Learning to Distort: Weakly-Supervised Image Quality Transfer for Prostate DWI Correction

Yucheng Tang[1], Wen Yan[1], Alexander Ng[2], Natasha Thorley[3], Pawel Rajwa[2], Yipei Wang[1], Aqua Asif[4,2], Clare Allen[5], Louise Dickinson[5], Francesco Giganti[2,5], David Atkinson[3], Shonit Punwani[5,3], Daniel C. Alexander[6,7], Shaheer Ullah Saeed[1], Veeru Kasivisvanathan[2,8], and Yipeng Hu[1]

[1] UCL Hawkes Institute and Department of Medical Physics and Biomedical Engineering, University College London, UK; yucheng.tang.24@ucl.ac.uk
[2] Division of Surgery and Interventional Science, University College London, UK
[3] Centre for Medical Imaging, University College London, UK
[4] British Urology Researchers in Surgical Training (BURST), UK
[5] Department of Radiology, University College London Hospitals NHS Foundation Trust, UK
[6] Centre for Medical Image Computing, University College London, UK
[7] Department of Computer Science, University College London, UK
[8] Department of Urology, University College London Hospitals NHS Foundation Trust, UK

Abstract. Single-shot echo-planar prostate diffusion-weighted imaging (DWI) is frequently complicated by geometric distortions, which impact the ability to derive reliable diagnoses from such images. Developing automated correction methods is challenged by the absence of paired distorted and undistorted clinical scans. In this paper, we first propose a novel weakly-supervised image quality transfer (IQT) framework from undistorted to distorted images that utilizes image quality assessment (IQA) signals to supervise the transfer process. Unlike traditional methods that require expensive, voxel-wise paired data or resort to developing unpaired algorithms, our approach utilizes image-level quality labels (here, distorted vs. undistorted) to establish latent quality prototypes within a pre-trained feature space. Recognizing that simulating realistic distortions is more reliable than direct unpaired correction, we describe a weakly-supervised prototype flow matching algorithm to explicitly regularize generative trajectories towards distorted prototypes, producing realistic susceptibility artifacts that mimic clinical degradations. By synthesizing these realistic pairs, we enable a second IQT model to be trained in the forward direction for distortion correction. Experimental results demonstrate that our generated images successfully mimic the diagnostic interference of real-world artifacts, which leads to more capable distortion correction IQT models. In addition to qualitative comparisons, we also conduct exhaustive quantitative evaluations that compare our approach with existing unpaired approaches (e.g., CycleGAN, UNIT-DDPM, and OT-FM) - as either forward or reverse alternatives - by assessing clinical downstream task performance in PI-RADS and Gleason score classification, using both in-distribution and external data sets. Code is available at https://github.com/yucheng722/ProtoFM-IQT.

Keywords: Prostate DWI · Flow Matching · Image Transfer.

1 Introduction

Prostate diffusion-weighted imaging (DWI) is playing a growing role in cancer detection, grading, risk assessment and their subsequent biopsy and treatment planning. However, single-shot echo-planar imaging is inherently prone to susceptibility artifacts caused by rectal air and patient motion [6]. These artifacts lead to geometric distortions, thereby misleading the spatial correspondence between DWI and anatomical T2-weighted images (T2WI) [16] and, when used alone, its true zonal and pathology localisation with respect to surrounding anatomy. In clinical practice, distorted DWI often needs repeated scans, extending examination time and increasing patient and healthcare burden.

Although acquisition-stage techniques [11, 4] and reverse phase-encoding tools such as TopUp [1] alleviate distortions effectively, they require additional scans or paired images with opposite encoding directions, which are often unavailable in routine clinical workflows. Fully-supervised image quality transfer (IQT) methods [12, 9, 3] rely on paired distorted-undistorted scans, which are generally infeasible. Unpaired IQT methods, including CycleGAN [17] and diffusion-based models [25, 18] may respectively face challenges in unstable training [10] and high inference latency [19], with additional risk in anatomical hallucinations [20].

To address these challenges, we propose a novel weakly-supervised IQT method that overcomes paired data scarcity. Using weak quality-based supervision, simulating distortions is found more reliable than direct correction (speculatively, generative mistakes (e.g., hallucinations) during degradation are less harmful than hallucinating non-existent or removing true anatomical structures during correction). Therefore, we first generate realistic distortions from undistorted inputs based on flow matching (FM) [13], where 'realistic' refers to clinically relevant DWI degradation patterns with measurable impact on downstream PI-RADS and Gleason scoring. The resulting synthetic pairs are then used to train a second supervised correction model, which maps generated distorted inputs back to clinically accepted undistorted DWI references, thereby reducing the risk of anatomically implausible hallucinations from direct unpaired correction.

Our contributions include: First, a novel weak-IQA-supervised prototype FM method to generate realistic distorted DWI from undistorted inputs. Second, an effective IQT algorithm reformulating the problem as realistic distortion generation followed by fully-supervised correction. Third, using two clinical downstream tasks on two independent data sets, we first demonstrate that classifiers evaluated on our synthesized distorted images exhibit significant performance degradation in PI-RADS and Gleason scoring, indicating strong correlation between the synthesized artifacts and diagnostic decision-making. Furthermore, we show that the generated distorted-undistorted pairs can be successfully used to train a second IQT model that effectively corrects real-world DWI distortions.

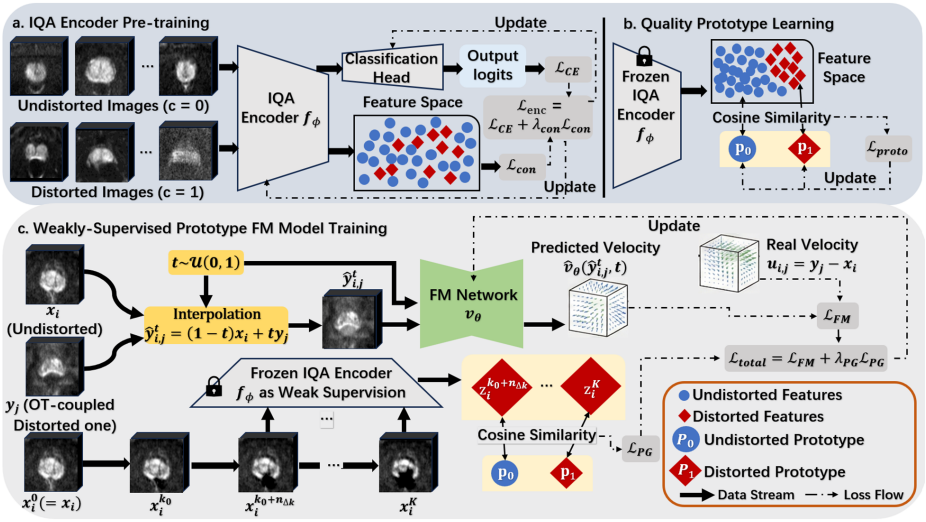

Fig. 1. Overview of the proposed Weakly-Supervised Prototype FM framework. (a) An IQA encoder is pre-trained to construct a feature space that distinguishes distorted and undistorted images. (b) Quality prototypes are learned to represent the centers of these domain distributions. (c) Training of the FM-based IQT model, where the generation trajectory is explicitly guided toward the distorted manifold using the pre-trained encoder and learned prototypes.

2 Method

The proposed IQT method comprises three stages: 1) constructing a weak supervision signal via an IQA encoder and quality prototypes, as shown in Fig. 2(a) and (b), and 2) training a FM network constrained by this signal, as shown in Fig. 2(c). 3) Inference via the trained FM network and use the generated paired images to fully supervise a second IQT model.

2.1 Construction of Weak Quality Supervision Signal

IQA Encoder Pre-training. We first pre-train an IQA encoder $f_\phi(\cdot)$, to discriminate distorted features from undistorted ones and extract distortion-related representations. The encoder accepts a 3D DWI volume x_i as input and produces a global feature vector $\mathbf{z}_i \in \mathbb{R}^d$, where d denotes the feature dimension. Consider a mini-batch $\mathcal{B}$ of samples $\{(x_i, c_i)\}_{i=1}^{|\mathcal{B}|}$, where $c_i \in \{0, 1\}$ is the binary label with $c_i = 0$ and $c_i = 1$ representing undistorted and distorted images, respectively. The IQA encoder extracts the feature vector $\mathbf{z}_i$, which is then mapped to the output logits $\mathbf{o}_i$ via a linear classification head. The encoder and classification head are optimized using the standard binary cross-entropy loss $\mathcal{L}_{\text{CE}}$ against the ground-truth labels c_i. To encourage a well-structured feature space with compact intra-class clusters and clear inter-class separation, we incorporate a

contrastive loss $\mathcal{L}_{\text{con}}$. This loss maximizes feature similarity among samples of the same class while separating features from different classes. Given a temperature parameter τ, the loss is formulated as in Eq. (1)

$$\mathcal{L}_{\text{con}} = \frac{1}{|\mathcal{B}|} \sum_{i=1}^{|\mathcal{B}|} \frac{-1}{|\mathcal{P}(i)|} \sum_{p \in \mathcal{P}(i)} \log \frac{\exp(\cos(\mathbf{z}_i, \mathbf{z}_p)/\tau)}{\sum_{j \neq i} \exp(\cos(\mathbf{z}_i, \mathbf{z}_j)/\tau)} \tag{1}$$

where $\cos(\mathbf{a}, \mathbf{b}) = \frac{\mathbf{a} \cdot \mathbf{b}}{|\mathbf{a}||\mathbf{b}|}$. $\mathcal{P}(i) = \{p \in \mathcal{B} \mid c_p = c_i, p \neq i\}$ is the set of samples in the same class as i, excluding itself. The total encoder training loss is defined as $\mathcal{L}_{\text{enc}} = \mathcal{L}_{\text{CE}} + \lambda_{\text{con}} \mathcal{L}_{\text{con}}$, where λ_{con} is a hyper-parameter weighting the contrastive loss.

Quality Prototype Learning. Subsequently, we train the quality prototypes with the IQA encoder frozen. We optimize two prototype vectors, $\mathbf{p}_0 \in \mathbb{R}^d$ and $\mathbf{p}_1 \in \mathbb{R}^d$, corresponding to undistorted and distorted images, respectively. For each sample (x_i, c_i), the extracted feature $\mathbf{z}_i$ is optimized to be closer to its class prototype and further from the opposing prototype. We utilize cosine similarity as the metric and apply a margin ranking loss with a constant margin m, defined as $\mathcal{L}_{\text{proto}} = \frac{1}{|\mathcal{B}|} \sum_{i=1}^{|\mathcal{B}|} \max(0, \cos(\mathbf{z}_i, \mathbf{p}_{1-c_i}) - \cos(\mathbf{z}_i, \mathbf{p}_{c_i}) + m)$.

2.2 Weakly-Supervised Prototype FM Model Training

Let $\mathbb{X} = \{x_i\}_{i=1}^{N_s}$ denote the source domain (undistorted DWI) and $\mathbb{Y} = \{y_j\}_{j=1}^{N_t}$ denote the target domain (distorted DWI), where N_s and N_t represent the number of samples in the source and target training sets, respectively. Since our ultimate clinical objective is distortion correction (forward direction, $\mathbb{Y} \to \mathbb{X}$), we formulate the distortion simulation as a reverse generation process ($\mathbb{X} \to \mathbb{Y}$). The objective of FM is to learn a time-dependent velocity field $v_\theta(\cdot, t)$ that approximates the true conditional flow at any time $t \in [0, 1]$. During training, mini-batches of samples are drawn independently from $\mathbb{X}$ and $\mathbb{Y}$, and a continuous time step t is sampled from the uniform distribution $\mathcal{U}(0, 1)$. We employ a minibatch optimal transport (OT) plan, where source and target samples within each mini-batch are coupled to minimize the overall transport cost, as detailed in [21]. Given an OT-coupled pair (x_i, y_j) and a time step t, the intermediate state is defined via linear interpolation as $\hat{y}_{i,j}^t = (1-t)x_i + ty_j$, with the ground-truth velocity field $u_{i,j} = y_j - x_i$. The FM network takes $\hat{y}_{i,j}^t$ and t as inputs to get the velocity field prediction $\hat{v}_\theta(\hat{y}_{i,j}^t, t)$. The training objective minimizes the mean squared error between the predicted and ground-truth velocity fields as $\mathcal{L}_{\text{FM}} = \mathbb{E}_{x_i, y_j, t}\left[\|\hat{v}_\theta(\hat{y}_{i,j}^t, t) - u_{i,j}\|_2^2\right]$.

In addition to learning the velocity field in the image space, we introduce a weak supervision strategy that explicitly guides the generated features toward the distorted feature manifold in the feature space. This guidance is intended to promote clinically observed distortion-like patterns, rather than to provide subject-specific biophysical deformation modelling. The FM network generates a trajectory $\{x_i^k\}_{k=0}^K$ starting from the initial state x_i^0 (where $x_i^0 = x_i$) with discrete time steps $t_k = k/K$. We define a threshold step k_0. For early generation

steps ($k < k_0$), we apply no quality constraints, as the early generated images are semantically unstable and noisy. For later steps ($k \geq k_0$), the generated image x_i^k is passed through the IQA encoder at each step to extract its feature $\mathbf{z}_i^k$. We calculate a margin ranking loss between this feature and the quality prototypes to formulate the guidance loss. Crucially, we introduce a time-dependent weight w_k to allocate stronger supervision as the generated image progressively reveals structural details, preventing the loss from misguiding the trajectory during the early, noisy stages. The prototype guidance loss is defined as:

$$\mathcal{L}_{\mathrm{PG}} = \sum_{k=k_0}^{K} w_k \cdot \max\left(0, \cos(\mathbf{z}_i^k, \mathbf{p}_0) - \cos(\mathbf{z}_i^k, \mathbf{p}_1) + m\right), \tag{2}$$

where $w_k = [(k - k_0)/(K - k_0)]^2$ is the quadratic time-dependent weight and m is the constant margin. To maintain memory efficiency while optimizing this continuous trajectory, we employ gradient accumulation, updating the model parameters at intervals of $n_{\Delta k}$ steps. The total training objective for the FM network is defined as $\mathcal{L}_{\mathrm{total}} = \mathcal{L}_{\mathrm{FM}} + \lambda_{\mathrm{PG}}\mathcal{L}_{\mathrm{PG}}$, where λ_{PG} is a hyper-parameter that balances the prototype guidance loss.

2.3 Inference and Second IQT Model Training.

During inference, the FM network generates a distorted image iteratively starting from an undistorted source image x_i. For each discrete time step $k = 1, \ldots, K$ (corresponding to $t_k = k/K$), the state x_i^k is updated from the previous state x_i^{k-1} using the predicted velocity field $\hat{v}_\theta(x_i^{k-1}, t_{k-1})$, as $x_i^k = x_i^{k-1} + \hat{v}_\theta(x_i^{k-1}, t_{k-1})\Delta t$, where $\Delta t = 1/K$. By applying this generation process to the undistorted dataset, we synthesize a paired dataset of undistorted and generated distorted images, denoted as $\{(x_i, \hat{y}_i)\}_{i=1}^{N_s}$, where $\hat{y}_i = x_i^K$ represents the final generated state. Utilizing these synthetic pairs, we train a supervised second IQT network denoted as G_ψ to map the distorted DWI $\hat{y}_i$ back to the undistorted one x_i. The network is trained to minimize the reconstruction loss between the predicted undistorted image and the ground truth as $\mathcal{L}_{\mathrm{corr}} = \mathbb{E}_{x_i, \hat{y}_i}[\|x_i - G_\psi(\hat{y}_i)\|_1]$, where $\|\cdot\|_1$ denotes the L_1 norm to preserve high-frequency textural details, which are critical for downstream clinical evaluations.

3 Experiments and Results

Implementation Details. The FM and the second IQT models both employ 3D U-Net backbones [5] featuring a three-level symmetric encoder-decoder structure. Specifically, we set the channel multipliers to (1, 2, 4) with two residual blocks per level and integrate a self-attention module at the second level. The IQA encoder is implemented using a 3D ResNet-18 architecture [8], which yields a 512-dimensional feature vector via global average pooling. For encoder pretraining, the temperature parameter τ and contrastive loss weight λ_{con} are set to 0.1 and 0.5, respectively. In the quality prototype learning and FM training

stages, the margin m is set to 0.2. FM training utilizes $K = 20$ total inference steps, with a threshold step $k_0 = 10$, an interval $n_{\Delta k} = 3$, and a guidance loss weight $\lambda_{\mathrm{PG}} = 1.0$. All models are optimized for 100 epochs using Adam with a learning rate of 1×10^{-4} and a batch size of 4. Experiments were conducted on an institutional GPU cluster with NVIDIA GPUs.

Datasets. A private dataset [23], referred to as 'Dataset 1', consisting of 1027 prostate DWI scans with multiple high b-values (≥ 1400 s/mm^2) from 851 patients, is used for training and internal testing. These cases are categorized into undistorted (790 scans) and distorted (237 scans). Dataset 1 is randomly partitioned into training and test sets with an 8 : 2 ratio at the patient level. The PRIME dataset [2] is used for external evaluation, comprising 483 cases, including 28 labeled as distorted by expert radiologists. The distortion labels were assigned independently under routine clinical conditions by 37 clinical experts with 5-20 years of experience, based on geometric consistency between T2WI and the imaging-coordinate-aligned DWI. Additionally, it includes expert-labeled PI-RADS [22] ($227 \geq 4$, $256 < 4$) and Gleason scores [7] ($209 \geq 3{+}4$, $274 < 3 + 4$) for downstream clinical validation. All volumes are preprocessed using SimpleITK [14] python library, resampled to a spatial resolution of $[1.0, 1.0, 1.5]$ mm, followed by center-cropping to $[128, 128, 32]$. Voxel intensities are normalized to the range $[-1, 1]$ via min-max scaling.

Downstream Evaluation Protocol. We utilize the ProFound foundation model[9] to evaluate the clinical value of generated images through downstream tasks. The model takes three modalities as input: high b-value DWI, axial T2WI, and the apparent diffusion coefficient (ADC) map. The ProFound model is fine-tuned on 386 undistorted PRIME cases and tested on the remaining 60 scans, which include 30 randomly selected samples per class for each task. During fine-tuning, affine transformations are applied jointly to all three modalities to preserve spatial alignment, including random rotation (up to 15°) and translation (up to 0.15 mm). For comparison and ablation study evaluation, we replace the original undistorted DWI with generated distorted DWI to measure the performance degradation. For correction evaluation, we use 37 real-world distorted volumes from the PRIME dataset and replace them with generated undistorted DWI. Performance is quantified using accuracy (ACC) and area under the curve (AUC). We follow the open-source ProFound fine-tuning code and use the same hyperparameter settings.

Comparison Study We compare the proposed weakly-supervised prototype FM method with CycleGAN (C-GAN) [17], UNIT-DDPM (U-DDPM) [18] using the SDEdit strategy [15], and optimal transport flow matching (OT-FM) [24] on the test sets. All models share the same 3D U-Net backbone. U-DDPM, OT-FM, and our method use 20 inference steps. Results are summarized in Table 1, and compared in Fig. 2 (a). As shown in Table 1, CycleGAN achieves relatively low classification scores, and primarily performs intensity and style transfer rather than learning geometric deformation as observed in Fig. 2 (a). U-DDPM with SDEdit fails to simulate geometric distortions. It maintains PI-RADS perfor-

[9] `https://github.com/pipiwang/ProFound`

Table 1. Quantitative evaluation on the PRIME dataset. Comparison and Ablation Study: Performance of the downstream classifier when the original undistorted DWI is replaced with generated distorted DWI. Lower ACC and AUC indicate more effective distortion simulation. Distortion Correction: Evaluation of diagnostic correction on real distorted images. For unpaired baselines (C-GAN, U-DDPM, OT-FM) and Ours (FG), correction is performed via forward generation (distorted-to-undistorted). Ours (UNet) refers to a supervised U-Net trained on the synthetic paired data produced by our distortion synthesis FM model. Higher ACC and AUC indicate superior correction. Ref. Un and Ref. Dis denote reference performance on original undistorted and distorted test images, respectively. All metrics are reported as mean ± standard deviation over five runs. For diffusion and FM-based methods, results with total inference steps of 20.

		PI-RADS Score		Gleason Score	
Category	Method	ACC (%)	AUC (%)	ACC (%)	AUC (%)
Comparison Study	Ref. Un	59.33 ± 3.83	67.22 ± 3.98	57.33 ± 2.98	62.89 ± 2.11
	C-GAN	50.33 ± 4.41	56.02 ± 4.42	54.67 ± 4.14	47.24 ± 4.12
	U-DDPM	60.00 ± 4.08	57.11 ± 4.25	52.67 ± 2.88	52.91 ± 3.49
	OT-FM	57.67 ± 4.14	61.44 ± 5.77	56.33 ± 4.82	54.67 ± 5.56
	Ours	49.67 ± 2.11	53.78 ± 3.45	46.67 ± 3.54	45.82 ± 3.21
Ablation Study	w.o. PG	57.67 ± 4.14	61.44 ± 5.77	56.33 ± 4.82	54.67 ± 5.56
	PG → CE	54.33 ± 2.36	57.67 ± 3.12	58.33 ± 4.63	59.33 ± 4.23
	Ours	49.67 ± 2.11	53.78 ± 3.45	46.67 ± 3.54	45.82 ± 3.21
Distortion Correction	Ref. Dis	42.86 ± 6.68	46.78 ± 8.76	48.57 ± 9.65	48.65 ± 7.71
	C-GAN	48.57 ± 6.96	49.56 ± 6.44	55.00 ± 6.49	50.41 ± 9.63
	U-DDPM	41.43± 8.22	42.78 ± 8.35	48.57 ± 8.22	50.53 ± 9.50
	OT-FM	45.71 ± 8.53	48.11 ± 8.97	43.57 ± 8.22	51.93 ± 9.50
	Ours (FG)	44.29 ± 5.98	47.33 ± 6.13	46.43 ± 7.99	46.08 ± 4.97
	Ours (UNet)	54.29 ± 4.66	53.98 ± 3.02	56.43 ± 5.42	51.11 ± 4.85

mance by preserving anatomical structures, but significantly degrades Gleason scoring. This is perhaps due to the denoising process smoothing out fine textures and intensities essential for grading, reflecting textural degradation rather than realistic distortion. For FM-based methods, the OT-FM yields higher ACC and AUC scores. Visually, OT-FM primarily introduces blurring effects through pixel-wise averaging and fails to simulate complex structural degradations. Without a defined target in the feature space, the standard flow trajectory tends to converge to a conservative mean representation of the target domain, failing to reach the regions of high distortion intensity within the manifold. Our method is able to push velocity field toward the core distortion features, producing PI-RADS-and-Gleason-impacting boundary deformation.

Ablation Study. We evaluate the weakly supervised prototype guidance (PG) mechanism by comparing our model with: (1) a version without the PG constraint (w.o. PG), which is equivalent to OT-FM, and a version where the PG loss is replaced by a standard binary cross-entropy (BCE) classification loss (PG → CE). The BCE loss is computed using the outputs of the pre-trained

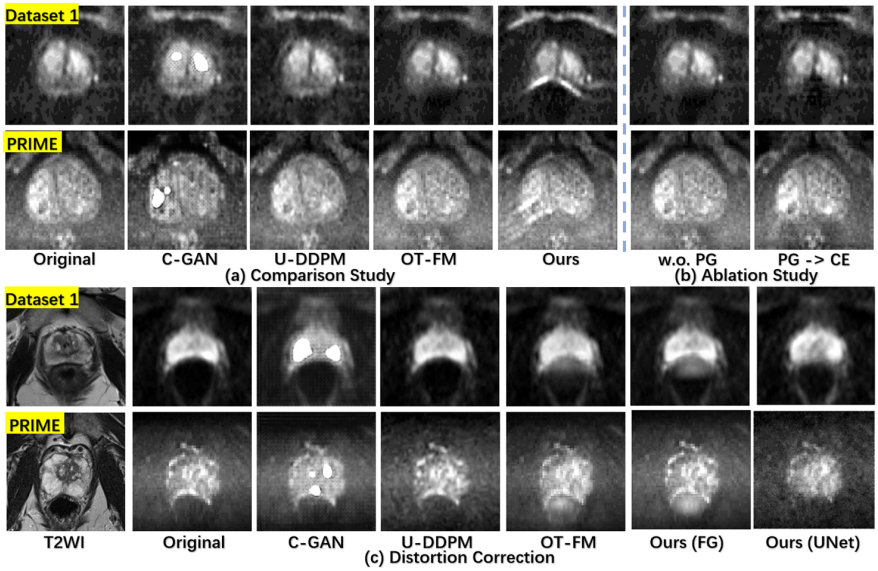

Fig. 2. Qualitative evaluation on the in-distribution and external test datasets. (a) Visual comparison of distortion generation with baseline methods. (b) Ablation study on the weakly-supervised PG mechanism. (c) Visualization of baseline methods for direct forward distortion correction and the U-Net model trained on synthetic paired data (Ours (UNet)).

classification head in Section 2.1 against the label for distorted images ($c = 1$). As shown in Table 1 and Fig. 2 (b), both removing and substituting the PG constraint lead to a reduction in distortion intensity. The CE approach is less effective than PG because the BCE loss is primarily optimized by pushing features across the decision boundary. Once the sample is classified as distorted, the gradient signal diminishes, often leaving the generated image near the boundary with insufficient geometric deformation. In contrast, by maximizing the cosine similarity to this cluster center, the PG constraint provides a continuous and stable directional force that pulls the generation trajectory deeper into the distorted distribution, resulting in more clinically realistic distortion.

Supervised Distortion Correction via Synthetic Paired Data. As shown in Table 1, direct forward generation (FG) ($\mathbb{Y} \to \mathbb{X}$) using the proposed approach (Ours (FG)) and other unpaired baselines yield poor downstream performance. Qualitatively, in Fig. 2(c), the Ours (FG) merely fills geometric voids with unrealistic intensities. On the other hand, as presented in Table 1, the U-Net trained on our synthetic data produced via reverse generation (Ours (UNet)) outperforms all baseline methods attempting direct correction on PI-RADS score task. However, we observe that the U-Net outputs tend to over-smooth high-frequency textures, resulting in blurrier images compared to the original undistorted references. This loss of fine-grained detail limits the model's performance

on the Gleason score task. We expect further development using generative modeling for this step may warrant improvement.

4 Conclusion

We proposed a novel and practical two-stage IQT framework for unpaired DWI distortion correction. Experiments show that generated images exhibit clinically relevant DWI degradation patterns with measurable impact on downstream diagnostic performance. Furthermore, using these synthetic paired samples, a simple supervised distortion correction model improves image quality in real-world scenarios. Our future work will focus on employing more powerful generative models (such as FM) for the supervised correction stage to enhance textural fidelity. Furthermore, we plan to incorporate corresponding T2WI as structural priors to provide stronger anatomical constraints during the correction process and explicitly quantify geometric alignment using T2WI-based overlap metrics.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Andersson, J.L., Skare, S., Ashburner, J.: How to correct susceptibility distortions in spin-echo echo-planar images: application to diffusion tensor imaging. Neuroimage **20**(2), 870–888 (2003)
2. Asif, A., Nathan, A., Ng, A., Khetrapal, P., Chan, V.W.S., Giganti, F., Allen, C., Freeman, A., Punwani, S., Lorgelly, P., et al.: Comparing biparametric to multiparametric mri in the diagnosis of clinically significant prostate cancer in biopsy-naive men (prime): a prospective, international, multicentre, non-inferiority within-patient, diagnostic yield trial protocol. BMJ open **13**(4), e070280 (2023)
3. Bian, Z., Shao, M., Carass, A., Prince, J.L.: Drdisco: Deep registration for distortion correction of diffusion mri with single phase-encoding. In: Medical Imaging 2023: Image Processing. vol. 12464, pp. 300–304. SPIE (2023)
4. Chien, N., Cho, Y.H., Wang, M.Y., Tsai, L.W., Yeh, C.Y., Li, C.W., Lan, P., Wang, X., Liu, K.L., Chang, Y.C.: Deep learning based multi-shot breast diffusion mri: Improving imaging quality and reduced distortion. European Journal of Radiology p. 112419 (2025)
5. Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O.: 3d u-net: learning dense volumetric segmentation from sparse annotation. In: International conference on medical image computing and computer-assisted intervention. pp. 424–432. Springer (2016)
6. Donato Jr, F., Costa, D.N., Yuan, Q., Rofsky, N.M., Lenkinski, R.E., Pedrosa, I.: Geometric distortion in diffusion-weighted mr imaging of the prostate—contributing factors and strategies for improvement. Academic radiology **21**(6), 817–823 (2014)
7. Epstein, J.I.: An update of the gleason grading system. The Journal of urology **183**(2), 433–440 (2010)

8. Hara, K., Kataoka, H., Satoh, Y.: Can spatiotemporal 3d cnns retrace the history of 2d cnns and imagenet? In: Proceedings of the IEEE conference on Computer Vision and Pattern Recognition. pp. 6546–6555 (2018)
9. Hu, Z., Wang, Y., Zhang, Z., Zhang, J., Zhang, H., Guo, C., Sun, Y., Guo, H.: Distortion correction of single-shot epi enabled by deep-learning. Neuroimage **221**, 117170 (2020)
10. Johnson, J.W.: Generative adversarial networks in medical imaging. In: State of the Art in Neural Networks and their Applications, pp. 271–278. Elsevier (2021)
11. Lawrence, E.M., Zhang, Y., Starekova, J., Wang, Z., Pirasteh, A., Wells, S.A., Hernando, D.: Reduced field-of-view and multi-shot dwi acquisition techniques: Prospective evaluation of image quality and distortion reduction in prostate cancer imaging. Magnetic resonance imaging **93**, 108–114 (2022)
12. Liao, P., Zhang, J., Zeng, K., Yang, Y., Cai, S., Guo, G., Cai, C.: Referenceless distortion correction of gradient-echo echo-planar imaging under inhomogeneous magnetic fields based on a deep convolutional neural network. Computers in biology and medicine **100**, 230–238 (2018)
13. Lipman, Y., Chen, R.T., Ben-Hamu, H., Nickel, M., Le, M.: Flow matching for generative modeling. arXiv preprint arXiv:2210.02747 (2022)
14. Lowekamp, B.C., Chen, D.T., Ibáñez, L., Blezek, D.: The design of simpleitk. Frontiers in neuroinformatics **7**, 45 (2013)
15. Meng, C., He, Y., Song, Y., Song, J., Wu, J., Zhu, J.Y., Ermon, S.: Sdedit: Guided image synthesis and editing with stochastic differential equations. arXiv preprint arXiv:2108.01073 (2021)
16. Rakow-Penner, R.A., White, N.S., Margolis, D.J., Parsons, J.K., Schenker-Ahmed, N., Kuperman, J.M., Bartsch, H., Choi, H.W., Bradley, W.G., Shabaik, A., et al.: Prostate diffusion imaging with distortion correction. Magnetic resonance imaging **33**(9), 1178–1181 (2015)
17. Sandfort, V., Yan, K., Pickhardt, P.J., Summers, R.M.: Data augmentation using generative adversarial networks (cyclegan) to improve generalizability in ct segmentation tasks. Scientific reports **9**(1), 16884 (2019)
18. Sasaki, H., Willcocks, C.G., Breckon, T.P.: Unit-ddpm: Unpaired image translation with denoising diffusion probabilistic models. arXiv preprint arXiv:2104.05358 (2021)
19. Song, J., Meng, C., Ermon, S.: Denoising diffusion implicit models. arXiv preprint arXiv:2010.02502 (2020)
20. Tivnan, M., Yoon, S., Chen, Z., Li, X., Wu, D., Li, Q.: Hallucination index: An image quality metric for generative reconstruction models. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 449–458. Springer (2024)
21. Tong, A., Fatras, K., Malkin, N., Huguet, G., Zhang, Y., Rector-Brooks, J., Wolf, G., Bengio, Y.: Improving and generalizing flow-based generative models with minibatch optimal transport. arXiv preprint arXiv:2302.00482 (2023)
22. Turkbey, B., Rosenkrantz, A.B., Haider, M.A., Padhani, A.R., Villeirs, G., Macura, K.J., Tempany, C.M., Choyke, P.L., Cornud, F., Margolis, D.J., et al.: Prostate imaging reporting and data system version 2.1: 2019 update of prostate imaging reporting and data system version 2. European urology **76**(3), 340–351 (2019)
23. Yan, W., Chiu, B., Shen, Z., Yang, Q., Syer, T., Min, Z., Punwani, S., Emberton, M., Atkinson, D., Barratt, D.C., et al.: Combiner and hypercombiner networks: Rules to combine multimodality mr images for prostate cancer localisation. Medical Image Analysis **91**, 103030 (2024)

24. Yazdani, M., Medghalchi, Y., Ashrafian, P., Hacihaliloglu, I., Shahriari, D.: Flow matching for medical image synthesis: Bridging the gap between speed and quality. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 216–226. Springer (2025)
25. Zou, S., Huang, Y., Yi, R., Zhu, C., Xu, K.: Cyclediff: Cycle diffusion models for unpaired image-to-image translation. arXiv preprint arXiv:2508.06625 (2025)