Received: 10 March 2023 | Revised: 18 July 2023 | Accepted: 15 August 2023

DOI: 10.1002/mrm.29851

RESEARCH ARTICLE

Magnetic Resonance in Medicine

# FD-Net: An unsupervised deep forward-distortion model for susceptibility artifact correction in EPI

**Abdallah Zaid Alkilani**[1,2] | **Tolga Çukur**[1,2,3] | **Emine Ulku Saritas**[1,2,3]

[1]Department of Electrical and Electronics Engineering, Bilkent University, Ankara, Turkey

[2]National Magnetic Resonance Research Center (UMRAM), Bilkent University, Ankara, Turkey

[3]Neuroscience Graduate Program, Bilkent University, Ankara, Turkey

**Correspondence**
Abdallah Zaid Alkilani, Department of Electrical and Electronics Engineering, Bilkent University, 06800 Ankara, Turkey.
Email: alkilani@ee.bilkent.edu.tr

**Funding information**
Türkiye Bilimsel ve Teknolojik Araştırma Kurumu, Grant/Award Number: 117E116

## Abstract

**Purpose:** To introduce an unsupervised deep-learning method for fast and effective correction of susceptibility artifacts in reversed phase-encode (PE) image pairs acquired with echo planar imaging (EPI).

**Methods:** Recent learning-based correction approaches in EPI estimate a displacement field, unwarp the reversed-PE image pair with the estimated field, and average the unwarped pair to yield a corrected image. Unsupervised learning in these unwarping-based methods is commonly attained via a similarity constraint between the unwarped images in reversed-PE directions, neglecting consistency to the acquired EPI images. This work introduces a novel unsupervised deep Forward-Distortion Network (FD-Net) that predicts both the susceptibility-induced displacement field and the underlying anatomically correct image. Unlike previous methods, FD-Net enforces the forward-distortions of the correct image in both PE directions to be consistent with the acquired reversed-PE image pair. FD-Net further leverages a multiresolution architecture to maintain high local and global performance.

**Results:** FD-Net performs competitively with a gold-standard reference method (TOPUP) in image quality, while enabling a leap in computational efficiency. Furthermore, FD-Net outperforms recent unwarping-based methods for unsupervised correction in terms of both image and field quality.

**Conclusion:** The unsupervised FD-Net method introduces a deep forward-distortion approach to enable fast, high-fidelity correction of susceptibility artifacts in EPI by maintaining consistency to measured data. Therefore, it holds great promise for improving the anatomical accuracy of EPI imaging.

**KEYWORDS**

deep learning, echo planar imaging, reversed phase-encoding, susceptibility artifacts, unsupervised learning

## 1 | INTRODUCTION

Echo planar imaging (EPI)[1] is the most commonly employed MRI sequence for diffusion-weighted imaging (DWI) and functional MRI (fMRI), due to its rapid k-space acquisition capability.[2,3] However, EPI is prone to susceptibility artifacts arising from $B_0$ field inhomogeneities, which are particularly prominent near tissue interfaces.[4] These artifacts manifest as intensity distortions from signal pileups/dropouts, and geometrical distortions due to compression/stretching of affected regions.[5] Severe artifacts can limit the clinical utility of EPI images. Therefore, artifact correction is an essential step to ensure accuracy of downstream qualitative and quantitative assessments, especially at high field strengths.[6-8]

A leading framework for susceptibility-artifact correction uses images acquired in reversed phase-encoding (PE) directions to estimate the susceptibility-induced displacement field directly from the resulting blip-up (BU) and blip-down (BD) EPI images.[5,9-11] An unwarping-based approach is commonly adopted for correction, where the reversed-PE images are nonlinearly transformed to alleviate artifacts based on the estimated displacement field. Either voxel-wise field estimates,[10,12,13] or weighted combination of basis spatial maps across the field-of-view (FOV)[9] can be used. Popular implementations of this framework include classical methods such as TOPUP from the FMRIB Software Library[9,14] and hyperelastic susceptibility correction of DTI data (HySCO) from the Statistical Parametric Mapping toolbox.[15,16] Since no additional data collection is needed beyond reversed-PE images, classical methods in the unwarping-based framework can offer notable benefits over measured-field-based, registration-based, or point spread function- (PSF) based approaches in the literature.[17,18] Nonetheless, these classical methods are based on iterative optimization techniques that introduce substantial computational burden, rendering them impractical under clinical settings.

Deep neural networks have recently been considered as a powerful alternative for artifact correction that can maintain high computational efficiency.[19] In the absence of ground-truth anatomically correct images, network training can be performed in a supervised fashion by using the corrected images generated by classical methods as reference. However, the lack of a physics-driven formulation in this approach can compromise generalization performance. Furthermore, the process of obtaining the corrected images can create extensive computational overhead for training, whereas the improvement gained in network performance may not scale with the computational overhead. Previous studies in this domain have addressed these problems via unsupervised learning strategies that aim to maximize the similarity of unwarped images across the two PE directions.[20,21] In this unwarping-based framework, reversed-PE images are first individually unwarped by the network, and then combined to produce a final estimate. Among such unsupervised learning-based methods, S-Net performs unwarping via bilinear interpolation and assesses the similarity between the corrected BU/BD images via a cross-modal loss.[20] Deepflow-Net instead performs unwarping via cubic interpolation and assesses the similarity between the corrected BU/BD images via a mean-squared error (MSE) loss.[21] While promising results have been reported, these previous methods define an unsupervised loss function in the output domain of unwarped images, for which no ground-truth data are available. Such lack of physical constraints in the loss function can cause suboptimal learning.[22,23] In turn, the network can produce low-fidelity images during inference, resulting in solutions that are notably inconsistent with the acquired reversed-PE images.[24]

Here, we propose a novel deep network model (FD-Net) based on a forward-distortion approach for correcting EPI susceptibility artifacts in reversed-PE image pairs. Unlike unwarping-based methods that average individually corrected reversed-PE images, FD-Net predicts a single anatomically corrected image along with a displacement field. Unlike previous deep-learning methods, FD-Net directly incorporates physical constraints in the input domain where measurements are available. Specifically, FD-Net forward-distorts the corrected image with the predicted field to reconstruct the reversed-PE image pair. Unsupervised learning is then achieved by enforcing consistency of the reconstructed versus acquired reversed-PE images. A multiresolution architecture is employed to maintain performance at both local and global scales. Comprehensive demonstrations are performed to assess the quality of corrected images and field estimates on EPI data from the Human Connectome Project (HCP) database.[25] FD-Net performs competitively with the reference TOPUP method, while enabling a leap in computational efficiency; and it significantly outperforms competing deep-learning methods based on the unwarping framework. These findings demonstrate the potential of FD-Net as a fast and effective method for susceptibility-artifact correction in EPI.

## 2 | THEORY

### 2.1 | Susceptibility-induced distortions

The relationship between the anatomically correct image and the distorted EPI image can be expressed as a linear system:

$$\underbrace{\mathbf{f}}_{n_{FE}n_{PE}\times 1} = \underbrace{\mathbf{K}}_{n_{FE}n_{PE}\times n_{FE}n_{PE}} \underbrace{\rho}_{n_{FE}n_{PE}\times 1}, \quad (1)$$

where $\mathbf{K}$ is a transformation matrix called the K-matrix, $\rho$ is the vectorized anatomically correct image, $\mathbf{f}$ is the vectorized EPI image, and $n_{PE}$ and $n_{FE}$ are the image dimensions in the PE and frequency encode (FE) directions, respectively. In general, $\mathbf{K}$ can be complex-valued given complex-valued images $\rho$ and $\mathbf{f}$, such that it performs a phase shift as well as interpolation.[9] In practice, however, magnitude images are more commonly utilized for convenience and $\mathbf{K}$ is real-valued. Ignoring the distortion along the FE-direction enables block diagonalization of the K-matrix, allowing the problem to be separated across FE lines as:

$$\underbrace{\mathbf{f}_i}_{n_{PE}\times 1} = \underbrace{\mathbf{K}_i}_{n_{PE}\times n_{PE}} \underbrace{\rho_i}_{n_{PE}\times 1}, \quad (2)$$

Here, $\mathbf{K}_i$, $i = 1, 2, \ldots, n_{FE}$, are the transformation submatrices acting along the PE-direction, and $\rho_i$ and $\mathbf{f}_i$ are the $i^{th}$ rows of the correct image and the EPI image, respectively. As shown in Figure 1, the K-matrix describes the mapping from the correct image to the EPI image. Deviations of the K-matrix from the identity matrix are representative of the amount of distortion, and multiple nonzero values on the same row indicate a many-to-one mapping (i.e., pileup/dropout distortions).

For reversed-PE acquisitions, Equation (2) can be written separately for the $i$th rows of the EPI images from BU/BD acquisitions. In that case, the associated K-matrices $\mathbf{K}_{i,BU}$ and $\mathbf{K}_{i,BD}$ are based on the same underlying field, with the difference of utilizing the negative of the field for BD acquisition.

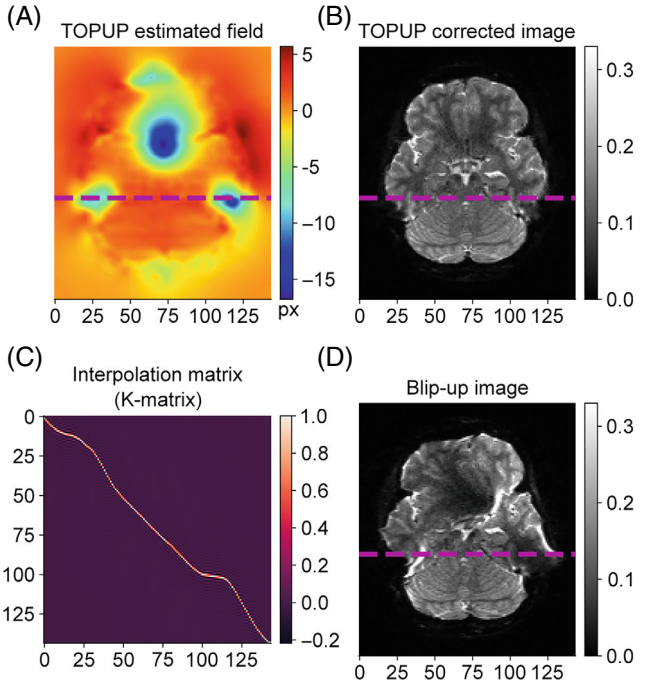


**FIGURE 1** Illustration of the image distortion characterized by the K-matrix. (A) The estimated displacement field (in units of pixels) and (B) the corrected image predicted by TOPUP are shown, with the magenta dashed lines highlighting a particular row along the PE direction (RL direction). (C) The K-matrix formed from the field for the highlighted row and (D) the corresponding blip-up EPI image. The deviations of the K-matrix from the identity matrix indicate the amount and direction of distortion, as can be understood by comparing the corrected image and the blip-up image for the highlighted row. The labeled axes correspond to the PE direction.

## 2.2 | Classical methods for distortion correction

Among classical methods for susceptibility-artifact correction in EPI, the predominant approach is correction based on reversed-PE acquisitions. TOPUP, a popular implementation of this approach, uses an alternating least-squares optimization to jointly solve the linear system of equations resulting from the reversed-PE acquisitions.[9,14] TOPUP first estimates the underlying field, which is taken as a compact linear combination of spatial basis functions across the image domain.[9] Next, transformation matrices that act on BU/BD acquisitions are generated based on the estimated field. Finally, to generate the anatomically correct image, unwarping is performed on BU/BD acquisitions by incorporating Jacobian modulation to compensate for intensity pileups. A main limitation of this method is that it relies on iterative optimization techniques that are computationally intensive.

Another classical method for correcting susceptibility artifacts is $B_0$ field map-based correction, which requires at least two additional acquisitions with different TE values for computing the field based on phase differences. This field is then used to correct the distorted EPI images by unwarping in image domain. However, erroneous field maps can elicit residual artifacts after correction, and phase unwrapping during field computation is prone to failure especially in regions with high $B_0$ inhomogeneities, such as air/tissue and bone/soft tissue interfaces.[26] Yet another classical method is registration-based correction, which requires an additional anatomical reference image to perform registration with the use of a cross-modal loss function.[24] A distortion-free $T_1$- or $T_2$-weighted image typically serves as an anatomical template for the EPI image in the presence of large distortions. Additional constraints are often incorporated to improve solutions, including diffusion tensor[27] and

fiber orientation distributions,[28] alignment of cortical surfaces[29] and synthesized anatomical images.[30] Popular implementations of the aforementioned methods provided in FMRIB Software Library are FUGUE and FLIRT, which perform $B_0$ field map-based correction and image registration-based correction, respectively.[31,32] However, in addition to requiring auxiliary scans, these approaches fall short at capturing more intricate distortions or compensating for signal intensity variations.[33] Alternatively, methods based on PSF measurements have been proposed for analytical correction based on regularized deconvolution,[18,34] where learning-based deconvolution methods can also be adopted to improve performance.[35,36] While PSF-based methods can correct a broad range of distortions in EPI images, they require voxel-wise PSF measurements via prolonged scans that must be repeated under notable changes in k-space trajectories.[37]

## 2.3 | Learning-based methods for distortion correction

In recent years, learning-based approaches have been adopted as a promising alternative for correction of susceptibility artifacts in EPI. A first group of methods have aimed to improve performance of classical methods via complementary data processing. Synthesis methods are applicable in cases where reversed-PE data are not available.[38,39] After an undistorted EPI image is synthesized given as input a structural MR image, synthesized and acquired EPI images are processed via TOPUP to unwarp the acquired image.[38,39] While suited for clinical data acquired under time limitations, synthesis methods can yield images with reduced resolution when compared to those based on reversed-PE acquisitions. Fiber-orientation distribution methods use latent features of fiber-orientation distribution images extracted from DWI data to further improve TOPUP-based correction of reversed-PE images.[40] Fiber-orientation distribution methods incorporate additional anatomical information to improve performance in problematic regions such as the brainstem. However, they still rely on the relatively slow TOPUP correction. Learning-based correction with multishot EPI sequences has also been considered to help minimize the distortions in acquired images. Low-rank reconstructions of a multishot EPI sequence based on simultaneous multislab acquisition have been proposed for DWI.[41] Self-supervised denoising of a multicontrast multishot EPI sequence based on reversed-PE acquisitions has been proposed for $T_2$, $T_2^*$, and susceptibility mapping.[42] Physics-driven reconstruction of an echo-shifting acquisition has been proposed for relaxometry along with $B_0$ and $B_1$ mapping.[43] Note that these methods involve advanced pulse sequence modifications that may not be available at all sites, and often leverage TOPUP for estimation of field maps.

A second group of methods have instead aimed to improve computational efficiency over classical correction methods. A common framework in this domain relies on field estimation followed by unwarping of EPI images. Earlier studies have considered supervised methods that train network models for correction assuming availability of ground truth for undistorted EPI images.[44-46] These ground truth images are typically obtained via simulations or from classical correction methods. Some supervised methods further cast estimation of the displacement field from a reversed-PE image pair as an optical flow estimation problem, and later use the estimated field for correction.[47,48] Although supervised methods benefit from the data-driven learning capabilities of network models, reliance on the availability of undistorted EPI images limits their utility in many applications where such ground truth is not available.

This has sparked interest in unsupervised methods that can learn to correct artifacts in the absence of ground truth. As in the case of classical methods, the predominant approach for unsupervised correction relies on reversed-PE acquisitions. Based on the assumption that displacements in non-PE directions are negligible,[10] the displacement field is estimated so as to maximize the similarity of unwarped images obtained by reverse distortion on the acquired PE image pair. The recently proposed S-Net[20] utilizes a three-dimensional U-Net model[19] to predict the field, followed by unwarping using bilinear interpolation inspired by the deformable image registration method VoxelMorph.[49] For unsupervised learning, S-Net uses a similarity loss taken as the local cross-correlation (LCC) between corrected BU/BD images, along with a diffusion regularizer to enforce field smoothness. Another recent method named Deepflow-Net[21] uses a two-dimensional (2D) U-Net model where field estimates are produced at multiple resolutions by extracting features from various stages of the decoder.[47,48] Deepflow-Net performs correction via cubic interpolation and adopts a density compensation similar to TOPUP[9] to handle pileups. For unsupervised learning, Deepflow-Net uses as similarity loss the MSE between the corrected BU/BD images, along with a total variation regularizer to enforce field smoothness. While these seminal methods have produced promising results, they enable unsupervised learning by assessing similarity of unwarped images in opposing PE directions. This indirect approach omits physics-driven constraints regarding the actual EPI measurements. Thus, performance of the learned correction can degrade under relatively large distortions and near tissue boundaries.

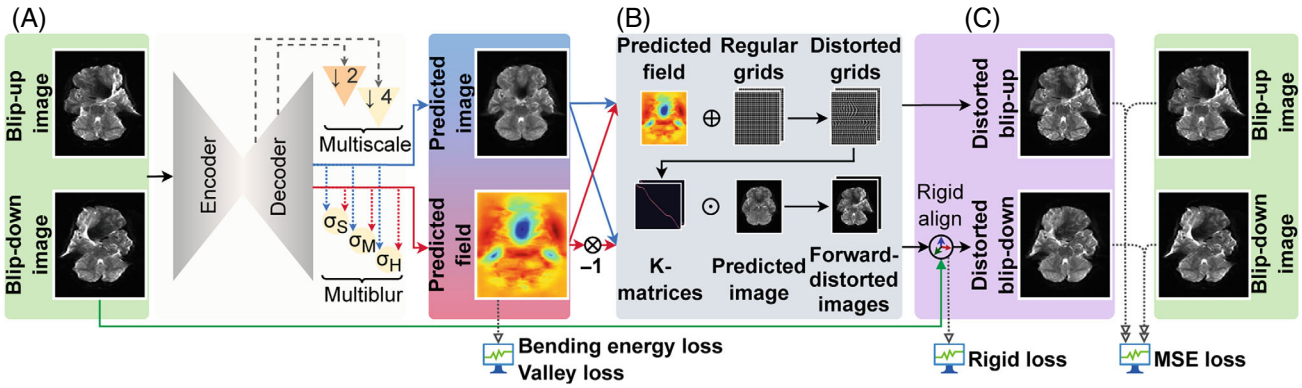


**FIGURE 2** Overview of the proposed FD-Net. (A) The input distorted blip-up/blip-down images are fed through an encoder-decoder in the prediction unit, which outputs a predicted image and a predicted field with optional multiresolution (multiscale and/or multiblur) schemes. The field is used to formulate the bending energy loss and valley loss. (B) The K-Unit applies forward-distortion in each PE direction, with the field negated for one of the directions. (C) A rigid alignment unit is included to improve registration, with the rigid loss formulated from the transformation parameters. The forward-distorted images are compared with the input images (redisplayed here for convenience) to formulate the mean-squared error loss. Training is performed over the aggregate of the shown losses.

Here, we propose a novel unsupervised deep-learning method for artifact correction in EPI to improve performance. Unlike previous unsupervised methods, the proposed FD-Net method directly constrains fidelity to the actual EPI measurements. This constraint is introduced by integrating the forward physical model of EPI distortions observed on measured images, so FD-Net benefits from the enhanced reliability of physics-driven deep learning.

## 2.4 | Proposed FD-Net

FD-Net is a novel unsupervised forward-distortion model that explicitly enforces measurement fidelity for enhanced correction performance, as outlined in Figure 2. The prediction unit, shown in Figure 2A, uses a 2D U-Net to produce both a predicted field and a predicted anatomically correct image from the input reversed-PE images. In contrast to unwarping-based methods that produce separate unwarped images for BU/BD acquisitions, predicting a single correct image can offer signal to noise ratio (SNR) benefits analogously to the sensitivity-encoding approaches in parallel imaging.[50] Code to implement FD-Net is available at: https://github.com/saritas-lab/FD-Net.

The K-Unit in FD-Net, illustrated in Figure 2B, forward-distorts the predicted anatomically-correct image using the predicted field to reconstruct the input PE images. The BU acquisition is reconstructed using the estimated field, whereas the BD acquisition is reconstructed using the negative of the estimated field. Distortions are efficiently emulated using the K-Unit that embodies a simple matrix multiplication with a separable formulation as in Equation (2). Afterwards, fidelity between reconstructed and measured data is enforced using a multiresolution scheme.

The rigid alignment unit in Figure 2C allows compensation for small movements between the input PE image in one direction (BD acquisition in this case) and its corresponding forward-distorted image. This allows the network to focus on displacements that are due to off-resonance via the field-based formulation of the K-Unit.

### 2.4.1 | Forward-distortion with K-unit

The K-Unit in FD-Net performs forward-distortion on the estimated anatomically correct image using the estimated field, as illustrated in Figure 3. The steps described below are given for the BU direction for brevity, but they are similarly conducted for the BD direction, with the difference of utilizing the negative of the displacement field. First, a uniform spatial grid $\mathbf{X}_{\text{grid}}$ is formed:

$$\underbrace{\mathbf{X}_{\text{grid}}}_{n_{\text{PE}}\times n_{\text{PE}}} = \begin{bmatrix} 1 & \cdots & 1 \\ 2 & \cdots & 2 \\ \vdots & & \vdots \\ n_{\text{PE}} & \cdots & n_{\text{PE}} \end{bmatrix}. \tag{3}$$

The distorted grid after interpolation, $\mathbf{X}_{i,\text{BU}}$, is formed by determining the new grid location for each pixel from the shift amount given in the displacement field, that is,

$$\underbrace{\mathbf{X}_{i,\text{BU}}}_{n_{\text{PE}}\times n_{\text{PE}}} = \begin{bmatrix} \mathbf{O}_{\text{field}}(i,1)+1 & \cdots & \mathbf{O}_{\text{field}}(i,n_{\text{PE}})+n_{\text{PE}} \\ \mathbf{O}_{\text{field}}(i,1)+1 & \cdots & \mathbf{O}_{\text{field}}(i,n_{\text{PE}})+n_{\text{PE}} \\ \vdots & & \vdots \\ \mathbf{O}_{\text{field}}(i,1)+1 & \cdots & \mathbf{O}_{\text{field}}(i,n_{\text{PE}})+n_{\text{PE}} \end{bmatrix}, \tag{4}$$

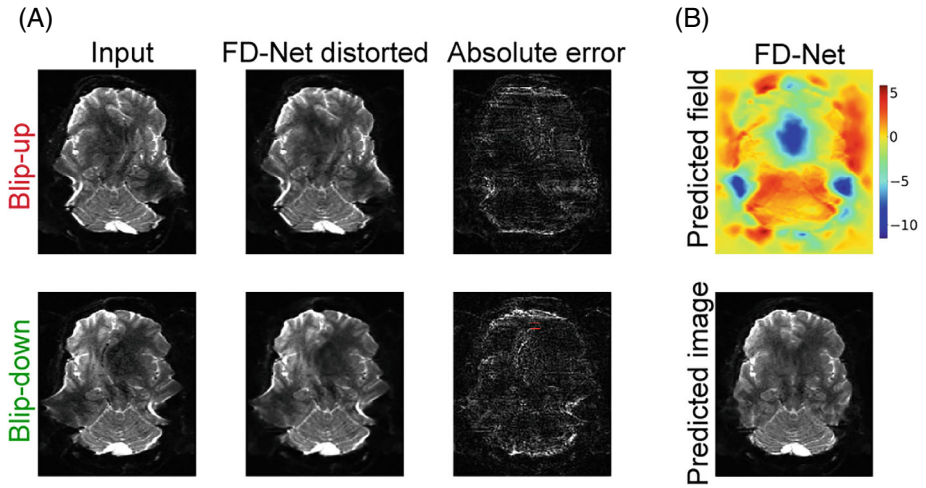


**FIGURE 3** Example of forward-distortion by using the K-Unit in FD-Net. (A) The input blip-up and blip-down echo planar imaging images are compared with the forward-distortion results of FD-Net. The intensities in absolute error maps are scaled up 2.5× for improved visualization. (B) The predicted field and predicted image outputs from FD-Net, which are input to the K-Unit to obtain the forward-distorted images in (A).

where $\mathbf{O}_{\text{field}}$ is the estimated field output of FD-Net in units of pixels and $i = 1, 2, \ldots, n_{\text{FE}}$ is the row index over the FE direction. For practical purposes, each entry in $\mathbf{X}_{i,\text{BU}}$ is kept limited between 1 and $n_{\text{PE}}$ (i.e., clipped to the valid range of interpolation). Taking the difference between the two grids and then applying an interpolation kernel, $\kappa(\xi)$, gives us the K-matrix that will act on the $i^{th}$ row as follows:

$$\underbrace{\mathbf{K}_{i,\text{BU}}}_{n_{\text{PE}}\times n_{\text{PE}}} = \kappa\left(\mathbf{X}_{i,\text{BU}} - \mathbf{X}_{\text{grid}}\right). \tag{5}$$

Using this K-matrix, the $i$th row of the forward-distorted image is reconstructed via a matrix multiplication:

$$\underbrace{\left[\mathbf{O}^T_{\text{dist,BU}}\right]_i}_{n_{\text{PE}}\times 1} = \mathbf{K}_{i,\text{BU}} \underbrace{\left[\mathbf{O}^T_{\text{image}}\right]_i}_{n_{\text{PE}}\times 1}, \tag{6}$$

where $(\cdot)^T$ denotes matrix transpose, $[\cdot]_i$ denotes the $i$th column of a matrix, and $\mathbf{O}_{\text{image}}$ is the predicted anatomically correct image. Finally, the forward-distorted image $\mathbf{O}_{\text{dist,BU}}$ can be formed by stacking the individually distorted rows:

$$\underbrace{\mathbf{O}_{\text{dist,BU}}}_{n_{FE}\times n_{PE}} = \left[\left[\mathbf{O}^T{}_{\text{dist,BU}}\right]_1 \middle| \left[\mathbf{O}^T{}_{\text{dist,BU}}\right]_2 \middle| \cdots \middle| \left[\mathbf{O}^T{}_{\text{dist,BU}}\right]_{n_{FE}}\right]^T. \tag{7}$$

Note that multiplication with K-matrix rows performs an interpolation across pixel neighborhoods with intensity modulations, so it can emulate signal pileups/dropouts.

### 2.4.2 | Network architecture

The architecture of FD-Net is detailed in Figure S1. As depicted in Figure S1A, the encoder in the prediction unit projects input reversed-PE images onto a latent representation across multiple stages. The receptive field is progressively refined by decreasing kernel size and using convolution with stride 2 for downsampling. The decoder then resolves the predicted field and predicted image from the latent representation through multiple stages of convolutional filtering and upsampling. Feature maps from the encoder stages are projected onto the decoder through skip connections to improve information flow.

A rigid-body motion may occur between the BU and BD acquisitions. As shown in Figure S1B, the rigid alignment unit in FD-Net applies motion-related transformations on one of the forward-distorted images only (BD distorted image in this case). This unit receives as input the measured BD acquisition along with the respective forward-distorted image, and uses convolutional and densely connected layers to predict the motion parameters $s_x$, $s_y$, and $r$, which capture the x-axis shift, y-axis shift, and in-plane rotation, respectively. These parameters are then used to apply a rigid transformation to the BD distorted image to improve its alignment with the corresponding BD acquisition. Note that a similar rigid alignment is also performed in TOPUP, and it offloads some burden from the nonrigid field-based alignment by accounting for subject movement between the two reversed-PE acquisitions.

As illustrated in Figure S2, FD-Net adopts a multiresolution scheme to improve performance by enforcing consistency across different spatial resolutions, in principle leading to faster convergence and more reliable performance. The multiresolution idea can be applied across spatial scales, spatial blurs, or both. In FD-Net, we refer to the multiresolution scheme applied at different spatial scales as multiscale and at different spatial blurs as multiblur. For multiscale, field and anatomically correct image estimates are produced at multiple spatial resolutions by extracting outputs from different stages of the decoder. For multiblur, the full resolution outputs are blurred with

Gaussian kernels at varying SDs. In both cases, the estimates obtained at multiple scales/blurs are processed with the K-Unit after proper scaling of their contribution to the overall loss function.

### 2.4.3 | Network loss

The overall loss function for FD-Net is given as:

$$\mathcal{L}_{\text{FD-Net}} = \sum_m \omega_m \left[ \mathcal{L}^{(m)}_{\text{MSE}} + \lambda_m \left( \mathcal{L}^{(m)}_{\text{BE}} + 10^3 \mathcal{L}^{(m)}_{\text{valley}} \right) \right] + \gamma \mathcal{L}_{\text{rigid}}, \tag{8}$$

where $m$ indicates the multiresolution step (full resolution by default; optionally includes multiscale and/or multiblur levels at progressively reduced resolutions), $\omega_m$ and $\lambda_m$ are the weighting and regularization parameter over the smoothness of the field for step $m$, superscript $(m)$ denotes the version of a parameter at step $m$, and $\gamma$ is the weight of the rigid alignment loss. Here, the first term denotes the sum of reconstruction losses, while the second term denotes the rigid loss, described in detail below.

First, $\mathcal{L}^{(m)}_{\text{MSE}}$ is MSE between the measured and forward-distorted images averaged across the two PE directions at the $m$th step:

$$\mathcal{L}^{(m)}_{\text{MSE}} = \frac{1}{2n^{(m)}_{\text{PE}} n^{(m)}_{\text{FE}}} \left[ \sum_{\mathbf{p}\in\Omega} \left( \mathbf{O}^{(m)}_{\text{dist,BU}}(\mathbf{p}) - \mathbf{I}^{(m)}_{\text{im,BU}}(\mathbf{p}) \right)^2 + \sum_{\mathbf{p}\in\Omega} \left( \mathbf{O}^{(m)}_{\text{dist,BD}}(\mathbf{p}) - \mathbf{I}^{(m)}_{\text{im,BD}}(\mathbf{p}) \right)^2 \right], \tag{9}$$

where $\mathbf{I}_{\text{im,BU}}$ and $\mathbf{I}_{\text{im,BD}}$ are the input EPI images for BU and BD acquisitions, respectively. For the multiscale scheme, these images are downsampled properly to avoid aliasing artifacts.

Next, $\mathcal{L}^{(m)}_{\text{BE}}$ is the bending energy regularizer[51] over the field at each step $m$ expressed as:

$$\mathcal{L}^{(m)}_{\text{BE}} = \sum_{\mathbf{p}\in\Omega} \left( \frac{\partial^2}{\partial x^2} \mathbf{O}^{(m)}_{\text{field}}(\mathbf{p}) \right)^2 + \left( \frac{\partial^2}{\partial y^2} \mathbf{O}^{(m)}_{\text{field}}(\mathbf{p}) \right)^2 + \left( \frac{\partial^2}{\partial xy} \mathbf{O}^{(m)}_{\text{field}}(\mathbf{p}) \right)^2 + \left( \frac{\partial^2}{\partial yx} \mathbf{O}^{(m)}_{\text{field}}(\mathbf{p}) \right)^2. \tag{10}$$

In practice, first- and second-order finite differences are used to approximate the gradients.[52]

$\mathcal{L}^{(m)}_{\text{valley}}$ is the valley loss for the field to prevent the overall loss function from exploding in earlier training iterations,[21] and is given as:

$$\mathcal{L}^{(m)}_{\text{valley}} = \sum_{\mathbf{p}\in\Omega} \max\left( \left| \mathbf{O}^{(m)}_{\text{field}}(\mathbf{p}) \right| - \tau_m, 0 \right), \tag{11}$$

where $\tau_m$ is a chosen threshold of maximum permissible field swing in units of pixels. $\mathcal{L}^{(m)}_{\text{valley}}$ sums the excess amount of field swing values when their magnitudes exceed $\tau_m$. These cases are penalized heavily by weighting $\mathcal{L}^{(m)}_{\text{valley}}$ with a large constant in Equation (8). In later stages of training, the effect of $\mathcal{L}^{(m)}_{\text{valley}}$ is negligible once the network converges towards reasonable solutions.

Finally, $\mathcal{L}_{\text{rigid}}$ is the rigid loss to find the smallest possible rigid transformation parameters for the alignment of measured and forward-distorted BD images, and is defined as follows:

$$\mathcal{L}_{\text{rigid}} = s_x^2 + s_y^2 + r^2. \tag{12}$$

Because the same rigid alignment applies to all multiresolution steps, a single rigid loss term is included in Equation (8).

## 3 | METHODS

### 3.1 | Experimental dataset and setup

#### 3.1.1 | Experiments on DWI dataset

For the main experiments in this work, unprocessed DWI data from HCP 1200 Subjects Data Release were used.[25] The images were acquired on a 3T MRI scanner (Siemens Skyra "Connectom"), using a multiband diffusion sequence with ss-EPI readouts in right-to-left (RL) and left-to-right (LR) reversed-PE polarities.[53] Other imaging parameters included: $210 \times 180\,\text{mm}^2$ FOV, 1.25 mm isotropic resolution, averages = 1, multiband acceleration factor 3; pulse repetition time/echo time = 5520/89.50 ms, flip angle = 78°, $168 \times 144$ acquisition matrix, bandwidth = 1488 Hz/Px, EPI factor = 144, echo spacing = 0.78 ms, and 6/8 partial Fourier acquisition.

A total of 24 subjects were selected randomly from the DWI dataset, with 12 reserved for training, four for validation and eight for testing. For each subject, a single b0-volume consisting of 111 axial slices with $168 \times 144$ image matrix was utilized. To obtain reference corrected images, the TOPUP method was applied on the data following the recommended guidelines by the toolbox.

To test the effect of training dataset size, the networks were trained separately with four subjects selected randomly from the original 12 subjects reserved for training. In addition, the networks were also trained separately with 42 subjects, by adding 30 new subjects to the training data. The validation and testing subjects were kept the same in all cases.

All networks were implemented in Keras with Tensorflow backend, on a machine with NVIDIA RTX 3070 GPU. Training was performed with the Adam optimizer

for a learning rate of $10^{-4}$ and a maximum of 1000 epochs, with early stopping when the change in the validation loss between consecutive epochs in the validation set fell below a threshold of $10^{-6}$.

#### 3.1.2 | Experiments on fMRI dataset

To test out-of-domain generalization, networks trained with EPI images from the DWI dataset were tested on EPI images from an fMRI dataset featuring different scan parameters. For this purpose, unprocessed task-evoked fMRI data from HCP 1200 Subjects Data Release was used,[25] collected for the task of emotion processing.[54] The images were acquired on a 3T MRI scanner (Siemens Skyra "Connectom"), using a BOLD sequence with a single-band spin-echo EPI readout in RL/LR reversed-PE polarities.[53] Other imaging parameters included: $208 \times 180$ mm$^2$ FOV, 2.00 mm isotropic resolution, averages = 1, pulse repetition time/echo time = 7060/58.00 ms, flip angle = 52°, $104 \times 90$ acquisition matrix, bandwidth = 2290 Hz/Px, EPI factor = 90, echo spacing = 0.58 ms, and 6/8 phase partial Fourier acquisition. For each subject, a single time frame consisting of 72 axial slices with $104 \times 90$ image matrix was utilized. Images were resampled using spline interpolation of order 3 to the size accepted by the pretrained networks (i.e., $168 \times 144$, the size of the DWI data).

For testing, evaluations were performed on the fMRI data corresponding to the same eight subjects assigned for testing in the DWI case. For fine-tuning, four additional training subjects were selected randomly from the fMRI dataset (nonoverlapping with the subjects chosen from the DWI dataset). The reference corrected images were obtained using TOPUP. The implementation and training procedures were the same as in the DWI case, with the difference that a maximum of 32 epochs was used with early stopping conditioned on the fine-tuning training loss.

### 3.2 | FD-Net implementation

The columns of the K-matrix in the K-Unit were generated using a sinc kernel, that is, $\kappa(\xi) = \mathrm{sinc}(\xi)$. All convolutional layers in the encoder-decoder (i.e., U-Net) utilized Leaky Rectified Linear Unit (ReLU) activation with a slope coefficient $\alpha = 0.2$, except at the final steps of the decoder as indicated in Figure S1A; the predicted image was output via a convolutional layer with ReLU activation and the predicted field was output via a convolutional layer with linear activation. For the multiscale case, convolutional layers akin to the full resolution case were employed to form the predicted field and image at 1/2 and 1/4 of the full scale. For each level of the multiblur case, the full resolution output was blurred with Gaussian kernels of standard deviation $\sigma$ and width $\lceil 4\sigma \rceil$. Three different blur levels were used: small (S), medium (M), and high (H) blurs of $\sigma_S = 0.5$, $\sigma_M = 1.5$, and $\sigma_H = 2.5$, respectively.

### 3.3 | Competing methods

Two unsupervised learning-based methods, S-Net and Deepflow-Net, were implemented for comparison. In addition, a supervised method was implemented to serve as a baseline for FD-Net. Implementations of competing methods were maintained as consistent to FD-Net as possible to facilitate fair comparisons:

1. *S-Net*: S-Net was implemented using a 2D U-Net. Only the field head at the end of the decoder in Figure S1A was necessary and correction was performed using a modified K-Unit approach as follows:

$$\underbrace{\left[\mathbf{O}^T_{\text{unwarp,BU}}\right]_i}_{n_{\text{PE}}\times 1} = \mathbf{K}^T_{i,\text{BU}} \underbrace{\left[\mathbf{I}^T_{\text{im,BU}}\right]_i}_{n_{\text{PE}}\times 1}, \tag{13}$$

   Here, $\mathbf{O}_{\text{unwarp,BU}}$ denotes the unwarped BU image. Note that no density compensation was incorporated by Duong et al.[20] Similarly, by transposing $\mathbf{K}_{i,\text{BU}}$, a standard unwarping interpolation was performed without density compensation. The BD acquisition was similarly treated, with the K-matrix formed after negation of the field. The average of the unwarped BU/BD images was taken as the corrected image. For training, LCC of the unwarped BU/BD images was utilized for similarity loss.[20,49] In place of the diffusion regularizer in Reference 20, bending energy from Equation (10) was used to facilitate comparison with FD-Net. In addition, the rigid alignment unit was utilized and the rigid loss from Equation (12) was incorporated.
2. *Deepflow-Net*: Deepflow-Net was implemented using a 2D U-Net. Only the field head at the end of the decoder in Figure S1A was needed and density-compensated correction was performed based on a modified K-Unit approach. The K-matrix for the BU acquisition was multiplied with a $n_{\text{PE}} \times 1$ vectorized image consisting of 1's, $\mathbf{1}$, to produce a density pileup map $\mathbf{W}_{\text{BU}}$. This map was inverted and used to weight the input PE image to enable density compensation akin to Zahneisen et al.[21]:

$$\underbrace{\left[\mathbf{O}^T_{\text{unwarp,BU}}\right]_i}_{n_{\text{PE}}\times 1} = \mathbf{K}^T_{i,\text{BU}} \underbrace{\left((\mathbf{1} \oslash \mathbf{W}_{\text{BU}}) \odot \left[\mathbf{I}^T_{\text{im,BU}}\right]_i\right)}_{n_{\text{PE}}\times 1}, \tag{14}$$

where

$$\underbrace{\mathbf{W}_{\mathrm{BU}}}_{n_{\mathrm{PE}}\times 1} = \mathbf{K}_{i,\mathrm{BU}} \underbrace{\mathbf{1}}_{n_{\mathrm{PE}}\times 1}. \tag{15}$$

Here, $\oslash$ and $\odot$ denote Hadamard division and product, respectively, and $(\mathbf{1} \oslash \mathbf{W}_{\mathrm{BU}})$ is limited in $[0, 1]$ to decrease the intensity in pileup regions.[21] The same process was also followed for the BD acquisition, with the K-matrix formed after negation of the field. In contrast to Equation (13), Equation (14) applies density compensation together with unwarping. The average of the two unwarped images was used as the corrected image. The same multiscale strategy as in FD-Net was adopted. MSE between the unwarped BU/BD images was used as the similarity loss. In place of total variation regularization in Zahneisen et al.,[21] bending energy loss was applied for the field as in Equation (10). The rigid alignment unit was also incorporated along with its loss term.

3 *Supervised baseline*: Finally, a supervised baseline was trained with an architecture identical to that of FD-Net, with the exception of the loss being fully supervised. For this purpose, MSE between the network predicted field/image and the results from TOPUP were utilized.

### 3.4 | Quantitative assessments

The qualities of the predicted image and field were assessed via Peak SNR (PSNR) and Structural Similarity Index Measure (SSIM) metrics, with the TOPUP results taken as reference. Before computing PSNR and SSIM, the field generated by each method was masked via a median Otsu threshold over the TOPUP image to remove background regions from consideration.[55]

For all methods, hyperparameters were chosen empirically to maximize PSNR and SSIM over the four subjects reserved as validation data. The selected hyperparameters are provided in Table 1. Performance assessments were reported on independent test data. Statistical evaluations were performed using paired Wilcoxon signed-rank test, with significance based on $p < 0.05$.

## 4 | RESULTS

### 4.1 | Computation time

All competing methods provided substantial computational advantage over TOPUP. While TOPUP took on average approximately 3086 s (∼ 51.5 min) to predict the field and an additional approximately 6 s to correct per volume, network-based methods merely took approximately 7.5 s. For performing distortion correction, this entails approximately 400× higher efficiency for a single volume, and approximately 5× higher efficiency for multivolume data such as multishell DWI with approximately 100 directions. Thus, network-based artifact correction enables significant speed up over classical methods. Furthermore, the long processing times of TOPUP also reflect poorly on supervised network models that are trained with TOPUP-corrected images as reference. Training the network models considered here took on average approximately 75 min per subject. As such, the unsupervised FD-Net offers approximately 2× higher efficiency in model training compared to the supervised baseline.

### 4.2 | Ablation studies for FD-Net

The choice of multiresolution strategy for FD-Net was first considered, followed by an ablation study on the combination of multiresolution components. The parameters were chosen empirically, with the purpose of maximizing quantitative image quality metrics with respect to TOPUP over the predicted field/image. Lastly, an ablation study was conducted to evaluate the contribution of each loss term in Equation (8).

**TABLE 1** Hyperparameter choices for the proposed FD-Net and the competing methods.

| | | $\omega$ | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Methods** | $\lambda$ | **FR** | **1/2** | **1/4** | **S** | **M** | **H** | $\gamma$ | $\tau$ |
| Proposed FD-Net | $10^{-5}$ | 0.4 | — | — | 0.3 | 0.2 | 0.1 | 0.01 | 32 |
| Deepflow-Net | $10^{-5}$ | 0.6 | 0.3 | 0.1 | — | — | — | 0.01 | 32 |
| S-Net | 10 | 1.0 | — | — | — | — | — | 0.01 | — |
| Supervised baseline | — | 1.0 | — | — | — | — | — | — | — |

*Notes*: For each method, irrelevant hyperparameters are marked with a dash (—). The hyperparameters considered are: $\lambda$ for field smoothness regularization, $\omega$ for multiresolution weighting parameter, $\gamma$ for rigid loss, and $\tau$ for valley loss threshold. $\omega$ is split into its constituent full resolution ("FR"), multiscale (1/2 and 1/4 scale), and multiblur (S, M, and H) components.

**TABLE 2** Performance comparison results for the loss ablation study for FD-Net.

| | Image quality | | Field quality | |
|---|---|---|---|---|
| **Loss terms** | **PSNR (dB)** | **SSIM (%)** | **PSNR (dB)** | **SSIM (%)** |
| $\mathcal{L}_{\text{FD-Net}}$ | 31.29 (2.79) | 86.71 (11.58) | 22.48 (5.69) | 83.09 (10.37) |
| $\mathcal{L}_{\text{FD-Net}} \setminus \{\mathcal{L}_{\text{rigid}}\}$ | 31.37 (2.93) | 86.65 (11.68) | 22.21 (5.86) | 83.02 (10.37) |
| $\mathcal{L}_{\text{FD-Net}} \setminus \{\mathcal{L}_{\text{valley}}\}$ | 31.23 (2.79) | 86.49 (11.63) | 22.36 (5.77) | 83.10 (10.27) |
| $\mathcal{L}_{\text{FD-Net}} \setminus \{\mathcal{L}_{\text{BE}}\}$ | 30.92 (2.72) | 85.51 (11.67) | 21.46 (5.82) | 81.47 (11.19) |
| $\mathcal{L}_{\text{MSE}}$ | 31.02 (2.84) | 85.50 (11.70) | 21.37 (5.92) | 81.35 (11.29) |

Notes: Peask signal-to-noise ratio (PSNR) and structural similarity index measure (SSIM) metrics are reported as mean (SD) across subjects. Removal of a loss component is indicated by "\" symbol followed by the removed loss term in curly braces. The full version of the loss is chosen for FD-Net as it provides the best overall performance.

#### 4.2.1 | Multiresolution ablation study for FD-Net

FD-Net was trained and subsequently evaluated for each multiresolution strategy, alongside a strategy with no multiresolution. The performances of the multiscale and multiblur schemes, as well their combination, were compared to determine the best multiresolution strategy. The hyperparameters chosen for each multiresolution scheme considered are provided in Table S1. PSNR and SSIM metrics are listed in Table S2. Overall, introducing a multiblur strategy provides a performance boost. Using the multiblur strategy, the image quality is improved by 0.67 dB PSNR/2.11% SSIM, and the field quality is improved by 1.68 dB PSNR/2.94% SSIM over the no multiresolution case. In contrast, the multiscale strategy underperforms in comparison to both the multiblur and the no multiresolution cases. A combination of multiblur and multiscale strategies does not improve over the multiblur case either, indicating that multiblur alone is sufficient to boost performance. Hence, the multiblur strategy was selected for FD-Net.

Next, combinations of three different blur amounts were considered for the multiblur case, with hyperparameters as listed in Table S3. The results in Table S4 show that including two or more blur levels boosts performance. Incorporating all blur levels (i.e., S-M-H multiblur combination) provides the best results, improving the image quality by 2.28 dB PSNR/5.23% SSIM and the field quality by 5.99 dB PSNR/8.80% SSIM over the worst performing M multiblur scheme. Hence, the S-M-H multiblur combination was chosen for FD-Net, as it provides a reliable generalization by incorporating all blur levels.

#### 4.2.2 | Loss ablation study for FD-Net

An ablation study was conducted by removing one loss term at a time from Equation (8) to investigate its contribution to the overall performance. Additionally, a version using only the MSE loss term (i.e., $\mathcal{L}_{\text{MSE}}$) was provided for reference. The results provided in Table 2 indicate that the proposed FD-Net provides the best overall performance. Removal of the rigid loss slightly decreases the predicted field quality, while removal of the valley loss slightly decreases the predicted image quality. Removal of the bending energy loss has the most detrimental effect on performance, leading to a significant drop in PSNR and SSIM down to the level of the MSE-only case. The proposed FD-Net improves the image quality by 0.27 dB PSNR/1.11% SSIM and the field quality by 1.21dB PSNR/1.74% SSIM over the MSE-only case.

### 4.3 | Comparison with Competing Methods

Comprehensive quantitative evaluations and visual assessments of the proposed FD-Net and the competing methods were conducted with respect to the reference TOPUP results.

*Slice-wise evaluations*: Figure 4 demonstrates the performance of each method across different slices of the dataset. Since the dataset captured the same anatomy at the same orientation for all subjects, a given slice number corresponds to approximately the same anatomical location in all subjects. Hence, no additional intersubject registration was conducted for this analysis. The underlying anatomy is illustrated in Figure 4A for a particular subject, where the $T_1$ weighted volume was registered to the corresponding b0 volume for display purposes, using FMRIB Software Library's FLIRT.[31,32] The results in Figure 4B show that all methods have dips/peaks in performance at the same slice indices, providing insight into which slices are more/less challenging in terms of distortion correction. FD-Net outperforms all competing methods in terms of the predicted image quality, especially at the problematic lower brain slices where severe distortions are present.

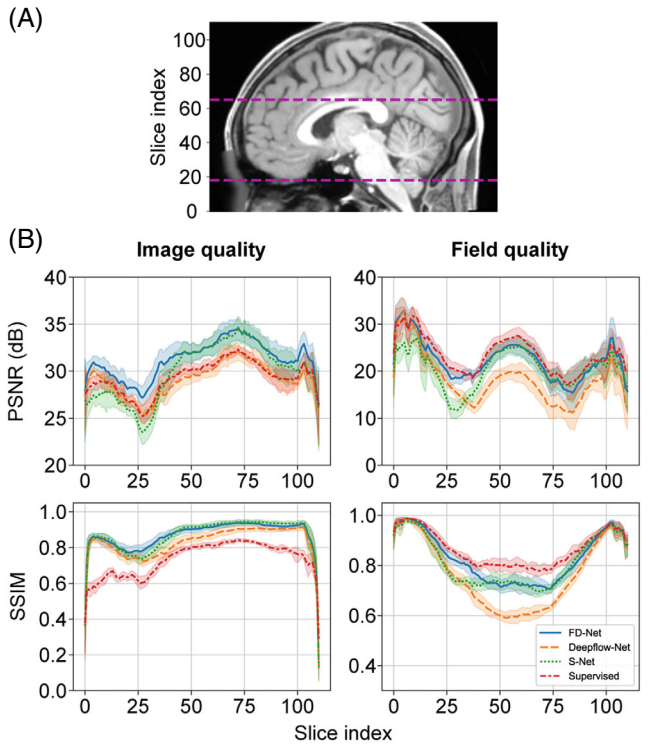


**FIGURE 4** Slice-wise performance comparison of FD-Net and competing methods. (A) An example $T_1$ weighted image registered to the b0 volume of an individual subject to illustrate the anatomical locations corresponding to the slice indices. Magenta dashed lines indicate the locations of more challenging (lower line) and less challenging (upper line) slices in terms of distortion correction (see visual results in Figure 6 and Figure 7). (B) Peak signal-to-noise ratio (PSNR;top row) and structural similarity index measure (SSIM; bottom row) metrics for predicted image (left column) and predicted field (right column). Results are shown for FD-Net and competing methods as a function of slice index. For each method, the mean metric is shown along with the 95% confidence interval.

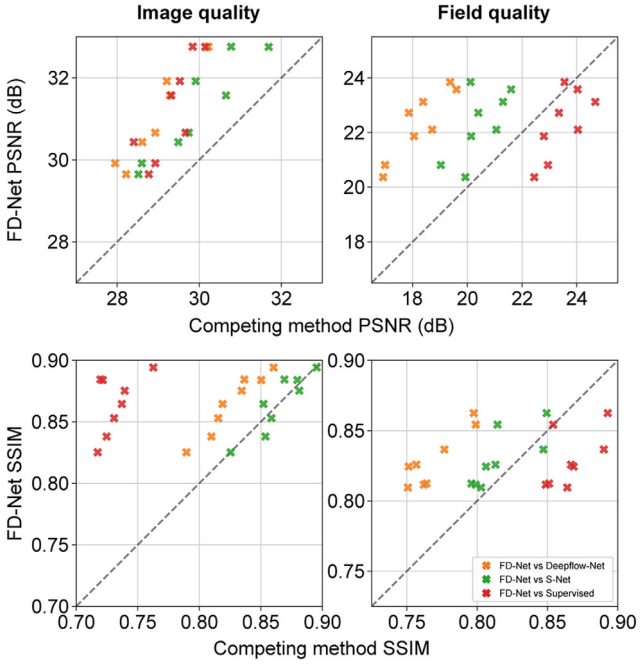


**FIGURE 5** Subject-wise performance comparison of FD-Net against competing methods. Peak signal-to-noise ratio (PSNR; top row) and structural similarity index measure (SSIM; bottom row) metrics for predicted image (left column) and predicted field (right column). Metrics are averaged across slices within individual subjects, and the mean metrics for the eight test subjects are displayed as scatter plots. The vertical axis denotes FD-Net performance, whereas the horizontal axis denotes competing method performance (see legend). The results above the dashed identity lines indicate superior performance by FD-Net.

Moreover, the predicted field quality from FD-Net exceeds the competing methods, except for the supervised baseline. It should be noted that while the supervised baseline is able to match the TOPUP field better, it performs the worst in terms of predicted image quality.

*Subject-wise evaluations*: The performance of each method was assessed over all slices in the volume of a given subject, for each of the eight subjects reserved for testing. Figure 5 gives the scatter plots of mean PSNR and mean SSIM of FD-Net versus each competing method for each subject, for a direct one-to-one performance comparison. In terms of image quality, FD-Net dominates over the competing methods, including the supervised baseline. While S-Net matches FD-Net in terms of SSIM over the predicted image quality, it lags behind in terms of PSNR. As for the predicted field quality, FD-Net is second only to the supervised baseline which was trained to directly fit the results from TOPUP.

*Overall performance evaluations*: The quantitative results in Table 3 summarize the overall performance of each method across all subjects. FD-Net significantly boosts image quality by 2.21 dB PSNR/4.01% SSIM ($p < 0.05$, paired Wilcoxon signed-rank test) over Deepflow-Net, by 1.37 dB PSNR/0.27% SSIM ($p < 0.05$) over S-Net, and by 1.97 dB PSNR/13.54% SSIM ($p < 0.05$) over the supervised baseline. It also boosts field quality by 4.24 dB PSNR/6.11% SSIM ($p < 0.05$) over Deepflow-Net, and by 2.03 dB PSNR/1.49% SSIM ($p < 0.05$) over S-Net, albeit incurs a cost of 1.00 dB PSNR/3.61% SSIM ($p < 0.05$) against the supervised baseline.

*Visual assessments*: To visually compare the qualities of the predicted images and the predicted fields, example results from the slices marked in Figure 4A are provided in Figure 6 for a lower brain slice and Figure 7 for an upper brain slice. These slices were chosen to represent the most and least challenging slices, corresponding to the dip and peak in PSNR in Figure 4B, respectively. The error maps, as well as visual inspection of the predicted image and predicted field, indicate that FD-Net outperforms the other methods. This is especially true at the problematic

**TABLE 3** Performance comparison of FD-Net and the competing methods.

| Methods | Image quality | | Field quality | |
|---|---|---|---|---|
| | PSNR (dB) | SSIM (%) | PSNR (dB) | SSIM (%) |
| **Proposed FD-Net** | **31.29** (**2.79**) | **86.71** (**11.58**) | 22.48 (5.69) | 83.09 (10.37) |
| Deepflow-Net | 29.08 (2.33) | 82.70 (11.72) | 18.24 (6.48) | 76.98 (14.19) |
| S-Net | 29.92 (3.63) | 86.44 (12.16) | 20.45 (5.43) | 81.60 (10.50) |
| Supervised baseline | 29.32 (2.24) | 73.17 (11.26) | **23.48** (**5.06**) | **86.70** (**7.63**) |

*Notes*: Peak signal-to-noise ratio (PSNR) and structural similarity index measure (SSIM) metrics are reported as mean (SD) across subjects. Bold font denotes the best performing method.

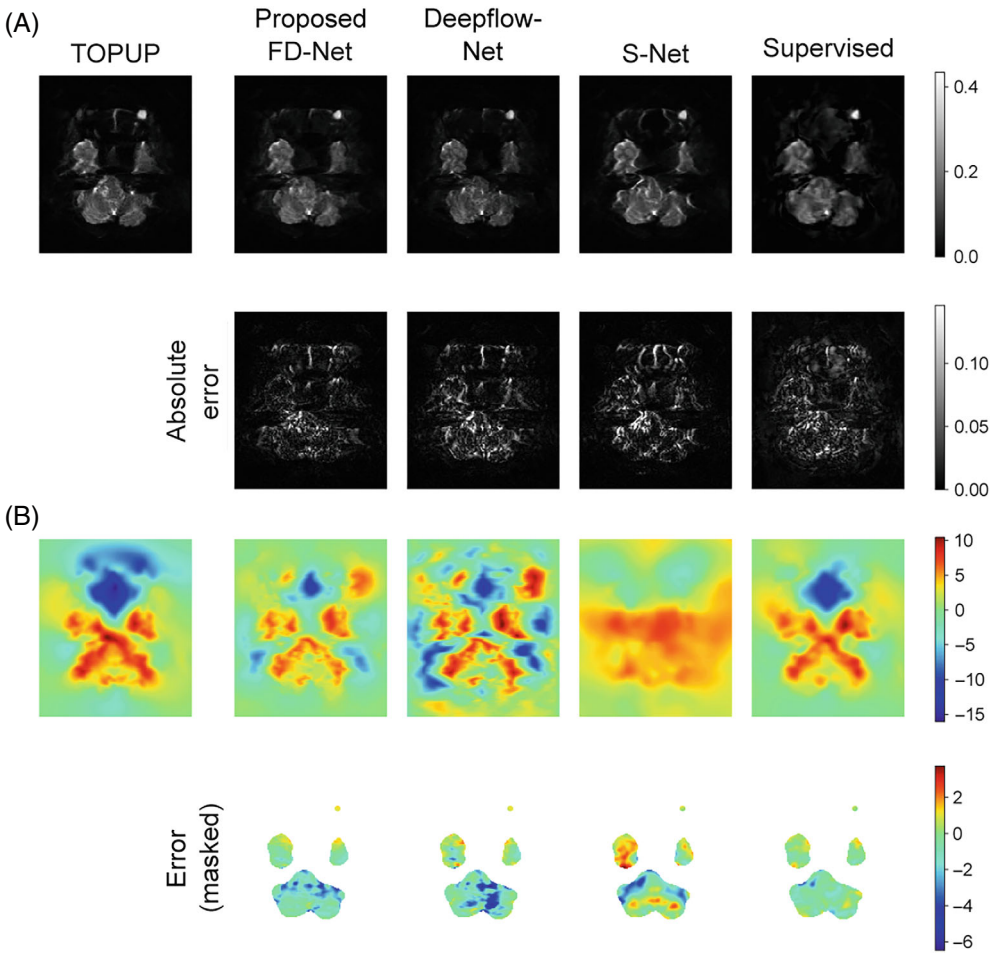


**FIGURE 6** Visual results for FD-Net and competing methods from a lower brain slice, corresponding to a more challenging location in terms of distortion correction. TOPUP results are taken as reference. (A) Predicted images and absolute error maps with respect to TOPUP. The error maps are scaled by 1.25× to a visibly discernible display window. (B) Predicted fields and the masked error maps with respect to TOPUP. The error maps were masked via a median Otsu threshold over the TOPUP image to remove the background regions. See the lower magenta dashed line in Figure 4A for the anatomical location of this slice.

lower brain slice example shown in Figure 6, where large distortions are present. The upper brain slice example in Figure 7 exhibits distortions that are not as severe, indicating a less challenging problem for all methods to solve. For both cases, the predicted images from FD-Net have higher overall similarity to the TOPUP corrected image, with less artifacts present than the other methods. The field results also demonstrate that FD-Net produces the highest fidelity field, with smoothness and details preserved in a coherent manner. Additionally, the forward-distorted images generated by FD-Net closely match the input distorted images for both the lower and upper brain slices, as shown in Figures S3 and S4, respectively.

*Performance evaluations for different training dataset size*: To assess the influence of training set size on model performance, evaluations were performed on training sets with 4 and 42 subjects. In both cases, FD-Net maintains the highest performance in terms of image quality and outperforms the competing unsupervised methods in terms of field quality. Quantitative results for models trained on four subjects are listed in Table S5. FD-Net boosts image quality by 4.36dB PSNR/4.59% SSIM ($p < 0.05$,

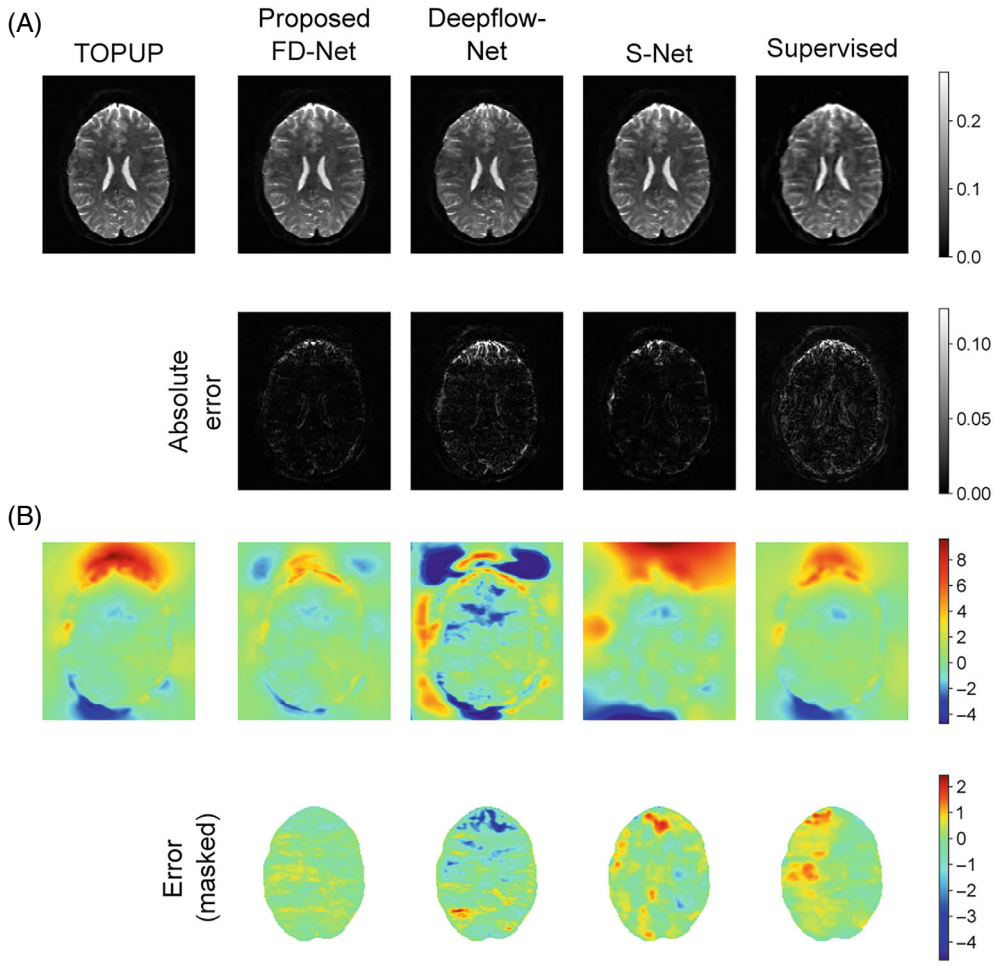


**FIGURE 7** Visual results for FD-Net and competing methods from an upper brain slice, corresponding to a less challenging location in terms of distortion correction. TOPUP results are takes as reference. (A) Predicted images and absolute error maps with respect to TOPUP. The error maps are scaled by 1.25× to a visibly discernible display window. (B) Predicted fields and the masked error maps with respect to TOPUP. The error maps were masked via a median Otsu threshold over the TOPUP image to remove the background regions. See the upper magenta dashed line in Figure 4A for the anatomical location of this slice.

paired Wilcoxon signed-rank test) over Deepflow-Net, by 3.71dB PSNR/2.11% SSIM ($p < 0.05$) over S-Net, and by 3.48dB PSNR/21.45% SSIM ($p < 0.05$) over the supervised baseline. It also boosts field quality by 4.24dB PSNR/6.61% SSIM ($p < 0.05$) over Deepflow-Net, and by 2.43dB PSNR/1.81% SSIM ($p < 0.05$) over S-Net, albeit incurs a cost of 0.91dB PSNR/0.89% SSIM ($p < 0.05$) against the supervised baseline. As listed in Table S6, quantitative results for models trained on 42 subjects are generally similar to those for the 12 subject case in Table 3. FD-Net boosts image quality by 2.24dB PSNR/4.50% SSIM ($p < 0.05$) over Deepflow-Net, by 1.24dB PSNR ($p < 0.05$) over S-Net (while offering similar SSIM), and by 1.39dB PSNR/9.92% SSIM ($p < 0.05$) over the supervised baseline. It also boosts field quality by 5.50dB PSNR/7.00% SSIM over Deepflow-Net, and by 2.13dB PSNR/0.28% SSIM ($p < 0.05$) over S-Net, albeit incurs a cost of 1.18dB PSNR/5.08% SSIM ($p < 0.05$) against the supervised baseline.

*Out-of-domain generalization on fMRI dataset*: Finally, FD-Net was evaluated on fMRI data obtained via the HCP, without and with fine-tuning on the fMRI dataset. Quantitative results in the absence of fine-tuning are listed in Table S7. In terms of image quality, FD-Net yields similar PSNR ($p > 0.05$, paired Wilcoxon signed-rank test) albeit moderately lower SSIM 0.48% ($p < 0.05$) than S-Net, it yields an improvement of 2.19dB PSNR/3.74% SSIM ($p < 0.05$) over Deepflow-Net, and it yields an improvement of 2.29dB PSNR/4.69% SSIM ($p < 0.05$) over the supervised baseline. In terms of field quality, FD-Net attains 1.51% higher SSIM ($p < 0.05$) albeit 0.21dB lower PSNR ($p < 0.05$) than S-Net, it yields an improvement of 0.79dB PSNR/2.27%SSIM ($p < 0.05$) over Deepflow-Net, and it yields an improvement of 0.35dB PSNR/1.89% SSIM ($p < 0.05$) over the supervised baseline. Meanwhile, quantitative results with fine-tuning are listed in Table S8. In this case, FD-Net consistently outperforms competing methods. It boosts image quality by 0.54dB PSNR/0.53% SSIM ($p < 0.05$) over S-Net, by 2.60dB PSNR/3.66% SSIM ($p < 0.05$) over Deepflow-Net, and by 2.80dB PSNR/7.57% SSIM ($p < 0.05$) over the supervised baseline. It also boosts field quality by 2.96dB PSNR/4.98% SSIM ($p < 0.05$) over S-Net, and by 0.80dB PSNR/4.80% SSIM ($p < 0.05$) over Deepflow-Net, while incurring a cost of 3.38dB PSNR/0.84% SSIM ($p < 0.05$) against the supervised baseline.

## 5 | DISCUSSION

In this work, we have proposed a deep forward-distortion model for unsupervised correction of susceptibility artifacts in EPI. FD-Net is based on a multiresolution network model that estimates a single anatomically correct image along with a displacement field, given a pair of reversed-PE acquisitions. Unsupervised learning is achieved by forward-distorting the anatomically correct image with the field, and enforcing consistency of the forward-distorted estimates to the input BU/BD acquisitions. Our results indicate that this forward-distortion approach improves estimation fidelity for both the corrected image and field across a broad range of cross sections in the brain. FD-Net outperforms competing unsupervised methods in image and field quality. It also achieves higher image quality than the supervised baseline, while maintaining the field quality.

Unwarping-based methods rely on similarity losses between corrected BU/BD images to enable unsupervised learning. As these losses are expressed in an inaccessible domain for which no explicit measurements are available, the resultant models can perform suboptimally under large displacements or intensity mismatches. In particular, S-Net uses LCC between corrected images. As a cross-modal similarity measure, LCC is known to be tolerant against intensity mismatches,[56] but places higher emphasis on global features that can incur spatial blur in field estimates. In turn, overly smooth field estimates and lack of density compensation in S-Net can limit its performance in regions of large displacements with abrupt susceptibility changes, particularly near the sinuses and ear canals. To improve reliability against large displacements, Deepflow-Net performs density compensation by estimating pileups via linear interpolation of the grid point density map.[21] However, the MSE loss that it adopts to measure similarity between corrected BU/BD images can lower tolerance against intensity mismatches and induce spatial blur in image estimates. In contrast to unwarping-based methods, the proposed FD-Net leverages a forward-distortion approach based on the K-matrix formulation where density compensation is not needed. For unsupervised learning, it uniquely measures the similarity between forward-distorted images, emulated from estimates of the anatomically correct image and the field, and acquired BU/BD images. As such, the similarity loss is expressed in the actual measurement domain, which can improve performance and reliability of FD-Net as suggested by our experimental results. Quantitative assessments on field quality indicate that the supervised baseline provides a closer match to the TOPUP-estimated displacement field than FD-Net. Yet, the apparent differences are relatively modest based on visual comparisons. On the other hand, FD-Net achieves a notable boost in image quality over the supervised baseline, which is best attributed to the physics-based forward-distortion approach in FD-Net contributing to generalization performance.[23]

Evaluations on different training set sizes indicate that FD-Net demonstrates fast learning and becomes readily performant in image quality with as few as four training subjects. At this relatively compact size, competing methods yield relatively poor image quality. When the size is increased, there is an initial boost from 4 to 12 subjects, yet the benefits for 42 subjects are rather marginal in image quality, more considerable in field quality. Among the competing methods, S-Net and the supervised baseline show consistent performance improvements for larger sizes, whereas Deepflow-Net does not improve notably from 12 to 42 subjects suggesting that its learning process might be saturated.

Evaluations on the fMRI dataset reveal that, without additional training, FD-Net outperforms the supervised baseline in out-of-domain generalization. FD-Net and S-Net show comparable performances and are capable of generalizing to the fMRI domain. Meanwhile, fine-tuning with only four additional subjects rapidly improves the image quality of FD-Net, underlining its rapid learning capability. While fine-tuning also boosts the field quality for the supervised baseline, it does not notably affect its image quality. Consistent with the DWI results, FD-Net surpasses all methods in terms of image quality and the competing unsupervised methods in terms of field quality.

Here, we implemented all unsupervised correction methods by including a rigid loss for consistent and fair comparisons with FD-Net. Based on Table 2, we observe that the rigid loss slightly influences image quality but achieves a modest boost in field quality. This improvement can be attributed to the benefits of spatial registration to account for possible patient motion. The empirical benefits of the rigid loss are expected to become more prominent for increasing levels of motion. We also observe a modest improvement in image quality by inclusion of the valley loss. This benefit can be attributed to the enhanced performance in regions of high field inhomogeneities by avoiding unrealistically large displacements. Similarly, we observe that the bending energy loss that enforces field smoothness is critical to the performance of FD-Net.

As common in deep-learning methods, the trained weights of the FD-Net model are kept fixed during inference. For models trained on limited datasets, this may results in suboptimal generalization to atypical anatomy. In such cases, subject-specific optimization of model weights during inference might improve generalization at the expense of prolonged inference times.[57,58] Here, modules within FD-Net were implemented based on

convolutional backbones given their training efficiency. To improve sensitivity to long-range context in brain images, self-attention based transformer backbones can be adopted.[59] In the current study, all deep-learning models were effectively trained from scratch on relatively modest sized datasets including only 12 subjects. In applications where training data are scarce, network modules can first be pretrained on large public datasets, and later fine-tuned on the application-specific target datasets.[60] Lastly, here we assumed that only reversed-PE images are available as inputs to FD-Net. In cases where additional measurements are viable to capture the field map and/or PSF, FD-Net could be modified to integrate these measurements for improved performance.

It is worth noting that the extent of susceptibility artifacts in EPI can also be reduced by modifying the imaging procedure. For example, methods such as parallel imaging[50,61] or reduced FOV imaging[62,63] decrease sensitivity to field inhomogeneities by encoding a smaller FOV in the PE direction during data acquisition. Similarly, multishot EPI,[64] such as interleaved EPI,[65] can also be performed to reduce field sensitivity. While powerful, these acquisition-based methods still require additional distortion correction in postprocessing. The proposed FD-Net is compatible with this class of methods, as long as a reversed-PE acquisition is performed during imaging.

## 6 | CONCLUSIONS

In this work, we introduced a novel deep-learning approach for efficient and performant correction of susceptibility artifacts in EPI. The proposed FD-Net estimates an anatomically correct image and a displacement field map. It achieves unsupervised learning by leveraging a forward-distortion model to enforce consistency of the estimates to measured reversed-PE images. FD-Net performs competitively with the reference TOPUP method, while offering notably faster inference as a deep learning approach. It also outperforms recent unsupervised correction methods that enforce similarity of unwarped reversed-PE images. Therefore, FD-Net holds great promise for susceptibility-artifact correction in EPI applications.

## ACKNOWLEDGMENTS

A preliminary version of this work was presented in the Annual Meeting of ISMRM in London, 2022. This work was supported by the Scientific and Technological Council of Turkey (TÜBİTAK) via Grant 117E116. Data were provided by the HCP, WU-Minn Consortium (Principal Investigators: David Van Essen and Kamil Ugurbil; 1U54MH091657).

## ORCID

*Abdallah Zaid Alkilani* https://orcid.org/0000-0001-8409-8444

*Tolga Çukur* https://orcid.org/0000-0002-2296-851X

*Emine Ulku Saritas* https://orcid.org/0000-0001-8551-1077

## REFERENCES

1. Mansfield P. Multi-planar image formation using NMR spin echoes. *J Phys C Solid State Phys*. 1977;10:L55-L58.
2. Holdsworth SJ, Bammer R. Magnetic resonance imaging techniques: fMRI, DWI, and PWI. *Semin Neurol*. 2008;28:395-406.
3. Deichmann R, Gottfried JA, Hutton C, Turner R. Optimized EPI for fMRI studies of the orbitofrontal cortex. *Neuroimage*. 2003;19:430-441.
4. Lüdeke KM, Röschmann P, Tischler R. Susceptibility artefacts in NMR imaging. *Magn Reson Imaging*. 1985;3:329-343.
5. Chang H, Fitzpatrick JM. A technique for accurate magnetic resonance imaging in the presence of field inhomogeneities. *IEEE Trans Med Imaging*. 1992;11:319-329.
6. Tournier J, Mori S, Leemans A. Diffusion tensor imaging and beyond. *Magn Reson Med*. 2011;65:1532-1556.
7. Jezzard P. Correction of geometric distortion in fMRI data. *Neuroimage*. 2012;62:648-651.
8. Gallichan D. Diffusion MRI of the human brain at ultra-high field (UHF): a review. *Neuroimage*. 2018;168:172-180.
9. Andersson JLR, Skare S, Ashburner J. How to correct susceptibility distortions in spin-echo echo-planar images: application to diffusion tensor imaging. *Neuroimage*. 2003;20:870-888.
10. Holland D, Kuperman JM, Dale AM. Efficient correction of inhomogeneous static magnetic field-induced distortion in echo planar imaging. *Neuroimage*. 2010;50:175-183.
11. Morgan PS, Bowtell RW, McIntyre DJO, Worthington BS. Correction of spatial distortion in EPI due to inhomogeneous static magnetic fields using the reversed gradient method. *J Magn Reson Imaging*. 2004;19:499-507.
12. Ardekani S, Sinha U. Geometric distortion correction of high-resolution 3 T diffusion tensor brain images. *Magn Reson Med*. 2005;54:1163-1171.
13. Embleton KV, Haroon HA, Morris DM, Ralph MAL, Parker GJM. Distortion correction for diffusion-weighted MRI tractography and fMRI in the temporal lobes. *Hum Brain Mapp*. 2010;31:1570-1587.
14. Smith SM, Jenkinson M, Woolrich MW, et al. Advances in functional and structural MR image analysis and implementation as FSL. *Neuroimage*. 2004;23:S208-S219.
15. Ruthotto L, Mohammadi S, Heck C, Modersitzki J, Weiskopf N. Hyperelastic susceptibility Artifact correction of DTI in SPM. In: *Bildverarbeitung für Die Medizin 2013*. Springer Berlin Heidelberg; 2013:344-349.
16. Modersitzki J. FAIR: Flexible Algorithms for Image Registration. *Society for Industrial and Applied Mathematics*; 2009:76-83.
17. Graham MS, Drobnjak I, Jenkinson M, Zhang H. Quantitative assessment of the susceptibility artefact and its interaction with motion in diffusion MRI. *PloS One*. 2017;12:1-25.

18. Patzig F, Mildner T, Schlumm T, Müller R, Möller HE. Deconvolution-based distortion correction of EPI using analytic single-voxel point-spread functions. *Magn Reson Med*. 2021;85:2445-2461.
19. Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. In: Navab N, Hornegger J, Wells W, Frangi A, eds. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. Springer; 2015.
20. Duong STM, Phung SL, Bouzerdoum A, Schira MM. An unsupervised deep learning technique for susceptibility artifact correction in reversed phase-encoding EPI images. *Magn Reson Imaging*. 2020;71:1-10.
21. Zahneisen B, Baeumler K, Zaharchuk G, Fleischmann D, Zeineh M. Deep flow-net for EPI distortion estimation. *Neuroimage*. 2020;217:116886.
22. Hammernik K, Klatzer T, Kobler E, et al. Learning a variational network for reconstruction of accelerated MRI data. *Magn Reson Med*. 2018;79:3055-3071.
23. Aggarwal HK, Mani MP, Jacob M. MoDL-MUSSELS: model-based deep learning for multishot sensitivity-encoded diffusion MRI. *IEEE Trans Med Imaging*. 2020;39:1268-1277.
24. Andersson JLR. Chapter 4 - geometric distortions in diffusion MRI. *Diffusion MRI*. 2nd ed. Academic Press; 2014:63-85.
25. Van Essen DC, Smith SM, Barch DM, Behrens TEJ, Yacoub E, Ugurbil K. The WU-Minn human connectome project: an overview. *Neuroimage*. 2013;80:62-79.
26. Zeng H, Constable RT. Image distortion correction in EPI: comparison of field mapping with point spread function mapping. *Magn Reson Med*. 2002;48:137-146.
27. Irfanoglu MO, Modi P, Nayak A, Hutchinson EB, Sarlls J, Pierpaoli C. DR-BUDDI (diffeomorphic registration for blip-up blip-down diffusion imaging) method for correcting echo planar imaging distortions. *Neuroimage*. 2015;106:284-299.
28. Qiao Y, Sun W, Shi Y. FOD-based registration for susceptibility distortion correction in brainstem connectome imaging. *Neuroimage*. 2019;202:116164.
29. Esteban O, Zosso D, Daducci A, et al. Surface-driven registration method for the structure-informed segmentation of diffusion MR images. *Neuroimage*. 2016;139:450-461.
30. Li Z, Fan Q, Bilgic B, et al. Diffusion MRI data analysis assisted by deep learning synthesized anatomical images (DeepAnat). *Med Image Anal*. 2023;87:102744.
31. Jenkinson M, Smith S. A global optimisation method for robust affine registration of brain images. *Med Image Anal*. 2001;5:143-156.
32. Jenkinson M, Bannister P, Brady M, Smith S. Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage*. 2002;17:825-841.
33. Tax CMW, Bastiani M, Veraart J, Garyfallidis E, Irfanoglu MO. What's new and what's next in diffusion MRI preprocessing. *Neuroimage*. 2022;249:118830.
34. In M, Posnansky O, Speck O. High-resolution distortion-free diffusion imaging using hybrid spin-warp and echo-planar PSF-encoding approach. *Neuroimage*. 2017;148:20-30.
35. Hu Z, Wang Y, Zhang Z, et al. Distortion correction of single-shot EPI enabled by deep-learning. *Neuroimage*. 2020;221:117170.
36. Ye X, Wang P, Li S, et al. Simultaneous superresolution reconstruction and distortion correction for single-shot EPI DWI using deep learning. *Magn Reson Med*. 2023; 89:2456-2470.
37. Paul D, Zaitsev M, Harsan L, et al. Implementation and application of PSF-based EPI distortion correction to high field animal imaging. *Int J Biomed Imag*. 2009;2009:946271.
38. Schilling KG, Blaber J, Huo Y, et al. Synthesized b0 for diffusion distortion correction (synb0-DisCo). *Magn Reson Imaging*. 2019;64:62-70.
39. Schilling KG, Blaber J, Hansen C, et al. Distortion correction of diffusion weighted MRI without reverse phase-encoding scans or field-maps. *PloS One*. 2020;15:1-15.
40. Qiao Y, Shi Y. Unsupervised deep learning for FOD-based susceptibility distortion correction in diffusion MRI. *IEEE Trans Med Imaging*. 2022;41:1165-1175.
41. Liao C, Bilgic B, Tian Q, et al. Distortion-free, high-isotropic-resolution diffusion MRI with gSlider BUDA-EPI and multicoil dynamic B0 shimming. *Magn Reson Med*. 2021;86:791-803.
42. Zhang Z, Cho J, Wang L, et al. Blip up-down acquisition for spin- and gradient-echo imaging (BUDA-SAGE) with self-supervised denoising enables efficient $T_2$, $T_2^*$, Para- and dia-magnetic susceptibility mapping. *Magn Reson Med*. 2022;88: 633-650.
43. So S, Park HW, Kim B, et al. BUDA-MESMERISE: rapid acquisition and unsupervised parameter estimation for $T_1$, $T_2$, $M_0$, $B_0$, and $B_1$ maps. *Magn Reson Med*. 2022;88:292-308.
44. Cao X, Yang J, Zhang J, et al. Deformable image registration based on similarity-steered CNN regression. *Medical Image Computing and Computer Assisted Intervention – MICCAI 2017*. Springer; 2017:300-308.
45. Krebs J, Mansi T, Delingette H, et al. Robust non-rigid registration through agent-based action learning. *Medical Image Computing and Computer Assisted Intervention – MICCAI 2017*. Springer; 2017:344-352.
46. Yang X, Kwitt R, Styner M, Niethammer M. Quicksilver: fast predictive image registration – a deep learning approach. *Neuroimage*. 2017;158:378-396.
47. Dosovitskiy A, Fischer P, Ilg E, et al. FlowNet: learning optical flow with convolutional networks. Paper presented at: 2015 IEEE International Conference on Computer Vision (ICCV). 2015; Santiago, Chile:2758-2766.
48. Ilg E, Mayer N, Saikia T, Keuper M, Dosovitskiy A, Brox T. FlowNet 2.0: evolution of optical flow estimation with deep networks. Paper presented at: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2017; Honolulu, HI:1647-1655.
49. Balakrishnan G, Zhao A, Sabuncu MR, Guttag J, Dalca AV. VoxelMorph: a learning framework for deformable medical image registration. *IEEE Trans Med Imaging*. 2019;38: 1788-1800.
50. Pruessmann KP, Weiger M, Scheidegger MB, Boesiger P. SENSE: sensitivity encoding for fast MRI. *Magn Reson Med*. 1999;42:952-962.
51. Staring M, Klein S, Pluim JPW. A rigidity penalty term for nonrigid registration. *Med Phys*. 2007;34:4098-4108.
52. Fornberg B. Generation of finite difference formulas on arbitrarily spaced grids. *Math Comput*. 1988;51:699-706.
53. WU-Minn Consortium of the NIH Human Connectome Project. 1200 subjects reference manual – Appendix I. Technical report. Human Connectome Project. 2017.

54. Barch DM, Burgess GC, Harms MP, et al. Function in the human connectome: task-fMRI and individual differences in behavior. *Neuroimage*. 2013;80:169-189. Mapping the Connectome.
55. Garyfallidis E, Brett M, Amirbekian B, et al. Dipy, a library for the analysis of diffusion MRI data. *Front Neuroinform*. 2014;8:1-17.
56. Avants BB, Epstein CL, Grossman M, Gee JC. Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. *Med Image Anal*. 2008;12:26-41.
57. Narnhofer D, Hammernik K, Knoll F, Pock T. Inverse GANs for accelerated MRI reconstruction. In: Descoteaux M, Maier-Hein L, Franz A, Jannin P, Collins D, Duchesne S, eds. *Wavelets and Sparsity XVIII*. Springer; 2019:111381A.
58. Korkmaz Y, Dar SUH, Yurt M, Ozbey M, Çukur T. Unsupervised MRI reconstruction via zero-shot learned adversarial transformers. *IEEE Trans Med Imaging*. 2022;41:1747-1763.
59. Dalmaz O, Yurt M, Çukur T. ResViT: residual vision transformers for multi-modal medical image synthesis. *IEEE Trans Med Imaging*. 2022;41:2598-2614.
60. Dar SUH, Özbey M, Çatli AB, Çukur T. A transfer-learning approach for accelerated MRI using deep neural networks. *Magn Reson Med*. 2020;84:663-685.
61. Griswold MA, Jakob PM, Heidemann RM, et al. Generalized autocalibrating partially parallel acquisitions (GRAPPA). *Magn Reson Med*. 2002;47:1202-1210.
62. Saritas EU, Cunningham CH, Lee JH, Han ET, Nishimura DG. DWI of the spinal cord with reduced FOV single-shot EPI. *Magn Reson Med*. 2008;60:468-473.
63. Barlas BA, Bahadir CD, Kafali SG, Yilmaz U, Saritas EU. Sheared two-dimensional radiofrequency excitation for off-resonance robustness and fat suppression in reduced field-of-view imaging. *Magn Reson Med*. 2022;88:2504-2519.
64. McKinnon GC. Ultrafast interleaved gradient-echo-planar imaging on a standard scanner. *Magn Reson Med*. 1993;30:609-616.
65. Nunes RG, Jezzard P, Behrens TEJ, Clare S. Self-navigated multishot echo-planar pulse sequence for high-resolution diffusion-weighted imaging. *Magn Reson Med*. 2005;53:1474-1478.


## SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

**Figure S1.** Details of the network architecture of FD-Net.
**Figure S2.** Illustration of two different multiresolution approaches, multiscale and multiblur, considered for FD-Net.
**Figure S3.** Visual results for the forward-distorted images from FD-Net for a lower brain slice.
**Figure S4.** Visual results for the forward-distorted images from FD-Net for an upper brain slice.
**Table S1.** Hyperparameter choices for the multiresolution strategy ablation study for FD-Net.
**Table S2.** Performance comparison of multiresolution strategies in FD-Net.
**Table S3.** Hyperparameter choices for the multiblur scheme ablation study for FD-Net.
**Table S4.** Performance comparison of multiblur schemes for FD-Net.
**Table S5.** Performance comparison between FD-Net and competing methods, trained on a set with 4 subjects.
**Table S6.** Performance comparison between FD-Net and competing methods, trained on a set with 42 subjects.
**Table S7.** Out-of-domain generalization comparison between FD-Net and competing methods.
**Table S8.** Out-of-domain generalization comparison between FD-Net and competing methods.

**How to cite this article:** Zaid Alkilani A, Çukur T, Saritas EU. FD-Net: An unsupervised deep forward-distortion model for susceptibility artifact correction in EPI. *Magn Reson Med*. 2024;91:280-296. doi: 10.1002/mrm.29851