Unsupervised Learning-Based Susceptibility Artifact Correction for Diffusion-Weighted MRI in Multiple Organs

Shihan Qiu[1†*], Radu Miron[2*], Yahang Li[1], Cornelius Eichner[3], Thorsten Feiweier[4], Nirmal Janardhanan[1], Bryan Clifford[5], Omar Darwish[4], Mahmoud Mostapha[1], Mariappan S. Nadar[1]

[1] Siemens Healthineers, Princeton, NJ, USA
[2] Siemens Industry Software România, Brasov, Romania
[3] Cancer Therapy Imaging, Siemens Healthineers AG, Forchheim, Germany
[4] Magnetic Resonance, Siemens Healthineers AG, Erlangen, Germany
[5] Siemens Medical Solutions USA, Inc., Boston, MA, USA
[†] shihan.qiu@siemens-healthineers.com

Abstract. Diffusion-weighted MRI (DWI) is widely used for assessing tissue microstructure, with echo-planar imaging (EPI) sequences being the preferred acquisition method due to their fast speed. However, EPI-based DWI is highly sensitive to field inhomogeneities, leading to susceptibility-induced distortions that compromise image quality. Traditional correction methods, such as TOPUP, estimate displacement fields from a pair of reversed phase-encoding (reversed-PE) images to mitigate these distortions. While effective, these approaches suffer from high computational cost, limiting their clinical utility. In this study, we propose an unsupervised learning method for susceptibility artifact correction in EPI. A transformer-style convolutional network enhanced with deformable convolutions is developed to estimate the displacement field from a pair of reversed-PE images, followed by image unwarping and intensity modulation to generate the distortion-free images. This approach surpasses the performance of conventional U-Net-based methods in accuracy. Additionally, a spatially weighted smoothness loss is introduced to enhance robustness against noise in the input data so that the predicted displacement fields from a pair of low b-value DWI can be applied to correct other images with different b-values and diffusion directions from the same subject, optimizing acquisition and computational efficiency. A single model was trained and evaluated on large datasets from multiple organs, acquired with diverse imaging sequences and parameters, at both 1.5T and 3T. Our results demonstrate that the proposed approach achieves generalizable high-quality distortion correction while significantly reducing processing time compared to TOPUP, highlighting its potential for clinical translation.

[*] Equal contribution.

Keywords: Susceptibility artifact correction, Deep learning, Diffusion MRI.

1 Introduction

Diffusion-weighted imaging (DWI) is a widely used MRI technique for characterizing tissue microstructure [1]. Echo-planar imaging (EPI) sequences [2] are commonly employed for DWI due to their rapid acquisition speed. However, EPI sequences are highly sensitive to magnetic field inhomogeneities, leading to susceptibility-induced distortions [3]. These artifacts manifest as geometric deformations and signal reduction or pileups along the phase encoding (PE) direction, affecting image interpretation.

Several post-processing techniques have been developed to mitigate susceptibility distortions in EPI [4–10]. Among them, the most widely established method is TOPUP [5]. Using a pair of reversed phase-encoding (reversed-PE) images as input, which exhibit reversed distortions, TOPUP iteratively estimates a displacement field to match the two images with each other. Despite its effectiveness, TOPUP is computationally expensive, leading to clinically infeasible processing times.

Deep learning (DL) has enabled data-driven methods for susceptibility artifact correction (SAC), where neural networks estimate the displacement field from reversed-PE images to unwarp and correct distortions [11–15]. Once trained, the model estimates the distortion field in a single forward pass, significantly reducing processing time compared to traditional methods. These existing DL-based methods, primarily based on U-Net [16], have been developed mainly for brain imaging at 3T or 7T with limited testing on other anatomies. In DWI applications, they were only evaluated on low b-value images, overlooking the performance on high b-value images which suffer from lower signal-to-noise ratio (SNR) but are critical for advanced diffusion imaging in the clinic.

In this study we propose an unsupervised learning-based SAC method that extends beyond previous works by incorporating a transformer-style deformable convolutional modulation as network building blocks and improves distortion correction accuracy in comparison with U-Net architectures. Additionally, a spatially weighted smoothness loss was introduced to improve the robustness to noise so that the estimated displacement field can be used to correct not only the input reversed-PE images but also other DWI of the same subject, including high b-value acquisitions, with minimum noise propagation. Our DL model was trained and evaluated on a large multi-organ dataset with diverse imaging sequences, enabling a single model for SAC in different DWI applications.

2 Methods

2.1 Susceptibility artifact correction with reversed-PE images

Susceptibility artifacts in EPI mainly manifest as distortions along the PE direction, which can be parametrized by a unidirectional displacement field shifting pixel coordinates. The displacement field U maps the distorted image I_1 to the corrected image I_c, while $-U$ maps the reversed-encoded image I_2 to the same corrected image:

$$I_c = (I_1 \circ (Id + U)) \cdot J_{det}(Id + U), \tag{1}$$

$$I_c = (I_2 \circ (Id - U)) \cdot J_{det}(Id - U), \tag{2}$$

where Id represents the identity transformation, and $\circ$ denotes spatial unwarping. The unwarping process $I_1 \circ (Id + U)$ can be implemented as taking pixel value of the distorted image $I_1(x, y + U(x, y))$ to generate the corrected pixel at (x, y), with the second dimension being PE direction. $J_{det}(\cdot)$ is the Jacobian determinant to correct for intensity variations caused by local expansion and compression of the image signal.

We trained a neural network to learn a mapping from the input reversed-PE images (I_1, I_2) to the displacement field U. With the estimated U, corrected images I_{c1} and I_{c2} can be generated according to Equations (1) and (2) from I_1 and I_2, respectively. Additionally, the displacement field can be used to correct other DWI images, e.g., with different b-values or diffusion directions, acquired with the same distortion pattern. Since low b-value DWI images typically have higher SNR, we estimate U using a pair of low b-value images and apply it to correct high b-value images for the same subject.

2.2 Network architecture for displacement estimation

A 2D encoder-decoder architecture is built to estimate the displacement. In EPI, distortions vary spatially, with large distortions at regions with strong susceptibility effects. Therefore, network modules with large and adaptive receptive fields are favorable.

We adopt the convolutional modulation block from Conv2Former [17], which mimics the self-attention in Vision Transformers [18] with a convolutional modulation:

$$\mathbf{A} = \mathrm{DConv}_{k\times k}\,(\mathbf{W_1 X_{in}}), \tag{3}$$

$$\mathbf{V} = \mathbf{W}_2 \mathbf{X}_{in}, \tag{4}$$

$$\mathbf{X}_{out} = \mathbf{W}_3(\mathbf{A} \odot \mathbf{V}) + \mathbf{X}_{in}, \tag{5}$$

where $\mathrm{DConv}_{k\times k}$ is a depth-wise convolution with $k \times k$ kernel, $\mathbf{W}_{1,2,3}$ are weight matrices of linear layers, and $\odot$ is the Hadamard product. This design enables each spatial location to interact with all the pixels within the $k \times k$ window centered at this location.

We further replace the depth-wise convolution with a deformable convolution [19], updating Equation (3) to:

$$\mathbf{A} = \mathrm{DefConv}_{k\times k}(\mathbf{W}_1 \mathbf{X}_{in}). \tag{6}$$

The deformable convolution layer adaptively augments the $k \times k$ square sampling region in convolution with additional learnable offsets for different spatial locations, making it suitable for characterizing the spatially varying distortions in EPI.

The complete network architecture is illustrated in **Fig. 1**.

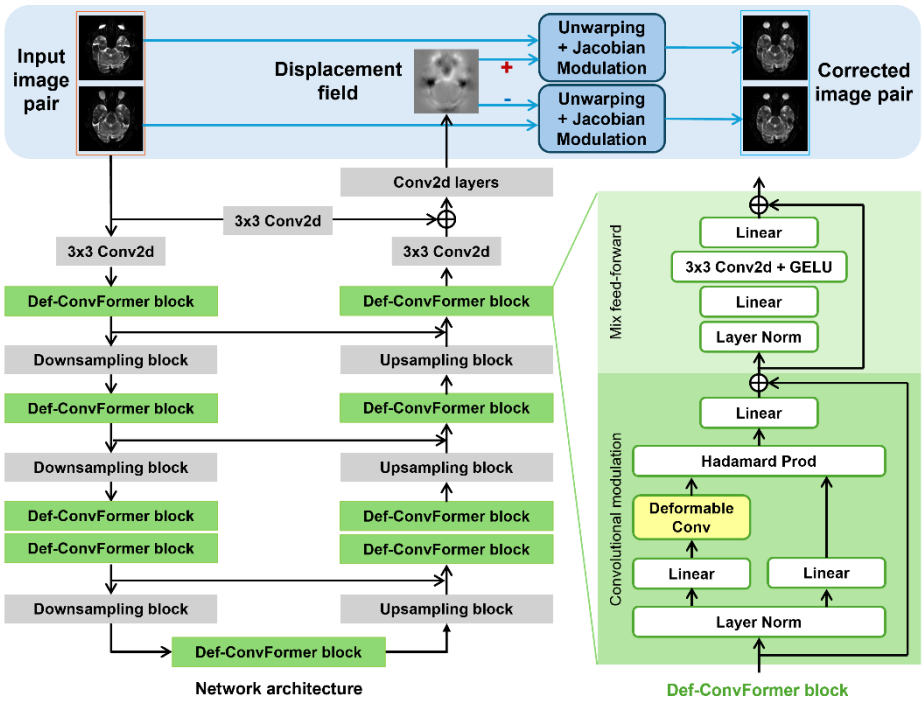

Fig. 1. Network architecture. The network employs an encoder-decoder architecture built with Def-Convformer blocks. It takes a reversed-PE image pair and outputs a displacement field, which is used to correct the distorted images through image unwarping and Jacobian modulation.

2.3 Unsupervised loss with adaptive smoothness constraint

Unsupervised learning is used, with the loss function consisting of two components:

$$L = L_{MSE}(I_{c1}, I_{c2}) + \lambda L_{smooth}(U, M). \tag{7}$$

$L_{MSE}(I_{c1}, I_{c2})$ measures the mean square error (MSE) between the two corrected images. $L_{smooth}(U, M)$ is a spatially weighted smoothness loss on the estimated field U:

$$L_{smooth}(U, M) = \frac{1}{N_x N_y} \sum_{x,y} M(x, y) \left(\left(\frac{\partial^2 U(x,y)}{\partial x^2} \right)^2 + \left(\frac{\partial^2 U(x,y)}{\partial y^2} \right)^2 \right). \tag{8}$$

The weighting matrix M is derived from the input images, adaptively regularizing the smoothness penalty across the image. In background regions with low signal intensity, dominated by noise, it enforces a stronger smoothness constraint to prevent noise from being transferred into the estimated displacement field. It is calculated as follows:

$$D = \text{Gaussion}(\max(I_1, I_2)), \tag{9}$$

$$M = \left(1 - \frac{D}{\max(D)}\right)(1 - a) + a, \tag{10}$$

where $\mathrm{Gaussion}(\cdot)$ denotes a Gaussian blurring operator, and a is a parameter that controls the relative smoothness penalty in high-intensity regions.

2.4 Datasets and training experiments

The datasets used for training and evaluating the network are summarized in **Table 1**. Multiple organs with various imaging sequences and parameters are included. Whereas the images are primarily low b-value DWI, a set of multi-shot EPI-based neural anatomical scans is also included since the distortion mechanism is the same. Data splitting was performed subject-wise within each dataset, resulting in 2300/137/244 scans for training/validation/testing. Additional six DWI scans at 3T with both low b-values and high b-values (800-1000 s/mm^2) in brain, prostate, abdomen, and spine were collected to validate and test the correction performance on high b-value DWI using the estimated displacement field from low b-value images. All data was acquired on clinical scanners (different MAGNETOM models, Siemens Healthineers AG, Forchheim, Germany), with informed consent from all subjects.

Table 1. Summary of datasets.

	Organ	FS	Sequences	Plane	Subj	Vol	Slices
Multi-organ DWI	Brain	3T	ss SE-EPI / RESOLVE, DWI (b=0)	Tra, Cor, Sag	29	188	8634
	Head-neck	3T	ss IR-EPI / RESOLVE, DWI (b=0 / 50 s/mm^2)	Tra	25	389	8344
		1.5T			21	167	3758
	Prostate	3T	ss SE-EPI / ZOOMit ss SE-EPI / RESOLVE, DWI (b=50 s/mm^2)	Tra	14	246	5412
		1.5T			11	149	3638
	Cervix	3T	ss SE-EPI / ZOOMit ss SE-EPI / RESOLVE, DWI (b=50 s/mm^2)	Tra, Sag	12	176	4992
		1.5T			11	80	2583
Prostate DWI	Prostate	3T	ss SE-EPI / ZOOMit ss SE-EPI, DWI (b=50 s/mm^2)	Tra	206	950	23750
Neurology EPI	Brain	3T	ms EPI, T1 / T2 / T2* / FLAIR	Tra, Cor, Sag	14	336	12096
Sum					343	2681	73207

To evaluate the effectiveness of the proposed network architecture and loss design, two sets of experiments were conducted. First, the proposed network with deformable convolution modulation blocks and 3×3 kernels (denoted as Def-Convformer) was compared against three ablated architectures: 1) a Convformer architecture without deformable convolutions but using depth-wise convolution with a 7×7 kernel, 2) a U-Net with deformable convolutions (Def-U-Net), and 3) a standard U-Net with 3×3 convolutions, which is commonly used in existing DL-based SAC methods. Second, the proposed spatially weighted smoothness loss, with empirically selected parameters $\lambda = 5$ and

$a = 0.2$, was compared to globally uniform smoothness losses with λ values of 0.1, 1, and 5. Additionally, TOPUP was used as a reference method for comparison. Normalized mean square error (NMSE), peak signal-to-noise ratio (PSNR), and structural similarity index (SSIM) were computed volume-wise between the two corrected images to evaluate the correction quality.

All networks were implemented in Pytorch, trained on a machine with an NVIDIA A100-SXM4-40GB GPU. Adam optimizer with a learning rate of 0.0001 was used.

3 Results

The similarity metrics between the two corrected images in the test set are summarized in **Table 2**. Among the four networks (model 4-7), the proposed Def-Convformer achieved the lowest NMSE and the highest PSNR and SSIM. The results in **Fig. 2** also show its superior performance in correcting high-resolution scans with large distortions.

As for loss function design (i.e., comparing model 1-4), the proposed Def-Convformer with spatially weighted smoothness loss (model 4) achieved comparable metrics with TOPUP. Whereas using a smaller smoothness penalty (e.g., $\lambda \leq 1$) improved the similarity between corrected low b-value images and achieved better metrics than TOPUP, the estimated displacement field tended to be noisy in the case of low-SNR inputs, which introduced substantial noise and artifacts into the high b-value images when it was applied to correct them (**Fig. 3**). The displacement field from TOPUP also showed a similar noise transfer issue. In contrast, the proposed spatially weighted smoothness loss suppressed the noise transfer in the low-SNR background region with a stronger smoothness penalty while maintaining the distortion correction performance in high-SNR regions using a smaller smoothness loss.

In terms of inference speed, the DL model took 3.15 sec to process a 224×224×32 image volume on an RTX A2000 GPU (4GB), whereas TOPUP took 44 min on CPU on an Intel Xeon Gold 6348H node with a memory allocation of 128GB.

Table 2. Similarity metrics between the two corrected images.

Id	Model	Loss	NMSE	PSNR (dB)	SSIM
1	Def-Convformer	$\lambda = 0.1$, w/o M	0.015	37.24	0.950
2	Def-Convformer	$\lambda = 1$, w/o M	0.030	34.25	0.914
3	Def-Convformer	$\lambda = 5$, w/o M	0.047	32.24	0.882
4	Def-Convformer	$\lambda = 5$, w/ M	0.036	33.48	0.898
5	Convformer	$\lambda = 5$, w/ M	0.038	33.19	0.894
6	Def-U-Net	$\lambda = 5$, w/ M	0.038	33.28	0.894
7	U-Net	$\lambda = 5$, w/ M	0.043	32.69	0.888
	TOPUP		0.035	33.34	0.907

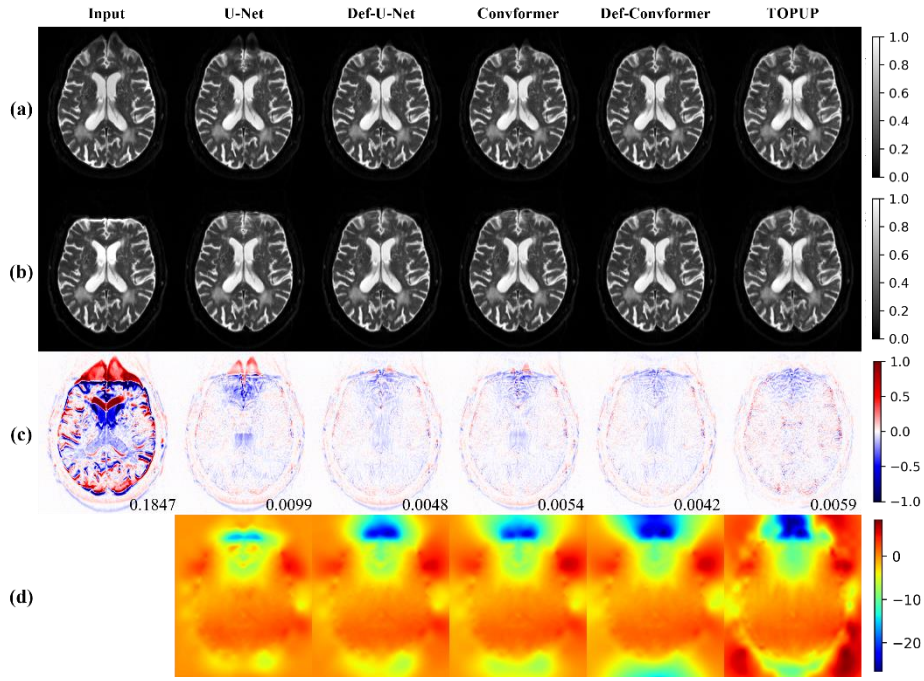

Fig. 2. Example results in a high-resolution brain scan (b-value=0 s/mm^2, matrix size = 384×384) with different network architectures. (a) and (b): a pair of images with opposite phase encodings along the anterior-posterior axis, (c): difference maps (×2) between the image pair, with NMSE labeled, (d): estimated displacement fields.

4 Discussion

An unsupervised learning-based SAC method was developed for EPI, with a novel Def-Convformer network for displacement field estimation from a reversed-PE image pair. A spatially weighted smoothness loss was designed to improve robustness to noise while maintaining correction quality. The method was evaluated in multiple organs with diverse sequences and protocols, showing generalizable performance.

EPI distortions vary spatially, with large displacements in regions of strong susceptibility effects. Existing DL-based correction methods predominantly rely on U-Net architectures, which use fixed 3×3 convolutional kernels with local receptive fields. While effective for many tasks, this design limits the network's ability to handle large, non-uniform displacements. To address this, our model incorporates deformable convolutional modulation blocks, allowing the receptive field to dynamically adjust and the features to be modulated based on local inputs. This design outperforms U-Net-based models, particularly in high-resolution images with severe distortions.

The correction method requires two EPI images with reversed PE directions. In DWI, acquiring such pairs for every b-value and diffusion direction is time-consuming, especially for high b-value DWI with low SNR requiring multiple averages or for large

numbers of diffusion directions. To improve scan efficiency and to ensure a consistent unwarping independent of diffusion weighting or direction, we estimate the displacement field from only one pair of reversed-PE low b-value DWI and apply it to other single-PE images, making the approach clinically applicable by only adding one fast EPI scan. This design necessitates a generalizable displacement field. Our experiments show that while lower smoothness loss achieves better similarity metrics in correcting the input low b-value images, the estimated displacement fields transfer noise from the inputs to the high b-value images. Conversely, a stronger smoothness penalty prevents the displacement field from being noisy, but compromises correction performance. To improve this trade-off, a spatially varying smoothness loss weighted by input intensities was designed, which achieved a better balance by enforcing stronger smoothness penalty in low-SNR background regions while applying a smaller smoothness loss in high-SNR regions.

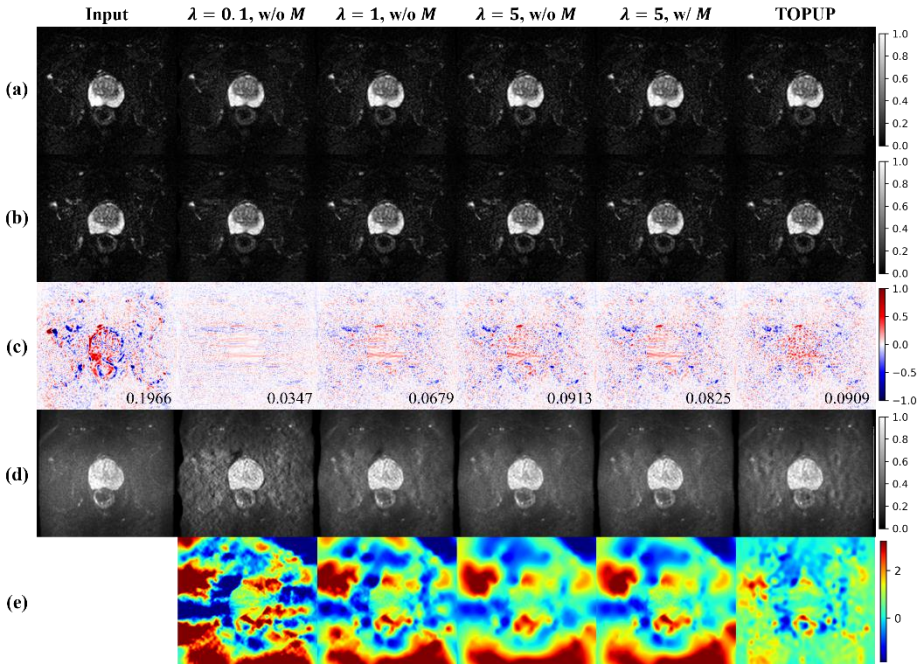

Fig. 3. Example results in prostate DWI with different loss designs. (a) and (b): a pair of low b-value images with opposite phase encodings along the left-right axis (b-value = 50 s/mm^2), (c): difference maps (×2) between the low b-value image pair, with NMSE labeled, (d): high b-value DWI (b-value = 800 s/mm^2) with same distortion characteristics as (a), (e): estimated displacement fields from the low b-value image pair, which were directly applied to the high b-value image with unwarping and Jacobian modulation to provide the corrected results in (d). The fifth column ($\lambda = 5$, w/ M) is the proposed method.

In this study, motion between scans is assumed to be small because of the short acquisition time of EPI. More evaluations on motion effects need to be done. A rigid

transformation can be further incorporated to be jointly estimated along with the displacement field in future studies [14]. This paper primarily focuses on DWI, but the correction method is also applicable to other EPI-based MRI, such as functional MRI.

5 Conclusion

An unsupervised learning-based susceptibility artifact correction method was developed for EPI. With the proposed Def-Convformer network and a spatially weighted smoothness loss, a displacement field is estimated from a pair of reversed-PE images and then used for correcting the distortions. The Def-Convformer architecture achieved superior performance in comparison with conventional U-Net-based models. The DL approach performs competitively with TOPUP across multiple organs, while providing notably faster inference speed, suggesting great potential for clinical applications.

Disclosure of Interests. Shihan Qiu, Yahang Li, Nirmal Janardhanan, Mahmoud Mostapha, and Mariappan S. Nadar are employees at Siemens Healthineers. Cornelius Eichner, Thorsten Feiweier, and Omar Darwish are employees at Siemens Healthineers AG. Radu Miron is an employee at Siemens Industry Software România. Bryan Clifford is an employee at Siemens Medical Solutions USA.

Disclaimer. The concepts and information presented in this paper are based on research results that are not commercially available. Future commercial availability cannot be guaranteed.

References

1. Baliyan, V., Das, C.J., Sharma, R., Gupta, A.K.: Diffusion weighted imaging: Technique and applications. World J. Radiol. 8, 785 (2016). https://doi.org/10.4329/wjr.v8.i9.785.
2. Mansfield, P.: Multi-planar image formation using NMR spin echoes. J. Phys. C Solid State Phys. 10, L55–L58 (1977). https://doi.org/10.1088/0022-3719/10/3/004.
3. Lüdeke, K.M., Röschmann, P., Tischler, R.: Susceptibility artefacts in NMR imaging. Magn. Reson. Imaging. 3, 329–343 (1985). https://doi.org/10.1016/0730-725X(85)90397-2.
4. Chang, H., Fitzpatrick, J.M.: A technique for accurate magnetic resonance imaging in the presence of field inhomogeneities. IEEE Trans. Med. Imaging. 11, 319–329 (1992). https://doi.org/10.1109/42.158935.
5. Andersson, J.L.R., Skare, S., Ashburner, J.: How to correct susceptibility distortions in spin-echo echo-planar images: application to diffusion tensor imaging. NeuroImage. 20, 870–888 (2003). https://doi.org/10.1016/S1053-8119(03)00336-7.
6. Ruthotto, L., Kugel, H., Olesch, J., Fischer, B., Modersitzki, J., Burger, M., Wolters, C.H.: Diffeomorphic susceptibility artifact correction of diffusion-weighted magnetic resonance images. Phys. Med. Biol. 57, 5715–5731 (2012). https://doi.org/10.1088/0031-9155/57/18/5715.
7. Irfanoglu, M.O., Modi, P., Nayak, A., Hutchinson, E.B., Sarlls, J., Pierpaoli, C.: DR-BUDDI (Diffeomorphic Registration for Blip-Up blip-Down Diffusion Imaging) method for correcting echo planar imaging distortions. NeuroImage. 106, 284–299 (2015). https://doi.org/10.1016/j.neuroimage.2014.11.042.

8. Duong, S.T.M., Schira, M.M., Phung, S.L., Bouzerdoum, A., Taylor, H.G.B.: Anatomy-Guided Inverse-Gradient Susceptibility Artifact Correction Method for High-Resolution FMRI. In: 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 786–790. IEEE, Calgary, AB (2018). https://doi.org/10.1109/ICASSP.2018.8461352.
9. Duong, S.T.M., Phung, S.L., Bouzerdoum, A., Boyd Taylor, H.G., Puckett, A.M., Schira, M.M.: Susceptibility artifact correction for sub-millimeter fMRI using inverse phase encoding registration and T1 weighted regularization. J. Neurosci. Methods. 336, 108625 (2020). https://doi.org/10.1016/j.jneumeth.2020.108625.
10. Holland, D., Kuperman, J.M., Dale, A.M.: Efficient correction of inhomogeneous static magnetic field-induced distortion in Echo Planar Imaging. NeuroImage. 50, 175–183 (2010). https://doi.org/10.1016/j.neuroimage.2009.11.044.
11. Duong, S.T.M., Phung, S.L., Bouzerdoum, A., Schira, M.M.: An unsupervised deep learning technique for susceptibility artifact correction in reversed phase-encoding EPI images. Magn. Reson. Imaging. 71, 1–10 (2020). https://doi.org/10.1016/j.mri.2020.04.004.
12. Zahneisen, B., Baeumler, K., Zaharchuk, G., Fleischmann, D., Zeineh, M.: Deep flow-net for EPI distortion estimation. NeuroImage. 217, 116886 (2020). https://doi.org/10.1016/j.neuroimage.2020.116886.
13. Legouhy, A., Graham, M., Guerreri, M., Stee, W., Villemonteix, T., Peigneux, P., Zhang, H.: Correction of Susceptibility Distortion in EPI: A Semi-supervised Approach with Deep Learning. In: Cetin-Karayumak, S., Christiaens, D., Figini, M., Guevara, P., Pieciak, T., Powell, E., and Rheault, F. (eds.) Computational Diffusion MRI. pp. 38–49. Springer Nature Switzerland, Cham (2022). https://doi.org/10.1007/978-3-031-21206-2_4.
14. Zaid Alkilani, A., Çukur, T., Saritas, E.U.: FD-Net: An unsupervised deep forward-distortion model for susceptibility artifact correction in EPI. Magn. Reson. Med. 91, 280–296 (2024). https://doi.org/10.1002/mrm.29851.
15. Bao, Q., Xie, W., Otikovs, M., Xia, L., Xie, H., Liu, X., Liu, K., Zhang, Z., Chen, F., Zhou, X., Liu, C.: Unsupervised cycle-consistent network using restricted subspace field map for removing susceptibility artifacts in EPI. Magn. Reson. Med. 90, 458–472 (2023). https://doi.org/10.1002/mrm.29653.
16. Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab, N., Hornegger, J., Wells, W.M., and Frangi, A.F. (eds.) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. pp. 234–241. Springer International Publishing, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28.
17. Hou, Q., Lu, C.-Z., Cheng, M.-M., Feng, J.: Conv2Former: A Simple Transformer-Style ConvNet for Visual Recognition. IEEE Trans. Pattern Anal. Mach. Intell. 46, 8274–8283 (2024). https://doi.org/10.1109/TPAMI.2024.3401450.
18. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., Houlsby, N.: An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale, https://arxiv.org/abs/2010.11929, (2020). https://doi.org/10.48550/ARXIV.2010.11929.
19. Dai, J., Qi, H., Xiong, Y., Li, Y., Zhang, G., Hu, H., Wei, Y.: Deformable Convolutional Networks. In: 2017 IEEE International Conference on Computer Vision (ICCV). pp. 764–773. IEEE, Venice (2017). https://doi.org/10.1109/ICCV.2017.89.