF is rather sparse. One of the reasons is the particularly strong  $B_1$ -field inhomogeneity, resulting in signal and contrast inhomogeneities within the image (see, e.g., Figure 1A). Such inhomogeneities clearly reduce the attractiveness for the clinical user and may hinder the detection of anomalies. In addition, automatic image post-processing (e.g., segmentation) can experience difficulties as a result.<sup>4,5</sup> Reduction of these inhomogeneities is expected to facilitate the adoption of 7 T for clinical body imaging applications.

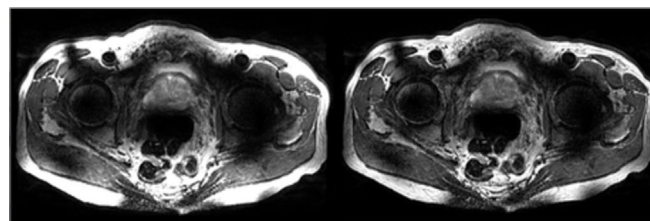
In the literature on bias field correction algorithms,<sup>6,7</sup> often an MR image  $V(x)$  is regarded as a multiplication of a homogeneous MR image of the anatomy  $U(x)$  with a distribution of inhomogeneities, referred to as the bias field  $B(x)$ , and an additive noise term  $\epsilon(x)$ .<sup>7</sup> The following equation is used as a model for MR images throughout this work:

$$V(x) = U(x) \cdot B(x) + \epsilon(x). \quad (1)$$

Bias field correction algorithms can be divided into two classes: prospective and retrospective methods. Prospective methods are based on altering the acquisition, potentially with additional hardware. At 7 T, the most used method is  $B_1$  shimming,<sup>8</sup> which can be combined with time interleaved acquisition of modes (TIAMO), where pairs of  $B_1$  shim settings are used to improve the homogeneity.<sup>9</sup> More sophisticated methods homogenize the distribution of the flip angle using an RF pulse that has been optimized for the interplay of gradients and RF waveforms.<sup>10</sup>

Retrospective methods (for example, the methods N3,<sup>11</sup> N4,<sup>12</sup> and BiCal<sup>13</sup>) are all based on the assumption that the bias field behaves as a smoothly varying and low frequency field. Other methods assume that the bias field follows a specific function or distribution.<sup>7,14-16</sup>

Most scanner vendors also offer a retrospective homogeneity algorithm such as CLEAR or PURE. These methods require a brief surface coil sensitivity calibration scan prior to imaging. The obtained receive array sensitivity maps are then used to correct the image for the inhomogeneous sensitivity patterns of each coil array element, resulting in an image with improved homogeneity. The transmit field inhomogeneities, however,



**FIGURE 1** Left, Example of a T2w prostate acquisition at 7 T. Note the hyper-intense regions in the subcutaneous fat layer and the low signal regions surrounding the femoral heads, which are characteristic of T2w images at 7 T. Right, Example of the mediocre effectiveness of the N4 algorithm on a 7 T image.

are not corrected by this method. Also, the method is less optimal at 7 T because of the absence of a so-called body coil to acquire the homogeneous reference image. For more examples we refer to Vovk et al.<sup>6</sup> for a complete review of classic bias field correction methods.

In recent years learning based methods have increased in popularity and it has been demonstrated that this methodology can excel in various tasks. In this work, we trained four neural networks for bias field correction of prostate images at 7 T to alleviate the  $B_1$ -field induced signal inhomogeneities. To train a neural network that can correct for the bias field, we require a dataset that pairs an inhomogeneous image to its homogeneous version or bias field. However, both the ground truth bias field  $B(x)$  and homogeneous image  $U(x)$  are unattainable via measurements at the MRI scanner at 7 T and above (the MRI scanners with a field strength of  $<7$  T can effectively use a body coil to acquire near-homogeneous images.) Therefore, an approximation of these variables is needed to facilitate the neural network training.

The following works all used a deep learning approach for bias field corrections but used different strategies to create a dataset for training. Venkatesh et al.<sup>17</sup> used a set of basis functions to generate a bias field for brain images from BrainWeb; the work of Dai et al.,<sup>18</sup> Chuang et al.,<sup>19</sup> and Gaillochet et al.<sup>20</sup> used the results from the N4 algorithm as substitute for a ground truth image to train a neural network; in the work of Wan et al.,<sup>21</sup> a neural network was trained to remove the bias field in conjunction with a segmentation task to avoid the need for a homogeneous ground truth image; in the work of Goldfryd et al.,<sup>22</sup> the bias field was generated by third order polynomials; Simkó et al.<sup>23</sup> used a Gaussian covariance model to generate a bias field; Nelamangala et al.<sup>24</sup> used a bias field obtained from the Human Connectome Project, which used the  $T_1$  weighted ( $T_1w$ ) and  $T_2$  weighted ( $T_2w$ ) brain images to approximate a bias field.<sup>25</sup>

In this work we created a synthetic 7 T dataset of  $T_2w$  prostate images, which was based on simulated  $B_1$  distributions of the eight-element coil array shown in the work of Raaijmakers et al.,<sup>26</sup> and  $T_2w$  prostate images obtained at 1.5 T. Note that this paper focuses only on  $T_2w$  spin echo sequences as they are the workhorse for diagnosis and treatment planning of prostate cancer.

The goal of this study was to correct the  $B_1$  induced bias field in prostate images acquired at 7 T. To this end, we trained four neural networks on the synthetic 7 T dataset. For evaluation we used patient data acquired at 7 T and to demonstrate the generalizability of this method we also evaluated the trained networks on a 3 T patient dataset.

A last note about terminology in this work. The words *neural networks* and *networks* are used interchangeably, as well as *input image*, *uncorrected image*, and *inhomogeneous image*. The *corrected image* will always refer to the bias field corrected image. When we refer to (prostate) images we refer to  $T_2w$  prostate images unless stated otherwise.

## 2 | METHODS

The goal of this study was to correct the  $B_1$  induced bias field in prostate images acquired at 7 T.

To this end, we compared the performance of four neural networks trained on the synthetic 7 T dataset with images corrected by the N4 algorithm.

These four neural networks differ from each other in two aspects. Two of them used a different input image: one used individual single-element receive array images as input (multi-channel network) and one required a coil-combined image as input (single-channel network). In addition to this, we varied the type of target image during training: the network either was trained to predict the homogeneous image  $U(x)$  directly (t-Image neural network) or predicted the bias field image  $B(x)$  instead to perform the bias field correction (t-Biasf neural network).

For evaluation we used the test split of the synthetic 7 T dataset, and two datasets acquired at 7 T: one contains prostate images from seven healthy volunteers, the other contains prostate images of four patients diagnosed with prostate cancer. Further, to demonstrate the generalizability to the clinical field strength, we also evaluated these networks trained on synthetic 7 T data on eight patients at 3 T.

For these datasets, the ground truth image is unknown and thus quantitative performance evaluation is challenging.<sup>27,28</sup> To address this problem, we used features derived from the gray level co-occurrence matrix (GLCM)<sup>29</sup> when a target image was absent. Additionally, two clinical experts graded the four patient images and seven volunteer images at 7 T by filling out a questionnaire about the contrast and homogeneity of the corrected images of two trained neural networks.

To validate the robustness of the neural networks we performed the bias field correction iteratively to measure the convergence behavior of our method.

The creation of the synthetic dataset is explained first, followed by the details of the training procedure. Next, we discuss the post-processing steps for the bias field correction, and we conclude with the evaluation procedure.

### 2.1 | Training data creation

In this section we will elaborate on the definition of each term of Equation (1), since this equation was used as the basis for the creation of the training data.

As a substitute for the homogeneous image  $U(x)$  we used T2w prostate images at 1.5 T. Here, we assumed that the 1.5 T image serves as a substitute for a perfectly homogeneous image at 7 T. Despite the change in contrast, we consider the difference between homogeneous images at 1.5 T and 7 T to be minor in comparison with the RF-field induced inhomogeneities at 7 T.

These images come from an existing dataset of 40 patients, which was obtained from our in-house patient database. Retrospective use of clinical image data for this study was approved by our local medical ethical committee (20-519). These images contained T2w prostate images with 90 slices, where the field of view ranged from (400 mm × 400 mm) to (446 mm × 446 mm). See Table 2 (later) for the acquisition parameters and Figure 2 for a schematic overview of the data creation.

The bias field  $B(x)$  was based on simulated RF transmit ( $B_1^+$ ) and receive ( $B_1^-$ ) field distributions of the eight-transceive-coil array by Raaijmakers et al.<sup>26</sup> These simulations were obtained using Sim4Life (ZMT, Zurich, Switzerland) and a set of 23 body models.<sup>30</sup> Each of the 23 simulations contained the  $B_1^+$  and  $B_1^-$  distribution of a fractionated dipole antenna array with eight elements.

These  $B_1^-$ -distributions were transformed to a bias field that is associated with a turbo spin echo acquisition and follows the body contours of a 1.5 T image.

To this end, the following steps were performed.

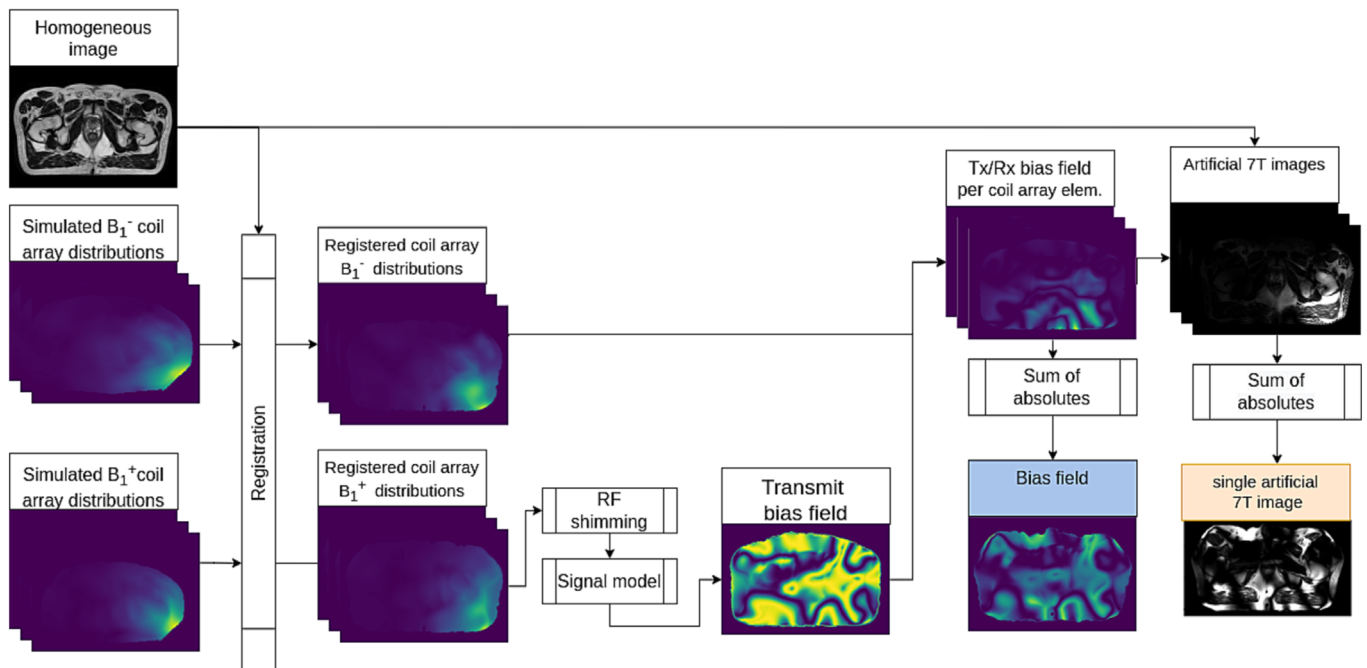
First, a registration step was needed to align a simulated  $B_1^-$ -distribution to one of the 1.5 T images. To accomplish this, we registered the body mask of the  $B_1^-$ -distributions to the body mask of the 1.5 T image using elastix.<sup>31,32</sup> The registration optimized a rigid and affine transformation map that was subsequently used to transform the  $B_1^-$ -distributions.

To ensure decent quality registrations the result was only accepted when a Dice score of 0.90 or higher was obtained between the body mask of the 1.5 T image and the body mask of the transformed  $B_1^-$ -field distribution.

Second, realistic  $B_1^+$  – field distributions were constructed from the individual coil array element  $B_1^+$  – field distributions. For this purpose, the phase distributions of the eight  $B_1^+$  – fields were normalized to that of the first coil array element, after which a shimming procedure was performed.<sup>8</sup> The following objective function was used to optimize the real valued phase settings  $\phi_n$ :

$$\max_{\phi_n} \frac{\mu(f(\phi_n))}{\sigma(f(\phi_n))}. \quad (2)$$

Here,  $\mu(\dots)$  is the mean operation and  $\sigma(\dots)$  the standard deviation of the spin echo signal function  $f(\phi_n)$ , which is defined as



**FIGURE 2** Schematic representation of the processing pipeline to create synthetic 7 T images from a homogeneous 1.5 T image. First, the simulated 7 T RF field distributions are registered to the body contour of the 1.5 T prostate image, followed by a RF shimming routine on the  $B_1^+$ -field distributions. Second, the obtained transmit field is input to the signal model to imitate the signal of a T2w image. By multiplying the obtained signal with the registered receive sensitivities ( $B_1^-$ -fields) we obtained the bias field per coil. Note that the colored (blue and orange) boxes act as target images for the neural networks.

$$f(\phi_n) = \sin^3 \left( \gamma \tau \left| \sum_{n=1}^8 \exp(\phi_n i) B_{1,n}^+(x_{ROI}) \right| \right) \quad (3)$$

where  $\gamma$  is the gyro-magnetic ratio and  $\tau$  is the pulse duration. These are set to scale the average flip angle to  $90^\circ$  in the region of interest  $x_{ROI}$ , which defines a close region around the prostate.

This ROI was chosen as a square region with size 10% of the maximum width or length of the image, where the center of the ROI coincides with the center of the pelvis.

Third, a flip angle map  $\alpha(x)$  was approximated from the shimmed  $B_1^+$  distribution by the following linear scaling:

$$\alpha(x) = C \left| \sum_{n=1}^8 \exp(\phi_n i) B_{1,n}^+(x) \right| \quad (4)$$

where the scalar  $C$  is set such that the region  $x_{ROI}$  had an average flip angle of  $90^\circ$ .

Fourth, the signal model shown in Equation (5) was used to approximate the signal received by a turbo spin echo with  $N$  refocusing pulses<sup>33</sup> given a certain flip angle map  $\alpha$ .

$$S(\alpha) = \exp(T_E/T_2) \frac{\sin(\alpha) \left( 1 - (\cos(\beta))^N \exp(-T_R/T_1) - M(N, \alpha, \beta) \right)}{1 - \cos(\alpha) (\cos(\beta))^N \exp(-T_R/T_1)} \quad (5)$$

where

$$M(N, \alpha, \beta) = \sum_{m=1}^N (1 - \cos(\beta)) (\cos(\beta))^{N-m} \exp(-(T_R - (2m-1)T_E/2)/T_1). \quad (6)$$

Here,  $\beta = 2\alpha$  and  $N$  is the TSE factor, which was set to 15. The  $T_1$  and  $T_2$  values in Equations (5) and (6) were based on the average between the values of subcutaneous fat and muscle at 7 T: a  $T_1$  value of 1067.5 ms and a  $T_2$  value of 34.5 ms.<sup>34</sup> The  $T_R$  was set to 2500 ms and the  $T_E$  to 90 ms,<sup>35</sup> which corresponds to the acquisition parameters of our T2w scans at 7 T.

Note that both Equations (3) and (5) model a signal response, where the former represents a spin echo and the latter a turbo spin echo with  $N$  refocusing pulses. For the sake of simplicity (independence from  $T_R$ ,  $T_E$ ,  $T_1$ , and  $T_2$ ) we used Equation (3) to optimize the RF shimming coefficients  $\phi_n$  and not Equation (5).

Finally, given the signal response to the flip angle map  $\alpha$  we obtain the bias field per coil  $j$  by multiplying this with the simulated  $B_1^-$  distributions.

$$B_j(x) = B_{1,j}^-(x) \cdot S(\alpha), j = 1, \dots, 8. \quad (8)$$

The noise term  $\epsilon_j(x)$  followed a complex valued normal distribution  $N(0, \sigma^2)$  with a fixed variance for all coils and no noise correlation, and the imaginary and real parts followed the same distribution. The variance of this distribution was set to  $\frac{2}{(4-\pi)n_{coil} \text{SNR}^2}$ , where the SNR ranged between 8 and 20. Here, we used no correlation between the individual coil elements; per coil element, the real and imaginary noise components had equal variances.

With the defined homogeneous image  $U(x)$ , bias field  $B_j(x)$ , and noise term  $\epsilon_j(x)$  per coil  $j$ , the synthetic 7 T image is given by

$$V_j(x) = B_j(x) \cdot U(x) + \epsilon_j(x), j = 1, \dots, 8. \quad (9)$$

The set of images defined by Equation (9) was used to train the multi-channel neural network. For training of the single-channel network we used the sum-of-magnitude image as input:

$$V(x) = \sum_{j=1}^8 |V_j(x)|. \quad (10)$$

Here, the sum of magnitudes of the individual coil array element images was taken to reduce any destructive interference over the receive channels.

## 2.2 | Training procedure

The resulting synthetic dataset was based on 23 sets of simulated  $B_1^-$ -field distributions and 40 prostate images at 1.5 T. We performed the data split such that each of the 23 models and 40 patients resides in either the training (70%), test (20%) or validation (10%) dataset. In case of

rounding errors, we favored the training set. For example, this means that the training set is created by combining 17 unique  $B_1$  distributions and data from 28 patients. Since each patient has 90 slices this resulted in 42 840 synthetic 7 T training images. By separating the  $B_1$  distributions and patient data before the creation of the synthetic 7 T data we guarantee that there is no information leaking into the test or validation set.

We trained four neural networks with the input and target according to Table 1. The neural networks used a ResNet-18 architecture<sup>36</sup> with EvoNorm normalization. This network had 22 048 449 parameters. The perceptual style loss<sup>37</sup> was used for optimization in combination with the  $L_1$  loss for the homogeneous networks and a relative difference  $L_{rel}$  loss with a smoothing factor in the denominator for the bias field networks:

$$L_{rel}(Y_{true}, Y_{pred}) = \frac{|Y_{true} - Y_{pred}|}{1 + Y_{true}}. \quad (11)$$

This relative difference loss was implemented to make the single- and multi-channel bias field network more robust to large intensity variations that can be present in the bias field. Since in some areas of the predicted bias field the value is equal or close to zero, the  $+1$  term in the denominator was added to avoid explosion of the loss term for small values of  $y_{true}$ .

The overall loss is then given by the following equation:

$$L(Y_{true}, Y_{pred}) = \lambda_1 H(i_{epoch} - n_{epoch}) L_X(Y_{true}, Y_{pred}) + \lambda_2 L_{percp}(Y_{true}, Y_{pred}). \quad (12)$$

Here, the parameters  $(\lambda_1, \lambda_2)$  were chosen empirically to be (5, 15) and (2, 24) when training a t-Image network and t-Biasf network, respectively. The function  $H(\cdot)$  is the Heaviside step function, which is active when  $i_{epoch} - n_{epoch}$  is larger than zero, so an additional loss was used in a later stage of the training process. The  $L_X$  loss is  $L_{rel}$  when training a t-Biasf network, or the  $L_1$  loss when training a t-Image network. When we trained the t-Biasf network we used a lag in the Heaviside step function of five epochs, which was needed for additional stability of the training process.

The Adam<sup>38</sup> optimizer was used with a starting learning rate of  $1 \times 10^{-4}$  on an NVIDIA Tesla V100 32 Gb. During training the following data augmentations were used: 10 phase shim settings that were optimized according to Equation (2) with a slight perturbation to  $x_{ROI}$  to prevent over-fitting on a single  $B_1^+$  distribution in the input; randomized order of individual coil images; TorchIO<sup>39</sup> random affine function for cropping and resizing of the image to make the network robust to size variations. For this transformation the following parameters were used: scale ranges from 0.25 to 4, degrees from  $-10^\circ$  to  $10^\circ$ , and image translation for both x and y from  $-50$  to  $50$  pixels. The rotation is performed around the center, the default padding value is zero, and isotropic scaling is used. Given the range of the random affine function and the varying shim settings, we expect to bridge the distance between the number of model parameters and the number of unique training examples.

Each network was trained for a maximum of 500 epochs. After each epoch the validation loss curve was used as a stopping criterion. If the current validation loss did not improve the average validation loss curve over the previous 20 epochs for 20 consecutive epochs, then the training was finished. In practice, this resulted in a training range of 250–500 epochs.

Finally, all images were resized to (256, 256) after data augmentation to limit GPU memory usage.

Note that the adapted phase shim settings for the data augmentation sometimes result in decreased  $B_1^+$  levels inside the prostate. Since the time duration of the pulse is set such that the prostate receives a  $90^\circ$  flip angle, for such cases the periphery of the anatomy would show more severe inversion bands and inhomogeneities.

## 2.3 | Post-processing

During inference, a patch and stitch approach is used, with a patch size of (256, 256) and stride of 128.

For the t-Biasf neural networks the sums of magnitudes of the input image(s) were divided by the output (bias field) to obtain a bias field corrected image.

**TABLE 1** Overview of the used models and their associated input and target images.

Model name	Input	Target
Single-channel t-Image (neural) network	$V = \sum_{n=1}^8  V_j(x) $	$U(x)$
Multi-channel t-Image (neural) network	$V_j(x), j = 1..8$	$U(x)$
Single-channel t-Biasf (neural) network	$V = \sum_{n=1}^8  V_j(x) $	$B(x)$
Multi-channel t-Biasf (neural) network	$V_j(x), j = 1..8$	$B(x)$

To reduce the chance of visible hallucinations from predictions of the t-Image neural networks, we used a smoothed bias field for the correction of the input image. This smoothed bias field was obtained by first dividing the sum of magnitudes of the input image(s) by the predicted homogeneous image, which resulted in a bias field. Then this bias field was smoothed by a square averaging kernel of size  $\lfloor M/32 \rfloor$ , where  $M$  is the maximum width or height of the image. This specific kernel size was obtained via trial and error. After this, the sums of magnitudes of the input image(s) were divided by the smoothed bias field to obtain the final bias field corrected image.

## 2.4 | Evaluation

Evaluation was performed on the test split from the synthetic dataset, two datasets acquired at 7 T and one at 3 T. Apart from the test split, we lacked a ground truth image, which posed a challenge to the evaluation task. To tackle this, we used three quantitative metrics and a qualitative assessment by two clinical experts. Both experts have more than 5 years of experience at the UMCU in abdominal radiology.

For the evaluation on the test split we used the structural similarity index measure (SSIM<sup>40</sup>), the Wasserstein distance (WD<sup>41</sup>), and the root mean squared error (RMSE) between the homogeneous target image and the corrected image. On datasets without a ground truth, we used features derived from the GLCM<sup>29</sup> and a relative change of these features.

The qualitative assessment consisted of a questionnaire on all 7 T data with three questions per subject that were related to the change of contrast in the prostate, homogeneity in the prostate and surrounding tissue, and homogeneity over the full body. The clinical experts graded images that were corrected by both the single-channel t-Biasf and t-Image neural networks.

A GLCM is calculated for non-overlapping patches, which have a size of 33% of the minimum width or height of the corrected image. Each patch is min-max normalized and cast to an 8-bit image to make it compatible with the implemented GLCM module.

Each GLCM contains a distribution of co-occurring pixel values at a given offset. These offsets are given by a distance and angle. In this work we chose a distance range from 1 to 5 mm with 1 mm increments and used six angles with increments of 60°.

This results in a matrix of size (256, 256, 5, 6), where the first two dimensions correspond to the number of pixel values ( $=2^8$ ) and the last two to the distance range and angles covered. Using this GLCM we derive the Homogeneity and Energy:

$$\text{Homogeneity} = \mu_{mn} \left( \sum \frac{P_{ijmn}}{1 + (i-j)^2} \right) \quad (13)$$

$$\text{Energy} = \mu_{mn} \left( \sum P_{ijmn}^2 \right) \quad (14)$$

where  $P_{ijmn}$  is a single element of the GLCM and the function  $\mu_{mn}(\cdot)$  takes the average over the patches and offsets denoted by  $m$  and  $n$ . Homogeneity is a feature that corresponds to images with a GLCM that has values that are close to the main diagonal, and the Energy feature captures a level of uniformity in the image. These features were chosen since they do not correlate with each other.<sup>42</sup>

These features derived from the GLCM offer no absolute interpretation as an RMSE would give. Therefore, we define a relative change between the GLCM-derived feature values of two images. In practice this means that we evaluate the relative change of the corrected image to its uncorrected image.

$$X_{\text{relative}} = \frac{X - X_{\text{reference}}}{X_{\text{reference}}} \quad (15)$$

where  $X$  and  $X_{\text{reference}}$  contain the GLCM-derived feature values of an image and its reference image. This feature shows how much an uncorrected image has improved in terms of the selected GLCM feature.

For the test split, for which the target image was present, we also calculate the relative change between the target images and the uncorrected images.

The volunteer dataset at 7 T used the coil setup as demonstrated in the work of Raaijmakers et al.<sup>26</sup> This dataset consists of seven volunteers, for whom we acquired one slice per subject. This small study was approved by the local medical ethics committee and all subjects signed informed consent prior to inclusion in the study. These images were acquired without SENSE or CLEAR reconstruction. The exported data contained the complex images per coil element to test the multi-channel networks.

The patient dataset at 7 T was acquired with a similar coil setup as to that demonstrated by Steensma et al.<sup>43</sup> This coil array contains additional loop receivers compared with the coil array by Raaijmakers et al.<sup>26</sup> This involves a multi-slice acquisition of four patients diagnosed with prostate cancer that contains 18 slices. During acquisition a SENSE factor of 2.5 was used, and CLEAR to improve homogeneity. For reconstruction all 24 receive elements were used. Here, the data were exported to the DICOM format, making it applicable only to the single-channel networks. A majority of these images are more heavily affected by RF inhomogeneities due to suboptimal RF shimming of the prostate.



The dataset acquired at 3 T consists of eight patients and 720 slices in total. This dataset was acquired as a treatment planning scan for patients who were scheduled for prostate cancer radiotherapy. We used this dataset to demonstrate the generalizability of the trained networks to a clinical field strength. Retrospective use of clinical image data for this study was approved by our local medical ethical committee (20-519). We refer to Table 2 for the acquisition parameters per dataset.

To demonstrate the robustness of the neural networks we performed the bias field correction in an iterative fashion for the trained neural networks. Convergence to zero of the difference between subsequent corrected images demonstrated the stability of the method.

### 3 | RESULTS

The results are presented per dataset, where we show the quantitative metrics and four examples of bias field corrected images. Note that the *target image* in this section refers to the homogeneous image  $U(x)$ , and not the bias field  $B(x)$ .

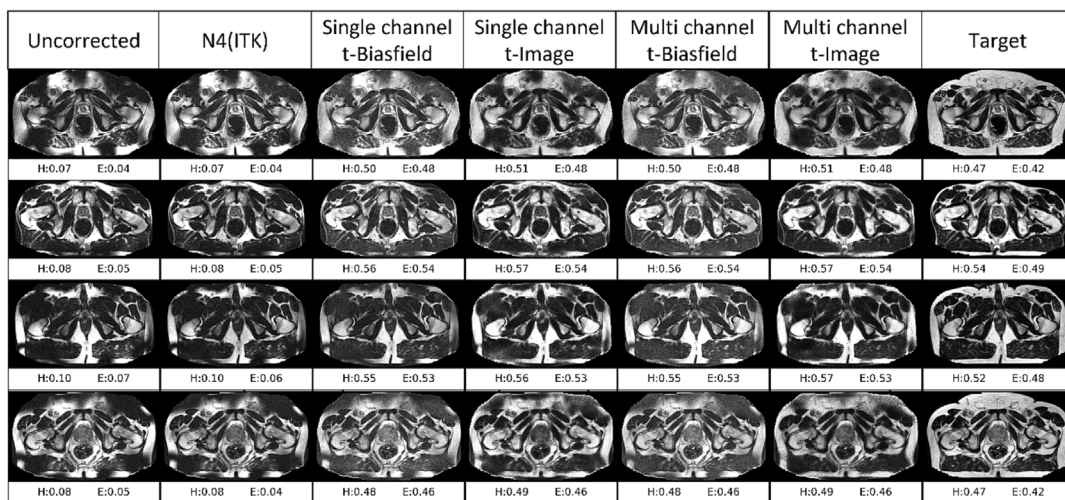
#### 3.1 | Test split

In Figures 3 and 4, we show four synthetic test images to give a visual impression of the correction. The bias field correction is most visible in the subcutaneous fat layer, where we observe a greater reduction from the results of the t-Image neural networks and a somewhat lesser reduction from the t-Biasf neural networks. However, in Figure 4, it is visible that the prostate tissue has changed for most results of the t-Image neural networks.

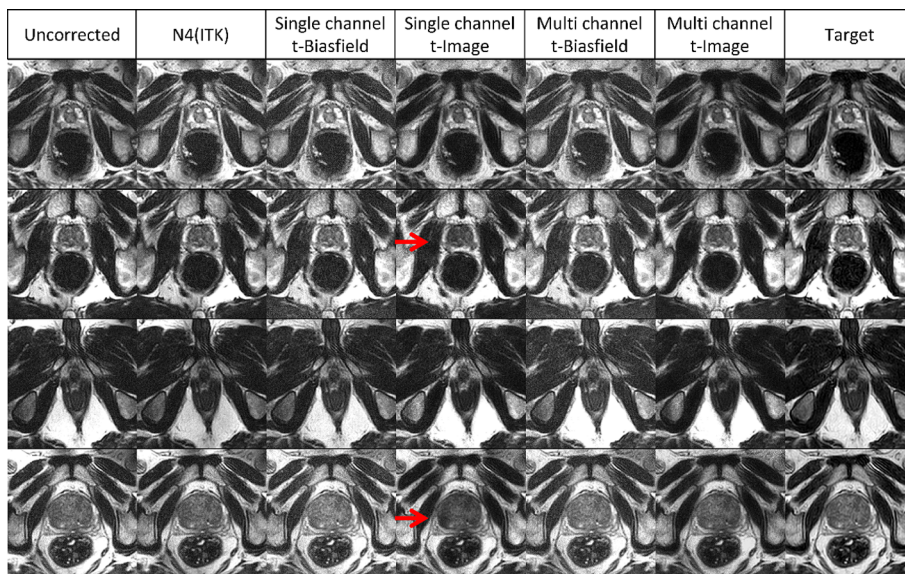
Performance metrics for the bias field correction on the test split are presented in Table 3. The inhomogeneous images in this dataset have an average Homogeneity and Energy of 0.07 and 0.04, respectively, whereas all the target images have an average Homogeneity and Energy of

**TABLE 2** Overview of the scan parameters for the test datasets. Note that ranges are indicated when the imaging parameters vary over the dataset.

Dataset	$T_R$ (ms)	$T_E$ (ms)	Pixel spacing (mm <sup>2</sup> )	Slice thickness (mm)	Number of images
1.5 T	1535–1635	120–278	0.58 <sup>2</sup> –0.88 <sup>2</sup>	2	3600
3 T	1635–5392	80–360	0.58 <sup>2</sup> –0.83 <sup>2</sup>	2–3	720
7 T volunteer	5000	90	0.49 <sup>2</sup> –0.68 <sup>2</sup>	3	7
7 T patient	10 000	140	0.28 <sup>2</sup>	3	70



**FIGURE 3** Four examples of corrected images from the test split data set. For each example, the sum of magnitudes of the input images, the result of the N4 algorithm, and the results of the four proposed neural networks are presented. In general, all investigated networks are able to significantly reduce the inhomogeneity patterns for the synthetic 7 T images. The results from the t-Image neural networks show fewer residual artifacts in the reconstructed image, although these methods can be affected by contrast changes inside the prostate, as in the second and fourth examples. The results from the t-Biasf neural networks are affected by reconstruction errors due to inaccuracies in the predicted bias field. These manifest themselves as hyperintensities.



**FIGURE 4** A close-up of the subjects shown in Figure 3. For each subject we present the sum of magnitudes of the input images, the result of the N4 algorithm, and the results of the four proposed networks. The arrow shows corrections where the tissue contrast has changed.

0.53 and 0.49, respectively. These large differences can only be attributed to the strong artificial  $B_1$  field induced inhomogeneities that were added during the creation of the data.

Looking at the first two columns of Table 3, we observe that the corrected images using the trained neural networks have Homogeneity and Energy levels that are comparable to those of the average target image. In contrast, the results from the N4 algorithm show little change compared with the Homogeneity and Energy levels of the inhomogeneous images.

The last two columns of Table 3 show the relative changes in Homogeneity and Energy between the corrected and inhomogeneous images. The relative change values for the trained neural networks are close to the target, compared with the N4 algorithm.

Expressing the performance in target based metrics, in Table 4 we present the SSIM, WD, and RMSE between the corrected image and the target image. The trained neural networks all have similar scores, where no clear “best” network variation can be chosen based on these metrics.

### 3.2 | Volunteer images at 7 T

By visual inspection of the images shown in Figure 5, we report a reduction of hypo-intense regions for all the trained neural networks; this is mostly noticeable in the subcutaneous fat layer. Further, we observe the introduction of minor hyper-intense regions in the second and last examples of the single-channel t-Biasf neural network.

In Figure 6, close-ups of the prostate from the examples shown in Figure 5 are shown to demonstrate the impact on a finer scale. For example, notice how all trained neural networks are affected by a degree of over-correction in the first example. Next, we see in the second example how the single-channel t-Biasf neural network is able to correct the hypo-intense region near the rectum wall. Further, it is clear that the multi-channel direct network was unable to correct the fatty tissue near the prostate, where a saturation of pixel intensity occurs for all examples.

In Table 5, the first two columns show the Homogeneity and Energy for the corrected images. Here, we observe an agreement between these values among the trained neural networks, whereas the results from the N4 algorithm deviate strongly from this.

The inhomogeneous images in this dataset have average Homogeneity and Energy levels of 0.13 and 0.06; we observe that these features were strongly increased by the correction using the neural networks. This is also reflected in the relative changes in Homogeneity and Energy shown in the last two columns. How close these values are to the true (relative change in) Homogeneity and Energy levels is unclear since no target image was available for this dataset.

### 3.3 | Patient images at 7 T

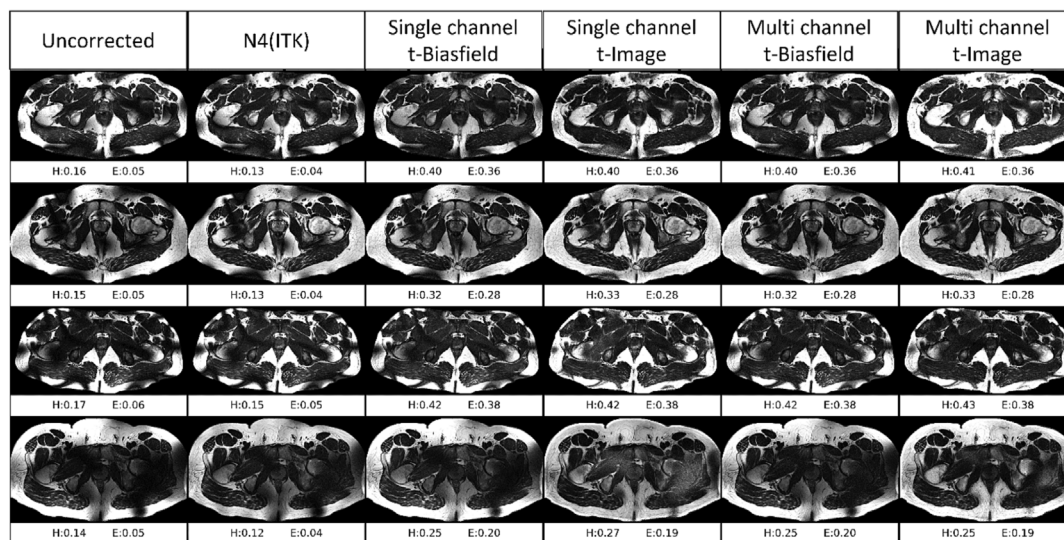
In Figure 7, we show a single slice of the patient images corrected by each model. The images corrected by the trained neural networks show a clear improvement in the visibility of the prostate. Notice how the correction of the single-channel t-Biasf neural network reduced the hypo-intense region in size but did not offer a correction for the null-signal regions in the image. This is in contrast with the results of the t-Image neural

**TABLE 3** The (relative) Homogeneity and Energy per model evaluated on the *test split* of the synthetic dataset. The feature values of the uncorrected input are used as reference for the calculation of the relative features.

Model	Metric			
	Homogeneity	Energy	Relative change in Homogeneity	Relative change in Energy
Uncorrected	0.09 ± 0.02	0.06 ± 0.02	0.0 ± 0.0	0.0 ± 0.0
N4 algorithm	0.09 ± 0.02	0.05 ± 0.02	-0.02 ± 0.03	-0.1 ± 0.09
Single-channel t-Biasf	0.56 ± 0.06	0.54 ± 0.06	5.32 ± 1.21	8.60 ± 2.28
Single-channel t-Image	0.57 ± 0.06	0.54 ± 0.06	5.46 ± 1.22	8.62 ± 2.29
Multi-channel t-Biasf	0.56 ± 0.06	0.54 ± 0.06	5.33 ± 1.21	8.60 ± 2.28
Multi-channel t-Image	0.57 ± 0.06	0.54 ± 0.06	5.49 ± 1.22	8.63 ± 2.28
Target	0.53 ± 0.06	0.49 ± 0.07	5.02 ± 1.22	7.73 ± 2.15

**TABLE 4** The quantitative metrics evaluated on the *test split* of the synthetic dataset with respect to the homogeneous target image.

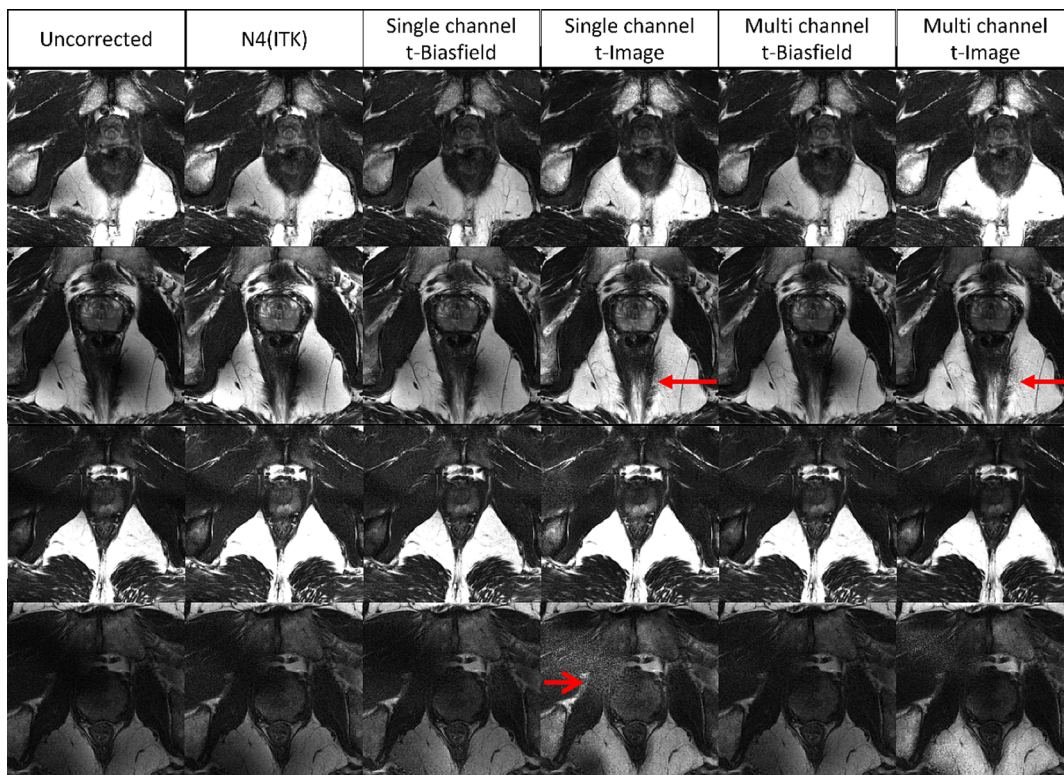
Model	Metric		
	WD	SSIM	RMSE
Uncorrected	0.45 ± 0.03	0.23 ± 0.04	9.83 ± 0.65
N4 algorithm	0.44 ± 0.04	0.24 ± 0.05	9.91 ± 0.65
Single-channel t-Biasf	0.08 ± 0.02	0.76 ± 0.05	6.33 ± 0.54
Single-channel t-Image	0.06 ± 0.01	0.73 ± 0.04	6.38 ± 0.54
Multi-channel t-Biasf	0.08 ± 0.02	0.77 ± 0.05	6.32 ± 0.56
Multi-channel t-Image	0.06 ± 0.02	0.75 ± 0.04	6.28 ± 0.51



**FIGURE 5** Corrected images of four volunteers at 7 T. For each subject, the sum of magnitudes of the input images, the result of the N4 algorithm, and the results of the four proposed neural networks are presented. In general, the inhomogeneity has been reduced for all networks except the N4 algorithm. Both single-channel networks show good performance, in terms of a sufficient bias field reduction as well as the preservation of the tissue contrast. A close-up of these images is given in Figure 6.

network, where these regions are substituted by noise. Also, similar to previous results, the t-Biasf neural network introduced minor hyper-intensities in the corrected areas.

In Figure 8, a close-up is given to emphasize the correction in the prostate itself. Notice how in the first example the lower left part of the prostate is obscured. Here, the trained neural networks are able to correct for this reduction in signal compared with the N4 algorithm. Similar corrections are found in the second and third examples, where the visibility of the rectum wall has been improved. The fourth example displays how small details are preserved after the bias field correction using the trained neural networks.



**FIGURE 6** A close-up of the subjects shown in Figure 5. For each subject, the sum of magnitudes of the input images, the result of the N4 algorithm, and the results of the four proposed networks are presented. Observe how the visibility of the fine structures in the subcutaneous fat layer is corrected by the single-channel t-Image network. The open arrow points to a region where noise has been introduced by the correction, and the normal arrow head points to a location of excellent bias field correction.

**TABLE 5** The (relative) Homogeneity and Energy per model evaluated on the volunteer images at 7 T. The feature values of the input are used as reference for the calculation of the relative features.

Model	Metric			
	Homogeneity	Energy	Relative change in Homogeneity	Relative change in Energy
Uncorrected	$0.16 \pm 0.02$	$0.05 \pm 0.01$	$0.0 \pm 0.0$	$0.0 \pm 0.0$
N4 algorithm	$0.14 \pm 0.02$	$0.05 \pm 0.01$	$-0.1 \pm 0.05$	$-0.15 \pm 0.07$
Single-channel t-Biasf	$0.39 \pm 0.06$	$0.36 \pm 0.07$	$1.51 \pm 0.56$	$5.71 \pm 1.72$
Single-channel t-Image	$0.40 \pm 0.06$	$0.35 \pm 0.07$	$1.58 \pm 0.57$	$5.67 \pm 1.76$
Multi-channel t-Biasf	$0.40 \pm 0.07$	$0.36 \pm 0.07$	$1.53 \pm 0.57$	$5.73 \pm 1.74$
Multi-channel t-Image	$0.40 \pm 0.07$	$0.35 \pm 0.07$	$1.57 \pm 0.59$	$5.68 \pm 1.77$

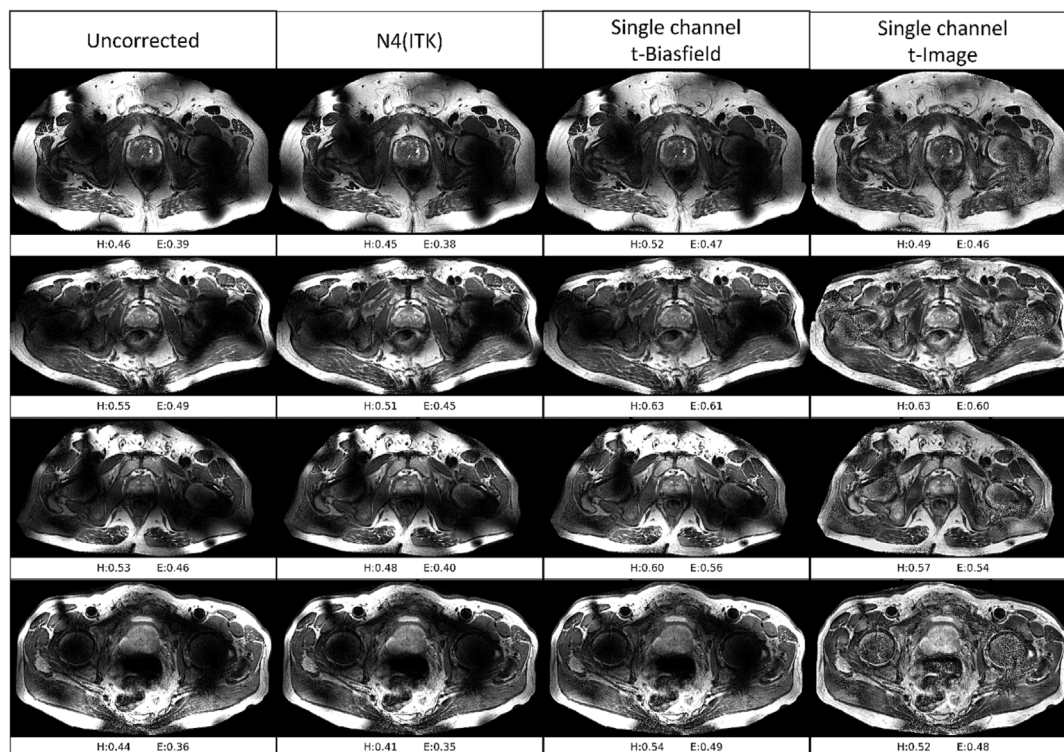
Given that the inhomogeneous images of this dataset have an average Homogeneity and Energy of 0.54 and 0.47, respectively, we observe an improvement for the trained neural networks. See Table 6 for the full details of the quantitative performance metrics.

### 3.4 | Patient images at 3 T

In Table 7, the first two columns show the Homogeneity and Energy for the corrected images. For this dataset, the average Homogeneity of the inhomogeneous images is 0.56, and the average Energy is 0.50.

Noticeable is the slight increase in Homogeneity and Energy compared to the average inhomogeneous image, as well as the agreement between the N4 algorithm and the trained neural networks.

In Figure 9 we demonstrate the bias field corrections on four prostate images at 3 T. These images show less severe inhomogeneities due to the decrease in field strength. This results in a comparable performance between the N4 algorithm and the proposed methods. The single-channel



**FIGURE 7** Bias field correction results for images of four prostate cancer patients at 7 T. Correction was performed retrospectively on DICOM data where multi-channel images are not available. For each subject, the original (CLEAR-reconstructed) input image and the results of the bias field correction method using the single-channel t-Biasf and t-Image neural networks are depicted. Both networks are able to reduce the inhomogeneity patterns, increasing the visibility of areas that were affected by the bias field.

t-Biasf and t-Image neural networks were able to correct these images, even though they were trained on much more extreme inhomogeneities. Note that the single-channel t-Biasf neural network does introduce minor hyperintensities in the second, third and fourth examples.

### 3.5 | Qualitative evaluation (questionnaire)

Two body-specialized radiologists were asked to grade the images qualitatively. They were graded on preservation of contrast and improvement of homogeneity.

The results are summarized in Figure 10, where the top row shows the evaluation of the results from the t-Biasf neural network and the bottom the t-Image neural network. Here, the red dotted line indicates the boundary between a deteriorated and a neutral or improved image quality. To assist in the reading of Figure 10, we present the score interpretations below.

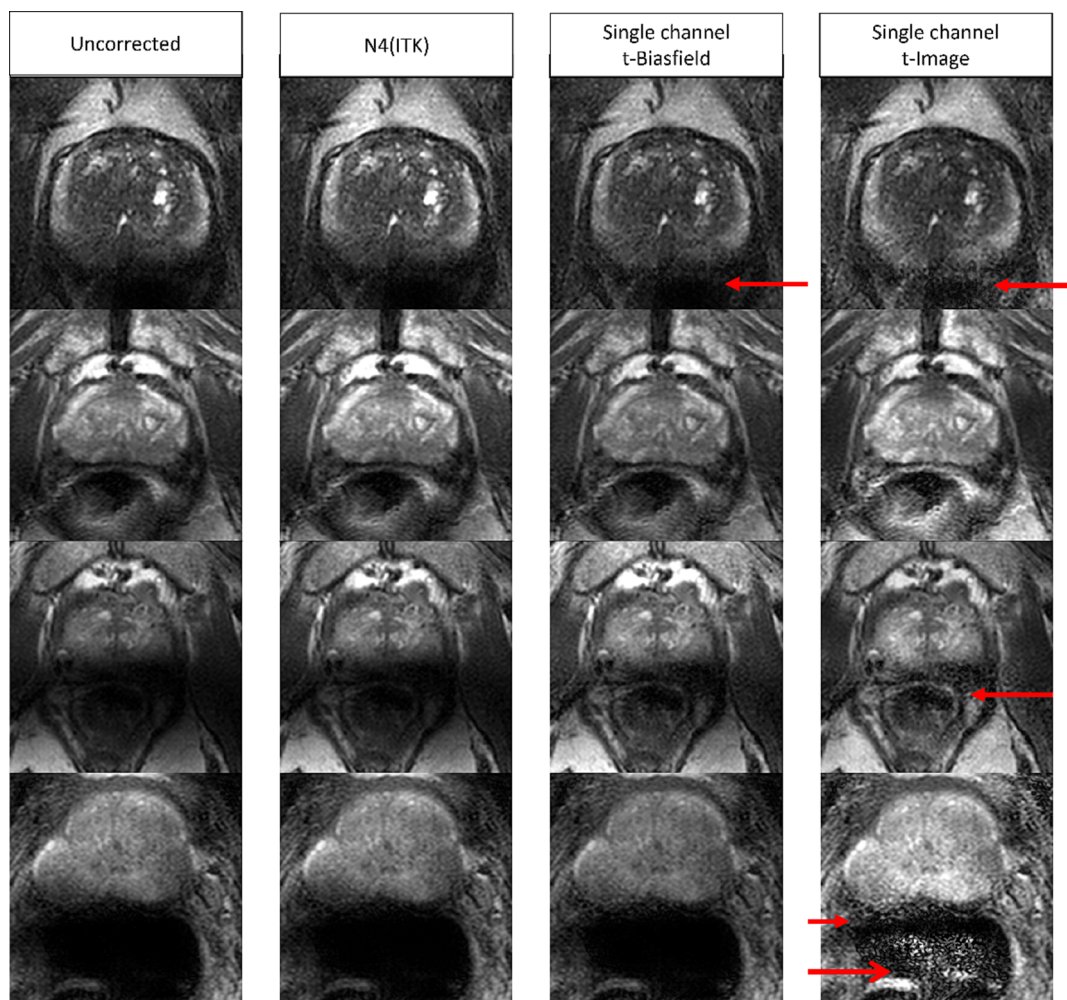
A comparison between the answers given shows an exact agreement of 35%. However, when we simplify the answers to “degraded image quality” and “not degraded image quality” (i.e., above or below the red dotted line), then an agreement of 83% is reached.

Overall, the radiologists scored positively on both contrast preservation and improvement of homogeneity. The average score on the contrast related question was 4.1 for the t-Biasf neural network, and 3.4 for the t-Image neural network. For the homogeneity-related questions an average score of 3.1 was achieved for the t-Biasf neural network, and 3.2 for the t-Image neural network.

However, one radiologist graded 3 out of 11 cases to have a deteriorated homogeneity in the prostate region. Also, both radiologists graded 2 or 3 cases to have deteriorated contrast with potential clinical impact.

The interpretation of the scores on degradation/improvement of contrast is the following.

1. Not clinically useful
2. Contrast deteriorated, possible consequences
3. Contrast deteriorated, no clinical relevance
4. Contrast has not changed
5. Contrast has improved, more visibility and possible improvement of clinical observations (improved visibility lesion, capsular invasion).



**FIGURE 8** A close-up of the subjects shown in Figure 7. For each subject, the sum of magnitudes of the input images, the result of the N4 algorithm, and the results of the four neural networks are presented. The first, third, and fourth images are prominent examples of the deep-learning based correction methods. Here, obscured anatomical structures have been recovered. The normal arrow heads point at locations with an excellent recovery of anatomical structures. The open arrow head shows an example where noise has been amplified.

**TABLE 6** The (relative) Homogeneity and Energy per model evaluated on the patient images at 7 T. The feature values of the input are used as reference for the calculation of the relative features.

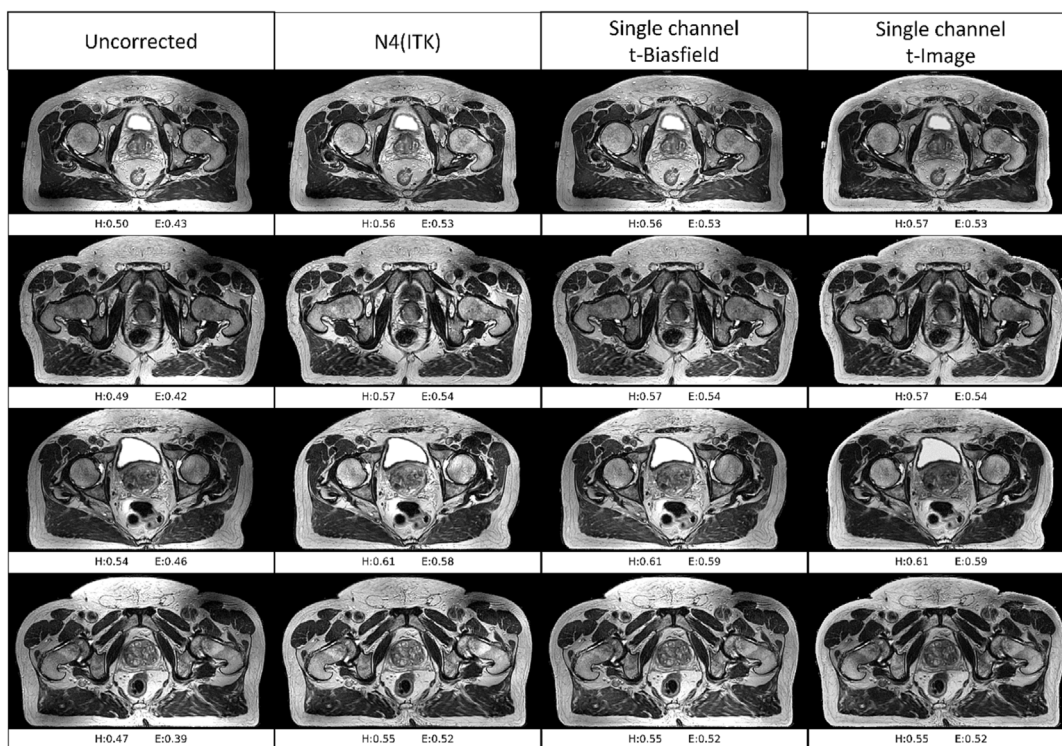
Model	Metric			
	Homogeneity	Energy	Relative change in Homogeneity	Relative change in Energy
Uncorrected	$0.52 \pm 0.05$	$0.45 \pm 0.05$	$0.0 \pm 0.0$	$0.0 \pm 0.0$
N4 algorithm	$0.49 \pm 0.04$	$0.42 \pm 0.04$	$-0.06 \pm 0.03$	$-0.07 \pm 0.04$
Single-channel t-Biasf	$0.61 \pm 0.05$	$0.57 \pm 0.06$	$0.17 \pm 0.05$	$0.26 \pm 0.06$
Single-channel t-Image	$0.58 \pm 0.06$	$0.56 \pm 0.06$	$0.12 \pm 0.06$	$0.23 \pm 0.07$

The interpretation of the scores on degradation/improvement of homogeneity (prostate region and whole body graded separately) is the following.

1. Homogeneity deteriorated, no clinical relevance
2. Homogeneity has not changed
3. Homogeneity has improved, no clinical relevance
4. Homogeneity has improved, improvement in ease of use
5. Homogeneity has improved, more visibility and possible improvement of clinical observations (improved visibility lesion, capsular invasion).

**TABLE 7** The (relative) Homogeneity and Energy per model evaluated on the patient images at 3 T. The feature values of the input are used as reference for the calculation of the relative features.

Model	Metric			
	Homogeneity	Energy	Relative change in Homogeneity	Relative change in Energy
Uncorrected	0.55 ± 0.06	0.49 ± 0.07	0.00	0.00
N4 algorithm	0.62 ± 0.06	0.59 ± 0.06	0.12 ± 0.03	0.22 ± 0.06
Single-channel t-Biasf	0.62 ± 0.06	0.59 ± 0.06	0.12 ± 0.03	0.23 ± 0.06
Single-channel t-Image	0.62 ± 0.06	0.59 ± 0.06	0.13 ± 0.03	0.23 ± 0.06



**FIGURE 9** Bias field correction results for images of four prostate cancer patients at 3 T. Correction was performed retrospectively on DICOM data where multi-channel images are not available. This dataset was used to demonstrate the applicability to lower field strength data and for a fair comparison with the result of the N4 algorithm.

### 3.6 | Iterative bias field correction

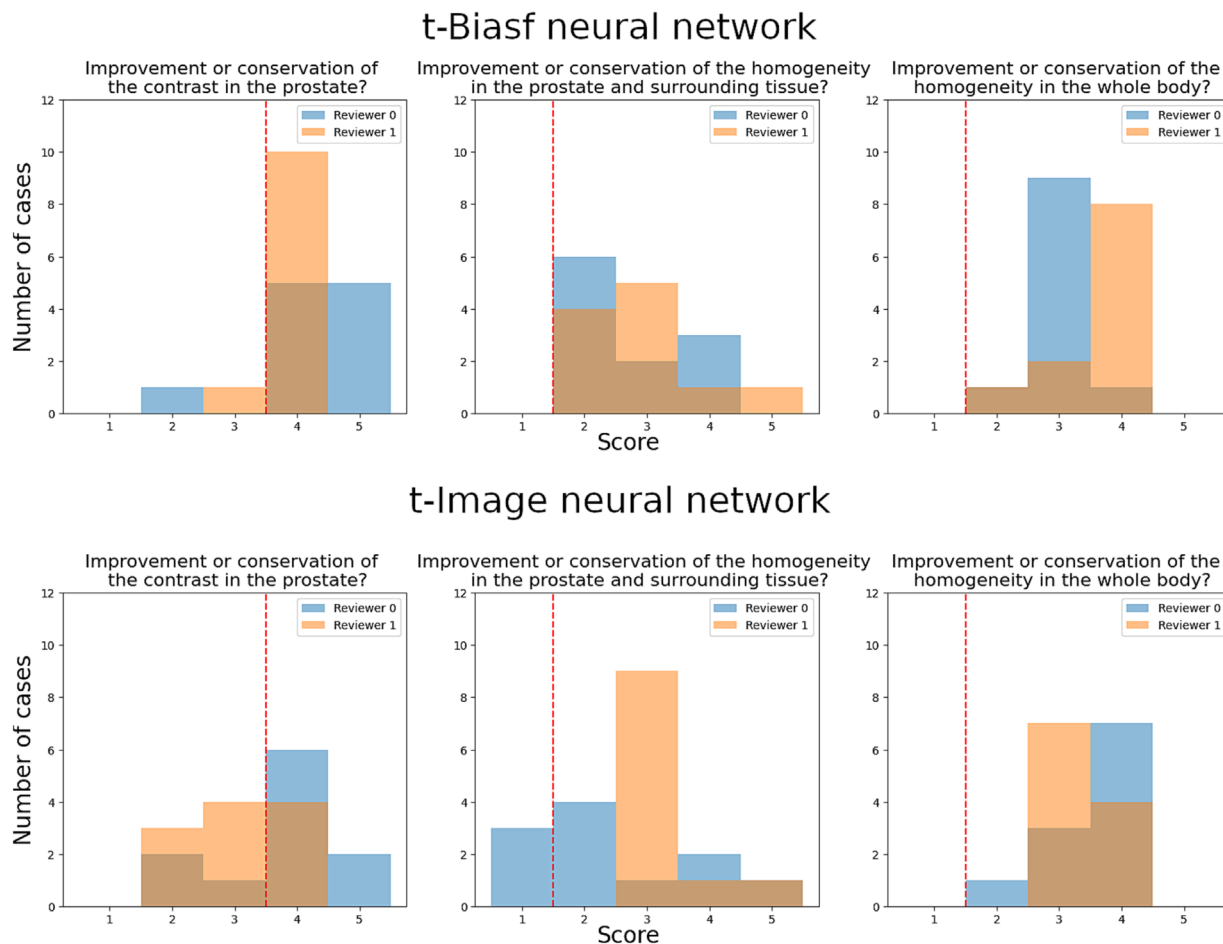
An iterative bias field correction is where we perform multiple bias field reconstructions where the input is updated by the previous bias field correction step. This is used to validate the robustness of the bias field correction as proposed by Hou,<sup>28</sup> where a desirable outcome is that the input and output will converge to the same image. To quantify this, we used the  $L_2$  norm and applied the bias field correction 10 times.

We used this method for all proposed networks on the available test datasets, where the convergence behavior is shown in Figure 11. These figures indicate that convergence is not attained on each dataset by each network, the exception being the single-channel t-Biasf neural network on the volunteer data.

Two others that stand out are the single-channel t-Biasf neural network on the test split, and the single-channel t-Image neural network on the patient data at 7 T. These converge to a non-zero value, which reflects an alternation between two solutions.

## 4 | DISCUSSION

The presence of a bias field in 7 T abdominal images is inevitable but decreases the attractiveness and the clinical applicability of the acquired images. Previous methods that correct for  $B_1$ -field induced inhomogeneities at lower field strengths were based on estimating a bias field from



**FIGURE 10** Distribution of scores over the questionnaire; the dashed (red) line indicates the boundary between deterioration and neutral or improvement. The top row shows the results from the t-Biasf neural network, and the bottom the t-Image neural network. The two colors indicate the two reviewers' choices.

the inhomogeneous image, with the assumption that the bias field is a smoothly varying and low frequency field. The bias field from acquisitions of T2w prostate images at 7 T does not satisfy these assumptions, hence conventional methods such as the N4 algorithm offer little to no improvement of the homogeneity. In this work four neural networks have been developed that retrospectively correct the bias field of T2w prostate images at 7 T.

To this end, we have created a synthetic 7 T dataset based on the assumed linear relation between a measured image, its “true” homogeneous image, and its bias field; see Equation (1). For the homogeneous image we used 1.5 T prostate images as a substitute, which were multiplied with a bias field that was based on simulated  $B_1^+$  and  $B_1^-$  fields at 7 T to create synthetic 7 T images.

Having both the ground truth homogeneous and bias field images allowed us to train two networks: one that predicts a homogeneous image (t-Image neural network), and the other that predicts the bias field (t-Biasf neural network). In addition, we explored different input dimensions: a coil-combined image (single-channel network), or the individual coil images of an eight-channel receive array (multi-channel network). The single-channel neural network is practical for retrospective use, whereas the multi-channel neural network can possibly benefit from the additional data.

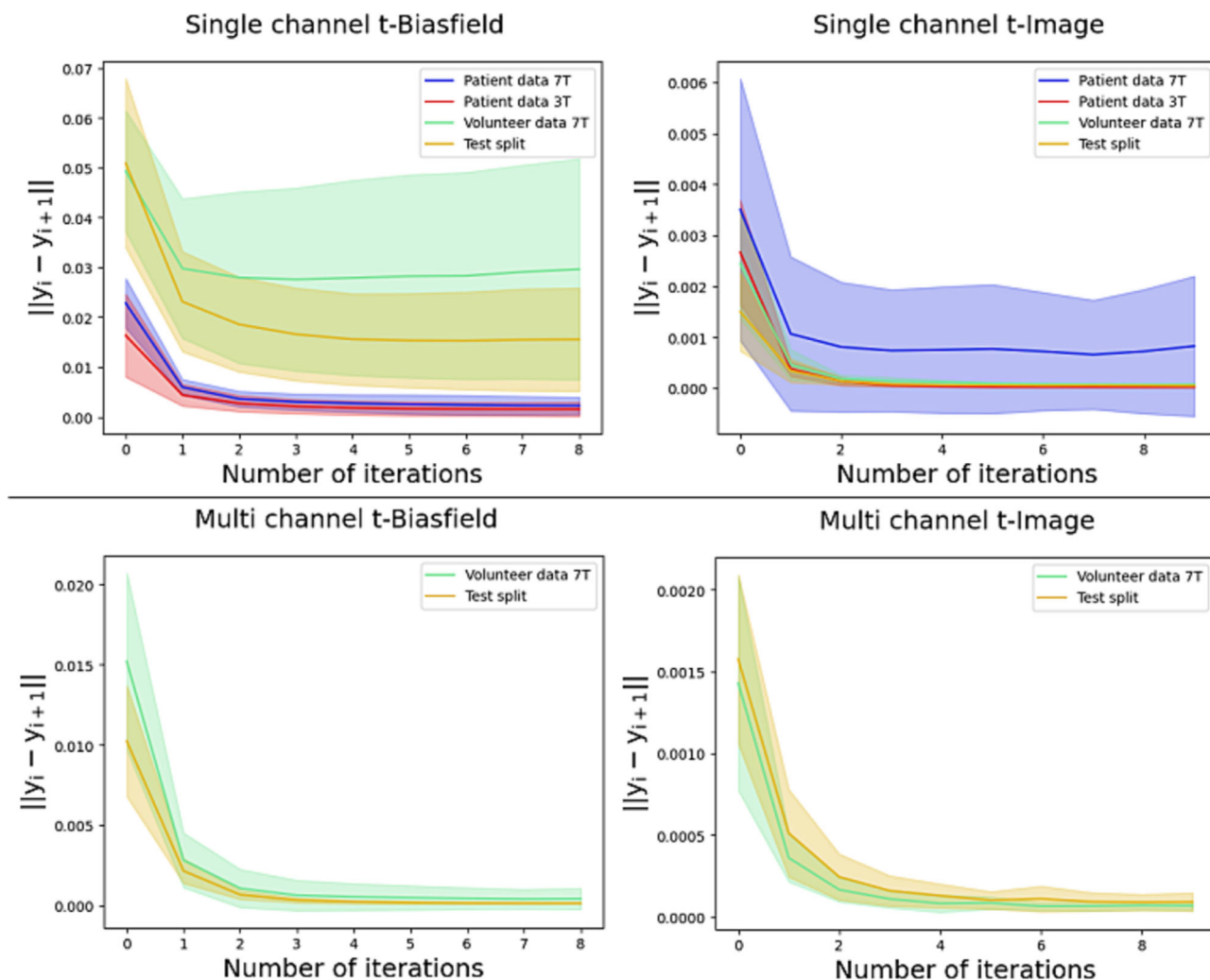
In general, the trained neural networks demonstrated a reduction of inhomogeneity for the images on the test split of the training data. Based on the SSIM, WD, and RMSE these networks show a clear improvement over the N4 algorithm. In addition, the increase in (relative) Homogeneity and Energy levels coincides with the performance gain expressed by the SSIM, WD, and RMSE.

Further, the visualization in Figure 3 affirmed the improved homogeneity expressed by the quantitative metrics.

The GLCM is a well known metric in some fields but is not often used for assessment of the bias field. From this matrix, the metrics Homogeneity and Energy can be derived as expressed by Equations (13) and (14). On average, the results on the test split showed that an increase in these metrics is in agreement with improved homogeneity.

However, the metric is not perfectly accurate. For example, Figure 3 shows that the Homogeneity is actually lower for the target image than for the corrected images, although visual inspection shows that the target image is more homogeneous. Since the relative increase in the metric





**FIGURE 11** For an iterative bias field reconstruction, we use the bias field corrected image as input to be corrected again. This procedure is repeated 10 times, where we plot the  $L_2$  norm of the difference between the input and output. For each trained network, we performed the iterative bias field correction on the available test datasets. The average norm is shown by a solid line and the shaded area shows the standard deviation of this norm.

averaged over all images does correlate well with the improvements expressed by SSIM, WD, and RMSE, these characteristics can serve as approximate indicators of the homogeneity.

The performance on both 7 T datasets was expressed using three methods. First, a visual inspection gave strong indication that the trained neural networks generalized to real world data and improved the quality of the images. Second, the increase in Homogeneity and Energy supported this observation. Third, the results from the questionnaire showed that both single-channel neural networks improved homogeneity and contrast as perceived by two clinical experts.

A minor note is that the patient images at 7 T were reconstructed with SENSE and 24 receive channels, instead of the eight-receive-channels combined sums of magnitudes that were used in the synthetic 7 T training dataset. Despite this discrepancy, the networks were able to perform the inhomogeneity correction as demonstrated by the images and quantitative results.

In fact, the corrected images from the 3 T dataset show that the trained neural networks are compatible with a completely different coil setup and less severe inhomogeneity patterns. Furthermore, both the (relative) Homogeneity and Energy in Table 7 and the images in Figure 9 demonstrate that the trained neural networks are on par with the N4 algorithm. However, the neural networks have an inference time that is 20 times faster than that of the N4 algorithm.

The current results indicate that there is little difference in the performance between the single- and multi-channel neural networks. Both the metrics displayed in Tables 3–5 and the visualized corrections in Figures 3–9 do not show dominance of one method over the other.

Only in terms of stability, shown in Figure 11, do we see a difference between the single- and multi-channel networks, with the multi-channel networks showing better convergence behavior.

With these developed networks we are able to drastically reduce the  $B_1$ -field induced inhomogeneity in prostate images at 7 T. A clear benefit of the neural networks is the fast inference time and the absence of hyper-parameters. However, care must be taken in the future implementation of such methods since the diagnostic value must be maintained at all times.

A crucial point in the bias field correction is that the contrast of the prostate must be maintained for diagnostic purposes. The metrics that have been used to assess the improvement in image quality were not sensitive to such contrast changes. Therefore, in an actual application, we would recommend the use of these bias field corrections as optional filters or check boxes in the diagnostic viewing environment. In this way, the radiologist always has the original, unprocessed image for verification.

Based on the current results, we advise the use of the t-Biasf network rather than the t-Image approach because of the apparent better preservation of contrast. In terms of single-channel or multi-channel input, our results do not show a clear difference. This will depend on the application. Retrospective application is often only possible using the single-channel network. For real-time application, arguably, one could decide to use single-channel input as well because the vendor-provided coil combination is often aided by measured coil sensitivities that allow for optimal coil combination and thereby optimize SNR in each region.

Although good results have been obtained, there are factors that can be improved. In particular, the generation of the synthetic 7 T images contains assumptions that limit the quality. An improved signal model that better reflects the behavior of the signal for very large flip angles would be one of the major potential improvements. On top of this, a signal model that includes the effect of  $B_0$  field inhomogeneity will create more realistic training data. Ultimately, training could be performed on Bloch simulated images from models with realistic distributions of  $T_1$  and  $T_2$  values so that the synthetic 7 T images used as training data would contain the loss of contrast for lower flip angles that is typical for T2w images. Improving on any of these factors will possibly raise the quality of the corrected image by making the training set more realistic.

Note that the assumption that 1.5 T images are applicable as ground truth data seems to be validated by the results. In general, we observed only an improvement of the homogeneity and no direct influence of 1.5 T-like structures in the results.

The performance evaluation can be improved by having access to tissue-specific segmentations on which the Homogeneity and Energy can be calculated. Furthermore, by having access to a tissue-segmentation network, the level of bias field correction can also be measured by the accuracy (or Dice score) when using this image as input to the tissue-segmentation network.

Further, we have not performed a thorough exploration of different neural network architectures. In this work a ResNet-18 architecture combined with the chosen losses produced promising results, and was sufficient for the current work. To explore the effect of more layers, we used a ResNet-50 architecture for a single experiment but this did not demonstrate an improvement over the ResNet-18 architecture (Figure S1). Given the limited number of training data, we resorted to the Resnet-18 architecture.

However, the current field of deep learning demonstrates that GANs or attention based mechanisms provide ample improvement for generic image-to-image-translation tasks. These novel architectures can be explored to further increase the performance of the bias field correction task. A downside is that these networks typically require larger numbers of training data, which we did not have available. As an example comparison we used the pretrained weights of the InhomoNet, provided by the work of Venkatesh et al.<sup>17</sup> This network was trained on data from BrainWeb that was made artificially more inhomogeneous by applying a mixture of basis functions. Additionally, the authors optimized their proposed network specifically for inhomogeneity correction, resulting in a model with approximately 100 M parameters. The evaluation of our test datasets showed that it is on par with the best performing network of our proposed methods (Supporting Information). Using the test split of our synthetic dataset, an SSIM of 0.70 was reached, in comparison with 0.76 for the single-channel t-Biasf approach. The metrics WD and RMSE showed similar results. Moreover, homogeneity and energy values on the measured test data set were similar to values obtained with our approach. Furthermore, the Inhomonet realized the same occasional artifacts as in our study: hyperintensities and loss of tissue contrast. All this demonstrates that a highly optimized network architecture can further benefit the quality of our inhomogeneity correction approach. However, in order to truly assess the performance gain achieved by the InhomoNet architecture combined with our approach, it would be necessary to retrain the network using a vastly larger synthetic dataset and subsequently compare its performance on our test datasets.

As a final remark, although the current methods are developed for T2w prostate images, with appropriate training the method could easily be applied to any other anatomical region or sequence as long as the appropriate  $B_1$  fields and signal equation are used to generate the bias fields for training.

## 5 | CONCLUSION

Visual inspection of the corrected images clearly shows an improvement in homogeneity for the images corrected by the neural networks with respect to the N4 algorithm. Based on the quantitative metrics we also observed an improvement for all neural networks over the correction performed by the conventional N4 algorithm. Little difference was observed between the investigated neural networks in terms of quantitative performance metrics.

However, visual inspection shows that in some cases the t-Biasf neural networks revealed fewer tissue contrast changes compared with the t-Image neural networks.

Finally, the clinical experts grading the corrected images at 7 T showed that there is a slight preference toward the single-channel t-Biasf neural network in terms of contrast preservation.

Therefore, in the possible future where a bias field removal model is used in the clinic, we advise the use of the single-channel bias field model. To preserve the original image for verification, we would recommend the correction as an optional view of the acquired image.

## ORCID

Seb D. Harrevelt  <https://orcid.org/0000-0001-8049-1898>

Daan Reesink  <https://orcid.org/0000-0002-8977-7013>

## REFERENCES

- Ladd ME, Bachert P, Meyerspeer M, et al. Pros and cons of ultra-high-field MRI/MRS for human application. *Prog Nucl Magn Reson Spectrosc*. 2018; 109:1-50. doi:10.1016/j.pnmrs.2018.06.001
- de Martino F, Yacoub E, Kemper V, et al. The impact of ultra-high field MRI on cognitive and computational neuroimaging. *NeuroImage*. 2018;168: 366-382. doi:10.1016/j.neuroimage.2017.03.060
- Barisano G, Sepehrband F, Ma S, et al. Clinical 7 T MRI: Are we there yet? A review about magnetic resonance imaging at ultra-high field. *Br J Radiol*. 2019;92(1094):20180492. doi:10.1259/bjr.20180492
- Chen J, Bokacheva L, Rusinek H. Image segmentation and nonuniformity correction methods. In: *3D Printing for the Radiologist*. Elsevier; 2022:31-43.
- Despotović I, Goossens B, Philips W. MRI segmentation of the human brain: challenges, methods, and applications. *Comput Math Methods Med*. 2015; 2015:450341. doi:10.1155/2015/450341
- Vovk U, Pernus F, Likar B. A review of methods for correction of intensity inhomogeneity in MRI. *IEEE Trans Med Imaging*. 2007;26(3):405-421. doi: 10.1109/TMI.2006.891486
- Styner M, Brechbuhler C, Szckely G, Gerig G. Parametric estimate of intensity inhomogeneities applied to MRI. *IEEE Trans Med Imaging*. 2000;19(3): 153-165. doi:10.1109/42.845174
- Metzger GJ, Snyder C, Akgun C, Vaughan T, Ugurbil K, de Moortele PF. Local  $B_1^+$  shimming for prostate imaging with transceiver arrays at 7T based on subject-dependent transmit phase measurements. *Magn Reson Med*. 2008;59(2):396-409. doi:10.1002/mrm.21476
- Orzada S, Maderwald S, Poser BA, Bitz AK, Quick HH, Ladd ME. RF excitation using time interleaved acquisition of modes (TIAMO) to address  $B_1$  inhomogeneity in high-field MRI. *Magn Reson Med*. 2010;64(2):327-333. doi:10.1002/mrm.22527
- Cloos MA, Boulant N, Luong M, et al. kT-points: short three-dimensional tailored RF pulses for flip-angle homogenization over an extended volume. *Magn Reson Med*. 2012;67(1):72-80. doi:10.1002/MRM.22978
- Sled JG, Zijdenbos AP, Evans AC. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. *IEEE Trans Med Imaging*. 1998;17(1):87-97. doi:10.1109/42.668698
- Tustison NJ, Avants BB, Cook PA, et al. N4ITK: improved N3 bias correction. *IEEE Trans Med Imaging*. 2010;29(6):1310-1320. doi:10.1109/TMI.2010. 2046908
- Mikheev A, Bokacheva L, Yang H, et al. Retrospective MRI non-uniformity correction: quantitative assessment of two methods. Poster presented at: ISMRM & SMRT Virtual Conference & Exhibition; August 8–14, 2020; Abstract 3661.
- Juntu J, Sijbers J, van Dyck D, Gielen J. Bias field correction for MRI images. In: *Computer Recognition Systems*. Springer; 2005:543-551. doi:10.1007/ 3-540-32390-2\_64
- Ardizzone E, Pirrone R, Gambino O. Frequency determined homomorphic unsharp masking algorithm on knee MR images. In: Roli F, Vitulano S, eds. *Image Analysis and Processing – ICIAP 2005*. ICIAP 2005. Lecture Notes in Computer Science. Vol 3617. Springer, Berlin, Heidelberg; 2005. doi:10.1007/ 11553595\_113
- Arriaga-García EF, Sanchez-Yanez RE, Garcia-Hernandez MG. Image enhancement using bi-histogram equalization with adaptive sigmoid functions. In: *2014 International Conference on Electronics, Communications and Computers (CONIELECOMP)*. Cholula, Mexico; 2014:28-34. doi:10.1109/ CONIELECOMP.2014.6808563
- Venkatesh V, Sharma N, Singh M. Intensity inhomogeneity correction of MRI images using InhomoNet. *Comput Med Imaging Graph*. 2020;84:101748. doi:10.1016/j.compmedimag.2020.101748
- Dai X, Lei Y, Liu Y, et al. Intensity non-uniformity correction in MR imaging using residual cycle generative adversarial network. *Phys Med Biol*. 2020; 65(21):215025. doi:10.1088/1361-6560/abb31f
- Chuang K, Wu P, Li Z, Fan K, Weng J. Deep learning network for integrated coil inhomogeneity correction and brain extraction of mixed MRI data. *Sci Rep*. 2022;12:8578. doi:10.1038/s41598-022-12587-6
- Gaillochet M, Tezcan KC, Konukoglu E. Joint reconstruction and bias field correction for undersampled MR imaging. In: Martel AL et al., eds. *Medical Image Computing and Computer Assisted Intervention—MICCAI 2020*. Lecture Notes in Computer Science. Vol.12262. Springer; 2020:44-52. doi:10.1007/ 978-3-030-59713-9\_5
- Wan F, Smedby Ö, Wang C. Simultaneous MR knee image segmentation and bias field correction using deep learning and partial convolution. *Proc SPIE*. 2019;10949:1094909.
- Goldfryd T, Gordon S, Raviv TR. Deep semi-supervised bias field correction of Mr images. In: *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*. IEEE; 2021:1836-1840. doi:10.1109/ISBI48211.2021.9433889
- Simkó A, Löfstedt T, Garpebring A, Nyholm T, Jonsson J. MRI bias field correction with an implicitly trained CNN. *Proc Mach Learning Res*. 2022;172: 1125-1138. doi:10.5281/zenodo.3749526
- Nelamangala Sridhara S, Akrami H, Krishnamurthy V, Joshi A. Bias field correction in 3D-MRIs using convolutional autoencoders. *Proc SPIE*. 2021; 11596:115962H. doi:10.1117/12.2582042
- Glasser M, Sotiropoulos S, Wilson J, et al. The minimal preprocessing pipelines for the Human Connectome Project. *NeuroImage*. 2013;80:105-124. doi:10.1016/j.neuroimage.2013.04.127

26. Raaijmakers AJE, Italiaander M, Voogt IJ, et al. The fractionated dipole antenna: a new antenna for body imaging at 7 Tesla. *Magn Reson Med*. 2016; 75(3):1366-1374. doi:10.1002/mrm.25596
27. Arnold JB, Liow JS, Schaper KA, et al. Qualitative and quantitative evaluation of six algorithms for correcting intensity nonuniformity effects. *NeuroImage*. 2001;13(5):931-943. doi:10.1006/nimg.2001.0756
28. Hou Z. A review on MR image intensity inhomogeneity correction. *Int J Biomed Imaging*. 2006;2006(1):49515. doi:10.1155/IJBI/2006/49515
29. Haralick RM. Statistical and structural approaches to texture. *Proc IEEE*. 1979;67(5):786-804. doi:10.1109/PROC.1979.11328
30. Meliadó EF, van den Berg CAT, Luijten PR, Raaijmakers AJE. Intersubject specific absorption rate variability analysis through construction of 23 realistic body models for prostate imaging at 7T. *Magn Reson Med*. 2019;81(3):2106-2119. doi:10.1002/mrm.27518
31. Klein S, Staring M, Murphy K, Viergever MA, Pluim JPW. elastix: a toolbox for intensity-based medical image registration. *IEEE Trans Med Imaging*. 2010;29(1):196-205. doi:10.1109/TMI.2009.2035616
32. Shamonin D, Bron E, Lelieveldt B, Smits M, Klein S, Staring M. Fast parallel image registration on CPU and GPU for diagnostic classification of Alzheimer's disease. *Front Neuroinform*. 2014;7:50. doi:10.3389/fninf.2013.00050
33. Dilorio G, Brown JJ, Borrello JA, Perman WH, Shu HH. Large angle spin-echo imaging. *Magn Reson Imaging*. 1995;13(1):39-44. doi:10.1016/0730-725X(94)00082-E
34. Jordan CD, Saranathan M, Bangerter NK, Hargreaves BA, Gold GE. Musculoskeletal MRI at 3.0 T and 7.0 T: a comparison of relaxation times and image contrast. *Eur J Radiol*. 2013;82(5):734-739. doi:10.1016/j.ejrad.2011.09.021
35. Steensma BR, Luttje M, Voogt IJ, et al. Comparing signal-to-noise ratio for prostate imaging at 7T and 3T. *J Magn Reson Imaging*. 2019;49(5):1446-1455. doi:10.1002/jmri.26527
36. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. IEEE; 2015:770-778. doi:10.48550/arxiv.1512.03385
37. Johnson J, Alahi A, Fei-Fei L. Perceptual losses for real-time style transfer and super-resolution. In: *European Conference on Computer Vision*. Springer; 2016:694-711.
38. Kingma DP, Ba JL. Adam: a method for stochastic optimization. In: *Third International Conference on Learning Representations, ICLR 2015—Conference Track Proceedings*; 2014. doi:10.48550/arxiv.1412.6980
39. Pérez-García F, Sparks R, Ourselin S. TorchIO: a Python library for efficient loading, preprocessing, augmentation and patch-based sampling of medical images in deep learning. *Comput Methods Programs Biomed*. 2021;208:106236. doi:10.1016/J.CMPB.2021.106236
40. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. Image quality assessment: from error visibility to structural similarity. *IEEE Trans Image Process*. 2004; 13(4):600-612. doi:10.1109/TIP.2003.819861
41. Kantorovich LV. Mathematical methods of organizing and planning production. *Manage Sci*. 1960;6(4):366-422. doi:10.1287/mnsc.6.4.366
42. Hall-Beyer M. Practical guidelines for choosing GLCM textures to use in landscape classification tasks over a range of moderate spatial scales. *Int J Remote Sens*. 2017;38(5):1312-1338. doi:10.1080/01431161.2016.1278314
43. Steensma BR, Voogt IJ, Leiner T, et al. An 8-channel Tx/Rx dipole array combined with 16 Rx loops for high-resolution functional cardiac imaging at 7 T. *Magn Reson Mater Phys Biol Med*. 2018;31(1):7-18. doi:10.1007/S10334-017-0665-5/TABLES/4

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Harrevelt SD, Meliadó EFM, van Lier ALHMW, et al. Deep learning based correction of RF field induced inhomogeneities for T2w prostate imaging at 7 T. *NMR in Biomedicine*. 2023;36(12):e5019. doi:10.1002/nbm.5019