Synthesizing High-*b*-Value Diffusion-weighted Imaging of the Prostate Using Generative Adversarial Networks

Lei Hu, PhD • Da-wei Zhou, MS • Yun-fei Zha, MD* • Liang Li, MD • Huan He, MS • Wen-hao Xu, MS • Li Qian • Yi-kun Zhang • Cai-xia Fu, MS • Hui Hu, PhD • Jun-gong Zhao, PhD*

From the Department of Diagnostic and Interventional Radiology, Shanghai Jiao Tong University Affiliated Sixth People's Hospital, 600 Yi Shan Road, Shanghai 200233, China (L.H., W.H.X., J.G.Z.); State Key Laboratory of Integrated Services Networks, School of Telecommunications Engineering, Xidian University, Xi'an, China (D.W.Z.); Department of Radiology, Renmin Hospital, Wuhan University, Wuhan, China (Y.F.Z., L.L., H. He, L.Q., Y.K.Z.); MR Application Development, Siemens Shenzhen MR, Shenzhen, China (C.X.F.); and Department of Radiology, The Affiliated Renmin Hospital of Jiangsu University, Zhenjiang, China (H. Hu). Received September 29, 2020; revision requested November 6; revision received April 11, 2021; accepted May 18. **Address correspondence to** J.G.Z. (e-mail: zhaojungongradio@hotmail.com).

*Y.F.Z. and J.G.Z. are co-senior authors.

Authors declared no funding for this work.

Conflicts of interest are listed at the end of this article.

Radiology: Artificial Intelligence 2021; 3(5):e200237 • <https://doi.org/10.1148/ryai.2021200237> • Content codes: **AI** **GU**

Purpose: To develop and evaluate a diffusion-weighted imaging (DWI) deep learning framework based on the generative adversarial network (GAN) to generate synthetic high-*b*-value ($b = 1500$ sec/mm²) DWI (SYN_{b1500}) sets from acquired standard-*b*-value ($b = 800$ sec/mm²) DWI (ACQ_{b800}) and acquired standard-*b*-value ($b = 1000$ sec/mm²) DWI (ACQ_{b1000}) sets.

Materials and Methods: This retrospective multicenter study included 395 patients who underwent prostate multiparametric MRI. This cohort was split into internal training (96 patients) and external testing (299 patients) datasets. To create SYN_{b1500} sets from ACQ_{b800} and ACQ_{b1000} sets, a deep learning model based on GAN (M_0) was developed by using the internal dataset. M_0 was trained and compared with a conventional model based on the cycle GAN (M_{cyc}). M_0 was further optimized by using denoising and edge-enhancement techniques (optimized version of the M_0 [Opt- M_0]). The SYN_{b1500} sets were synthesized by using the M_0 and the Opt- M_0 were synthesized by using ACQ_{b800} and ACQ_{b1000} sets from the external testing dataset. For comparison, traditional calculated ($b = 1500$ sec/mm²) DWI (CAL_{b1500}) sets were also obtained. Reader ratings for image quality and prostate cancer detection were performed on the acquired high-*b*-value ($b = 1500$ sec/mm²) DWI (ACQ_{b1500}), CAL_{b1500}, and SYN_{b1500} sets and the SYN_{b1500} set generated by the Opt- M_0 (Opt-SYN_{b1500}). Wilcoxon signed rank tests were used to compare the readers' scores. A multiple-reader multiple-case receiver operating characteristic curve was used to compare the diagnostic utility of each DWI set.

Results: When compared with the M_{cyc} , the M_0 yielded a lower mean squared difference and higher mean scores for the peak signal-to-noise ratio, structural similarity, and feature similarity ($P < .001$ for all). Opt-SYN_{b1500} resulted in significantly better image quality ($P \leq .001$ for all) and a higher mean area under the curve than ACQ_{b1500} and CAL_{b1500} ($P \leq .042$ for all).

Conclusion: A deep learning framework based on GAN is a promising method to synthesize realistic high-*b*-value DWI sets with good image quality and accuracy in prostate cancer detection.

Supplemental material is available for this article.

© RSNA, 2021

Diffusion-weighted imaging (DWI) is a key component of multiparametric MRI in the detection and characterization of prostate cancer (1). Compared with standard-*b*-value (800–1000 sec/mm²) DWI, high-*b*-value (>1000 sec/mm²) DWI enables detection of small changes in diffusion because of better contrast between malignant and background tissue (2). According to the Prostate Imaging Reporting and Data System version 2, higher *b* values (≥ 1400 to 2000 sec/mm²) are considered to result in better image quality for differentiating between these tissue types (3). However, there is currently no widely accepted optimal high *b* value because such a *b* value depends on the magnetic field strength, software, and hardware.

High-*b*-value DWI is technically challenging to obtain in clinical practice; acquired DWI frequently suffers from poor in-plane spatial resolution, distortion, and artifacts. Moreover, an increase in the *b* value is generally accompanied by a significant reduction in the signal-to-noise ratio and requires a prolonged

acquisition time (4–6). To overcome these shortcomings, calculated DWI and zoomed field of view (FOV) DWI technology have been introduced. In calculated DWI, high-*b*-value images are calculated based on acquired DWI sets with lower *b* values through voxel-by-voxel extrapolation of the fitted signal decay curves (7). Calculated DWI is potentially less prone to producing artifacts because the longer echo times required to accommodate the strong gradient pulses needed for high-*b*-value acquisitions are avoided, but the image quality of calculated DWI with high *b* values still relies on the image quality of the DWI with lower *b* values. Zoomed FOV DWI excites a small FOV that covers only the region of interest and consequently increases the in-plane resolution and reduces the geometric distortion artifacts (8,9). However, zoomed FOV DWI techniques depend on the gradient performance, radiofrequency design, and software platform (5,10–14), and their clinical application remains limited.

Abbreviations

ACQ_{b1000} = acquired standard-*b*-value (*b* = 1000 sec/mm²) DWI, ACQ_{b1500} = acquired high-*b*-value (*b* = 1500 sec/mm²) DWI, ACQ_{b50} = acquired (*b* = 50 sec/mm²) DWI, ACQ_{b800} = acquired standard-*b*-value (*b* = 800 sec/mm²) DWI, ADC = apparent diffusion coefficient, AUC = area under the ROC curve, CAL_{b1500} = calculated (*b* = 1500 sec/mm²) DWI, CS = clinical site, DWI = diffusion-weighted imaging, FOV = field of view, FSIM = feature similarity, GAN = generative adversarial network, M₀ = deep learning model based on the GAN, M_{cyc} = conventional model based on the cycle-GAN, MSD = mean squared difference, Opt-M₀ = optimized version of the M₀, Opt-SYN_{b1500} = SYN_{b1500} set generated by the optimized M₀, PSNR = peak signal-to-noise ratio, ROC = receiver operating characteristic, SSIM = structural similarity, SYN_{b1500} = synthesized high-*b*-value (*b* = 1500 sec/mm²) DWI

Summary

Deep learning based on the generative adversarial network is a promising method to generate high-*b*-value diffusion-weighted images with good quality and detection capability for prostate cancer.

Key Points

- Compared with diffusion-weighted imaging (DWI) sets created by using the cycle generative adversarial network, the synthesized high-*b*-value (*b* = 1500 sec/mm²) DWI (SYN_{b1500}) sets created by the deep learning framework achieved a lower mean squared difference (0.002 vs 0.004) but achieved higher peak signal-to-noise ratio (0.819 vs 0.793), structural similarity (0.901 vs 0.873), and feature similarity (27.40 vs 24.67) scores (*P* < .001 for all).
- Compared with the acquired high-*b*-value (*b* = 1500 sec/mm²) DWI (ACQ_{b1500}) set, the SYN_{b1500} sets demonstrated better imaging features, such as the suppression of normal or benign prostate tissue (reader 1: 2.3 vs 4.2; reader 2: 2.8 vs 4.1), absence of ghosting (reader 1: 2.3 vs 2.9; reader 2: 2.5 vs 3.2), and improved overall image quality (reader 1: 2.3 vs 4.3; reader 2: 2.4 vs 4.4), with ratings being based on a five-point Likert scale (*P* < .001 for all comparisons).
- The use of SYN_{b1500} sets from the framework allowed for a better mean area under the curve across readers than the use of ACQ_{b1500} sets (0.79 vs 0.89; *P* < .001) or calculated (*b* = 1500 sec/mm²) DWI sets (0.85 vs 0.89; *P* < .05) and demonstrated the most value among less-experienced radiologists.

Keywords

Prostate Cancer, Abdomen/GI, Diffusion-weighted Imaging, Deep Learning Framework, High *b* Value, Generative Adversarial Networks

The development of generative adversarial networks (GANs) has provided momentum in the field of medical image generation and translation (15). A GAN is a generative model consisting of a generator (learns a map from random noise to synthesize a target image) and a discriminator (distinguishes between the generated and real images) (16). Because of their ability to simulate data distribution and synthesize data with unprecedented realism, GANs have been used to solve challenging medical image analysis problems, including image reconstruction, segmentation, denoising, detection, and classification (15). GANs have also been proposed as a generic solution for cross-modality image synthesis, such as producing PET or MRI sets from CT sets or synthesizing CT sets from MRI sets (17,18).

We hypothesized that a deep learning algorithm based on the GAN may be a promising technique to generate synthetic high-*b*-value DWI sets. This study aimed to establish and evaluate a

semisupervised learning framework based on GAN and multi-level semantic features to generate synthetic high *b*-value DWI from acquired standard *b*-value DWI.

Materials and Methods

Patient Data

This retrospective multicenter study was approved by the local ethics committee at our institution. Informed consent was obtained from all patients. We collected data from patients who underwent 3-T multiparametric prostate MRI and underwent a subsequent targeted MRI–transrectal US fusion biopsy because they were suspected of having prostate cancer between November 2018 and December 2019. Exclusion criteria were as follows: a history of treatment for prostate cancer (ie, hormone therapy, radiation therapy, cryotherapy, or surgery), an interval of more than 2 weeks between MRI and the biopsy procedure, and unavailability of the final prostate cancer diagnosis.

A total of 395 patients from six clinical sites (CSs, referred to as CS1–CS6) were included in our study (Fig 1) and were further split into an internal training set (from CS1) and an external testing set (from CS2 to CS6). The CS1 dataset consisted of 96 patients examined with zoomed FOV DWI. Because the images acquired by using zoomed FOV DWI had the highest resolution, the fewest artifacts, and the least distortion, zoomed FOV acquired with standard-*b*-value (*b* = 1000 sec/mm²) DWI (ACQ_{b1000}) and zoomed FOV acquired with high-*b*-value (*b* = 1500 sec/mm²) DWI (ACQ_{b1500}) sets from CS1 were used to develop, train, and perform a pretest of our proposed deep learning model based on the GAN (called M₀).

The test dataset from CS2–CS6 consisted of the remaining 299 patients examined with full FOV DWI in a multicenter setting. To assess the performance of images synthesized by our proposed framework, we input DWI with *b* = 800 sec/mm² (ACQ_{b800}; from CS2 and CS3) and ACQ_{b1000} (from CS4, CS5, and CS6) into our framework to generate synthetic DWI with *b* = 1500 sec/mm² (SYN_{b1500}) sets. The SYN_{b1500} sets were then compared with ACQ_{b1500} and calculated (*b* = 1500 sec/mm²) DWI (CAL_{b1500}) sets in terms of image quality and the accuracy of prostate cancer depiction.

Detailed clinical and tumor characteristics of patients, including age, tumor locations, Prostate Imaging Reporting and Data System scores, and Gleason scores are shown in Table 1 and Table E1 (supplement). The parameters and equipment information of the DWI sequences performed for each dataset are shown in Table 2.

Semisupervised Deep Learning Framework

A diagram of the M₀ is shown in Figure 2. The model consists of three modules: a synthesis adversarial module, a feature representation–consistency verification module, and a potential feature representation module. The synthesis adversarial module was used to determine whether the SYN_{b1500} images belonged to the target domain (ie, the actual ACQ_{b1500} images). The feature representation–consistency verification module was used to evaluate the loss

Figure 1: Diagram for inclusion of patients in the study. CS = clinical site, PCa = prostate cancer.

Table 1: Clinical Characteristics of the Patient Cohort

Characteristic	Without Cancer	With Cancer
No. of patients	195	200
Age (y)	69 (52–83)	70 (50–83)
Position		
Peripheral zone	64	117
Transition zone	131	83
PI-RADS score		
1	0	0
2	73	13
3	60	13
4	41	92
5	21	82
Gleason score		
6	NA	25
7	NA	117
8	NA	33
9	NA	25

Note.—Unless otherwise indicated, data are numbers of patients. PI-RADS = Prostate Imaging Reporting and Data System, NA = not applicable.

between the SYN_{b1500} images and the actual ACQ_{b1000} images that were input into the framework in the semantic structure. The potential feature representation module composes a potential feature representation to supervise the generator network with the potential feature representation of the actual ACQ_{b1000} images. Details about each module can be found in Appendix E1 (supplement).

Model Training, Selection, and Optimization

To better understand the capability of our proposed framework over other existing deep learning techniques that can be used for synthesizing a high- b -value DWI set, we implemented a conventional model based on the cycle-GAN (M_{cyc}) as a comparison.

Before model training, the front and last sections that did not cover the prostate were removed manually. The most representative DWI map sections were selected, and the number of selected diffusion-weighted images for each patient ranged from five to 15. The CS1 dataset included 832 matched ACQ_{b1000} and ACQ_{b1500} pairs, of which 300 pairs were selected randomly as a reference set and were used to train the discrimination network and generated potential feature representations. The ACQ_{b1000} images were randomly partitioned into a training set (50 patients, 438 pairs) and an internal test set (46 patients, 394 pairs) for the M_0 and M_{cyc} . The zoomed FOV ACQ_{b1500} images were used as a benchmark for the SYN_{b1500} images to evaluate the performance

of the M_0 and M_{cyc} . The synthetic images generated by the M_0 and M_{cyc} were compared and assessed quantitatively by using the peak signal-to-noise ratio (PSNR), mean squared difference (MSD), structural similarity (SSIM), and feature similarity (FSIM). The MSD is the variance of the estimator under an unbiased estimator, and an MSD value closer

Table 2: Diffusion-weighted Imaging Sequence Parameters

Dataset	Dataset 0	Dataset 1	Dataset 2	Dataset 3	Dataset 4	Dataset 5
Manufacturer	Siemens	Siemens	Siemens	General Electric	General Electric	Siemens
Model	Skyra	Skyra	Verio	750 plus	750 w	Skyra
Field strength (T)	3	3	3	3	3	3
Field of view (mm)	196×109	306×380	306×380	300×240	300×210	380×281
Imaging matrix	100×100	134×166	134×166	64×128	64×128	178×178
Thickness (mm)	3	4	4	3	3	3
Distance fact (mm)	0.3	0.5	0.5	0	0.5	0.1
b values (sec/mm ²)	50, 1000, 1500	0, 800, 1500	0, 800, 1500	0, 800, 1000, 1500	0, 1000, 1500	50, 1000, 1500
Echo time (msec)	76	65	65	76.4	61.6	79
Repetition time (msec)	3800	3400	3400	5000	6000	4200
Bandwidth (Hertz per pixel)	1612	753	753	250	250	1872
Acquisition time (min:sec)	2:27	4:43	4:43	3:52	4:06	3:05

Note.—GE = General Electric, px = pixel.

to 0 indicates more similarity between synthesized images and actual acquired images. The PSNR is used to calculate the ratio between the maximum possible signal and the distorting noise. The SSIM and FSIM analyses were used to obtain the normalized mean values of SSIM and FSIM between the SYN_{b1500} images and reference images (ACQ_{b1500} images), respectively (19). Because the M_0 showed better performance for generating synthetic images than the M_{cv} , it was further optimized in terms of denoising and edge enhancement, creating an optimized version of the M_0 (Opt- M_0). To suppress noise and improve the sharpness of the synthesized images, a feed-forward denoising convolutional neural network (20) and the Laplacian operator were implemented. More details about image optimization are shown in Appendix E2 (supplement).

All algorithms were developed by using Python, version 3.5 (<https://www.python.org/downloads>) and MATLAB, version R2016b (<https://www.mathworks.com>) on an Ubuntu, version 16.04 system with an NVIDIA TITAN Xp GPU. A detailed description of the experimental setting is shown in Appendix E3 (supplement).

Figure 2: The overview of our framework, which includes three modules: the synthesis adversarial (SA) module, the feature representation–consistency verification (FRCV) module, and the potential feature representation (PFR) module. The SA module is used for judging whether a synthesized image (S) from the synthesized DWI (syn-DWI) set belongs to the target domain; the target random zoomed field of view reference image (R_z) is from the random zoomed field of view DWI with a b value of 1500 sec/mm² ($z\text{-DWI}_{b1500}$) set. The FRCV module is used to evaluate loss between S and the image from the standard- b -value DWI set (X). The PFR module comprises potential feature representation $P(X)$, which supervises the synthesis network and the potential feature representation of X together. The loss functions $\mathcal{L}^{sa}$, $\mathcal{L}^{frcv}$, and $\mathcal{L}^{pfr}$ supervise the training of the synthesis network. R denotes the reference set. DWI = diffusion-weighted imaging.

Performance Assessment

To test and verify the generalization and effectiveness of our framework, the external testing set from CS2 to CS6

Figure 3: A 69-year-old man with prostatic hyperplasia with an initial prostate-specific antigen level of 11.54 ng/mL (Prostate Imaging Reporting and Data System category 4). Arrows indicate the location of the lesion, and rectangles indicate normal prostate tissue. The lesion was determined to be benign through a targeted MRI–transrectal US fusion biopsy procedure. These images are from the (left) acquired standard- b -value ($b = 1000 \text{ sec/mm}^2$) diffusion-weighted imaging (DWI) (ACQ_{b1000}) set, (second from left) zoomed field of view high- b -value ($b = 1500 \text{ sec/mm}^2$) DWI (ACQ_{b1500}) set, (third from left) zoomed field of view synthesized high- b -value ($b = 1500 \text{ sec/mm}^2$) DWI (SYN_{b1500}) set generated by the deep learning model based on the generative adversarial network (GAN) (M_0), and (right) SYN_{b1500} set generated by the conventional model based on the cycle-GAN (M_{cyc}), respectively. This example shows that M_0 model provided images with a more accurate signal suppression of the benign lesion (especially on the right peripheral zone of the lesion) than the image from the M_{cyc} in reference to ACQ_{b1500} . For normal prostate tissue, the distribution of signal intensity on the M_0 image was very similar to that on the ACQ_{b1500} image. The distribution of signal intensity on the M_{cyc} image was very dissimilar to that of the ACQ_{b1500} image.

was used to compare the image quality and tumor detection among different DWI sets (the ACQ_{b1500} , CAL_{b1500} , and SYN_{b1500} sets [the latter of which was generated from M_0] and the SYN_{b1500} set generated from the Opt- M_0 [Opt- SYN_{b1500}]). For this purpose, the CAL_{b1500} images were calculated by using full FOV acquired DWI ($b = 50 \text{ sec/mm}^2$) (ACQ_{b50}) and ACQ_{b800} sets (from CS2 and CS3) and the full FOV ACQ_{b50} and ACQ_{b1000} sets (from CS4, CS5, and CS6) by using a standard monoexponential fit (21). The SYN_{b1500} and Opt- SYN_{b1500} images were generated by using the full FOV ACQ_{b800} and ACQ_{b1000} sets.

Image Quality Assessment

Two radiologists (L.Q. and Y.K.Z.) with about 3 years of experience in prostate MRI reporting independently evaluated the image quality of the four DWI sets (ACQ_{b1500} , CAL_{b1500} , SYN_{b1500} , and Opt- SYN_{b1500}). To minimize recognition bias, the radiologists evaluated these DWI sets in four different sessions at intervals of at least 2 weeks. In each session, only one of the four DWI sets for each patient was evaluated. They reviewed all DWI sets independently in a random order and rated the image quality of the DWI sets in terms of the suppression of benign prostate tissue, anatomic distortion, artifacts, and overall image quality by using a five-point Likert scale, with 5 indicating the highest quality (22). All of them were blinded to all clinical details and biopsy results. Axial T2-weighted images were used as a reference for guiding anatomic localization of findings on the DWI sets (13).

Prostate Cancer Detection Assessment

Another four radiologists (two 3rd-year radiologic residents [L.H. and H. He] and two radiologists [H. Hu and J.G.Z.] with more than 20 years of experience in interpreting prostate MRI sets) who were blinded to all clinical details and

biopsy results independently performed the prostate cancer detection assessments. These readers reviewed images covering the whole prostate for each patient including the axial T2-weighted images and the apparent diffusion coefficient (ADC) maps and DWI sets at four different time points with an interval of 2 weeks. The readers gave a final score that indicated their confidence in the diagnosis of prostate cancer according to a five-point scale (1, definitely benign; 2, probably benign; 3, indeterminate; 4, probably cancer; and 5, definitely cancer) (23,24).

To evaluate the effects of the synthetic DWI sets on the derivation of the ADC map, ADC calculations were performed for all DWI sets with all b values by using the monoexponential model. The section of the ADC map containing the greatest lesion extent was selected for analysis; a circular region of interest of 0.5–0.8 cm² was drawn in the center of the lesion, excluding its edges, and the mean ADC value for each contour was recorded.

Statistical Analysis

Wilcoxon signed rank tests were performed by using MedCalc, version 19.2 (MedCalc Software) to compare DWI sets for image quality rating scores. The area under the receiver operating characteristic (ROC) curve (AUC) was used to quantify the predictive accuracy of using each ADC set to differentiate between benign and malignant prostate lesions. A multiple-reader multiple-case ROC method was used to compare the diagnostic capabilities of the ACQ_{b1500} , CAL_{b1500} , SYN_{b1500} , and Opt- SYN_{b1500} sets.

The AUC differences of different readers were tested by using DeLong tests. ROC curves were determined for each of the four readers by using the MRMCaov package for R Studio (<https://github.com/brian-j-smith/MRMCaov>). A two-sided P value of less than .05 was considered to indicate a significant difference.

Figure 4: Violin plots of distributions of the quantitative metrics between the deep learning model based on the generative adversarial network (GAN) (M_0) and the conventional model based on the cycle-GAN (M_{cyc}). FSIM = feature similarity, MSD = mean squared difference, PSNR = peak signal-to-noise ratio, SSIM = structural similarity.

Results

Patient Clinical Characteristics

The internal dataset included 96 patients with a mean age of 71 years $\pm$ 7 (standard deviation): 40 patients had benign lesions, and 56 patients had malignant lesions. The external testing datasets contained a total of 299 patients with a mean age of 68 years $\pm$ 8: 144 patients had malignant lesions, and 155 patients demonstrated no histologic evidence of cancer. There was no difference in mean patient age between those with and those without prostate cancer (69 years $\pm$ 8 vs 70 years $\pm$ 8, $P = .71$).

Performance of M_0 and M_{cyc} on the CS1 Dataset

As shown in Figure 3, images generated by using the M_0 demonstrated better performance in terms of visual perception than those generated by using the M_{cyc} . Neither M_0 images nor M_{cyc} images showed obvious anatomic distortion in terms of the overall prostate shape; however, some structural semantics and textural details from M_{cyc} -generated images were missing, which led to high levels of undesirable signals. Therefore, the shape and contrast of the lesions and the distribution of noise on M_{cyc} images were not visually similar to those on the real zoomed FOV ACQ_{b1500} images. Images generated by using the M_0 were much closer to the real zoomed FOV ACQ_{b1500} im-

ages, particularly in terms of the signal suppression of normal prostate tissue and benign lesions, than those generated by using the M_{cyc} .

The quantitative evaluation results are consistent with the visual evaluation. For quantitative evaluation parameters, the M_0 images had a lower MSD (0.002 vs 0.004) but had higher mean SSIM (0.819 vs 0.793), FSIM (0.901 vs 0.872), and PSNR (27.40 vs 24.76) scores ($P < .001$ for all comparisons). As shown in violin plots (Fig 4), the M_0 also performed better than the M_{cyc} in terms of the distribution, median, and interquartile range of the MSD, SSIM, FSIM, and PSNR scores.

Image Quality Assessment

As shown in Figure 5, among the four DWI sets (the ACQ_{b1500} , CAL_{b1500} , SYN_{b1500} , and Opt- SYN_{b1500} sets), the Opt- SYN_{b1500} demonstrated the clearest contours of the prostate and lesions, the highest suppression of normal prostate tissue, and the least amount of noise and the fewest artifacts as assessed by using visual perception.

Table 3 and Table E2 (supplement) show the results of the image quality assessments of the four DWI sequences from the two reviewers. The CAL_{b1500} , SYN_{b1500} , and Opt- SYN_{b1500} sets had better image quality than the ACQ_{b1500} set in terms of the suppression of benign prostate, anatomic distortion, artifacts, and overall image quality for both readers ($P < .001$ for all). The SYN_{b1500} and Opt- SYN_{b1500} sets had higher overall image

Figure 5: T2-weighted images (T2WI), apparent diffusion coefficient (ADC) images, images from the acquired standard- b -value ($b = 1000 \text{ sec/mm}^2$) diffusion-weighted imaging (DWI) (ACQ_{b1000}) set, images from the acquired high- b -value ($b = 1500 \text{ sec/mm}^2$) DWI (ACQ_{b1500}) set, images from the calculated ($b = 1500 \text{ sec/mm}^2$) DWI (CAL_{b1500}) set, images from the synthesized high- b -value ($b = 1500 \text{ sec/mm}^2$) DWI (SYN_{b1500}) set generated by the deep learning model based on the generative adversarial network (M_0), and images from the SYN_{b1500} set generated by the optimized M_0 (Opt-SYN $_{b1500}$) depicting four different patients. Patient A: A 76-year-old man with chronic prostatitis. Compared with ACQ_{b1500} and CAL_{b1500} images, the SYN_{b1500} and Opt-SYN $_{b1500}$ images' susceptibility artifacts due to air in the rectum are greatly reduced. Patient B: A 63-year-old man with prostatic hyperplasia. No restricted diffusion is found in any DWI sets, but the contour of the prostate is depicted more clearly within the SYN_{b1500} and Opt-SYN $_{b1500}$ images. Patient C: An 87-year-old man with prostatic cancer in the right transition zone (Gleason score, 4 + 3). Patient D: A 69-year-old man with prostatic cancer in the peripheral zone invading the rectum (Gleason score, 4 + 5). In patients C and D, all DWI sets show focally restricted diffusion; however, the contour of the prostate and lesion is depicted more clearly within the SYN_{b1500} and Opt-SYN $_{b1500}$ images.

quality ($P < .001$) and greater suppression of benign prostate ($P < .001$), as well as fewer ghost artifacts, than the CAL_{b1500} set for both reviewers ($P < .05$). There was no difference in the absence of distortion among the CAL_{b1500} , SYN_{b1500} , and Opt-SYN $_{b1500}$ sets for either reader ($P \geq .05$). Compared with the SYN_{b1500} set, the Opt-SYN $_{b1500}$ set had greater suppression of benign prostate and better overall image quality ($P < .001$), but no other comparison between these two DWI sets was significant for either reader ($P \geq .05$).

Tumor Detection

The five-point Likert score was used to compute the ROC curves (Fig 6). The AUCs for differentiating cancerous from noncancerous lesions by using the ACQ_{b1500} , CAL_{b1500} , SYN_{b1500} , and Opt-SYN $_{b1500}$ sets for each reader and the means across the four readers are summarized in Table 4. The mean AUCs of the different DWI sets were as follows: ACQ_{b1500} , 0.79; CAL_{b1500} , 0.85; SYN_{b1500} , 0.88; and Opt-SYN $_{b1500}$, 0.89. The mean Opt-SYN $_{b1500}$ AUC was higher than the mean ACQ_{b1500} ($P < .001$) and CAL_{b1500} ($P < .05$) AUCs. The AUC values between the ACQ_{b1500} and SYN_{b1500} sets were similar ($P = .12$); the AUC values for the SYN_{b1500} and Opt-SYN $_{b1500}$ sets were also similar ($P = .50$).

The AUCs for the different DWI sets for each individual reader were also compared (Fig 6). For the two resident radiologists, the AUC for the Opt-SYN $_{b1500}$ set exceeded those of the

ACQ_{b1500} and CAL_{b1500} sets ($P < .05$ for all both comparisons); the AUCs for the SYN_{b1500} and Opt-SYN $_{b1500}$ sets were similar. For the two more experienced radiologists, there were no differences in the AUCs among the four DWI sets.

The mean ADC values of cancerous versus noncancerous lesions derived from the ACQ_{b1500} , CAL_{b1500} , SYN_{b1500} , and Opt-SYN $_{b1500}$ sets are shown in Table E3 (supplement). The AUCs of the ADC values calculated for the image sets were as follows: ACQ_{b1500} , 0.751; CAL_{b1500} , 0.849; SYN_{b1500} , 0.881; and Opt-SYN $_{b1500}$, 0.921. As shown in Figure E4 (supplement), the AUC of the ADC values derived from the Opt-SYN $_{b1500}$ set exceeded the AUC of the ADC values derived from the ACQ_{b1500} , CAL_{b1500} , and SYN_{b1500} sets ($P \leq .05$ for all).

Discussion

The main result of our study is that we developed a deep learning framework based on GAN and multilevel semantic features to generate high- b -value DWI sets, which, in a reader study, had better image quality and tumor detection capability than acquired DWI and calculated DWI sets.

With the ability to conduct data-distribution simulation, GAN-based methods have been used to synthesize realistic-looking medical images, such as MRI sets of prostate lesions (25) or CT images of lung cancer nodules (26), that are difficult for experienced radiologists to distinguish from real acquired images. Generation of these images can be used to solve clinical problems,

such as data scarcity and class imbalance. GANs have also been used for image-to-image translation tasks. For example, Wolterink et al (27) used cycle-GAN to transform MRI sets into CT images to reduce the radiation exposure to patients. Similar to Wolterink et al (27), Chartas et al (28) generated pairs of cardiac MRI sets by using cardiac CT sections. The major limitation of these GANs is the lack of a guarantee that lesions are rightly preserved during image-to-image translation (29). To overcome this shortcoming, we designed a potential feature representation module and a feature representation-consistency verification module to improve the authenticity of the generated image on texture and structure, respectively. Compared with the cycle-GAN, which uses the distance between the reconstructed image and the original image at the pixel level to supervise the network, our model uses the distance between the synthesized image and the original image at the level of multiple depth features (30) in the feature representation-consistency verification module and pays more attention to the depth semantic features. Moreover, our model uses reference sets to generate pseudofeature representations for each input image (31) in the potential feature representation module, which can be regarded as references for synthesizing high- b -value images. In addition, in our synthesis network, skip connections (32) are used to input the active output of each convolutional layer into the corresponding deconvolutional layer again, and more feature information is used in this process. By comparing our framework with the cycle-GAN through subjective evaluation, we found that given zoomed FOV ACQ_{b1000} images as inputs, the M_0 has better performance in terms of synthesizing high- b -value DWI sets than the M_{cyc} . For objective evaluations, the M_0 model achieved a lower MSD and higher SSIM, FSIM, and PSNR scores than images from the M_{cyc} , which means that images generated by the M_0 were more similar to the actual zoomed FOV ACQ_{b1500} set. The results suggest that the module designed in our model is effective.

According to the validation of the multicenter dataset, the Opt-SYN_{b1500} set showed better image quality than the ACQ_{b1500} and CAL_{b1500} sets. In our study, a zoomed FOV ACQ_{b1500} set was chosen as the reference to train the synthesis discriminant network because it has been widely shown that using zoomed FOV DWI results in better image quality, including fewer artifacts, less blurring and distortion, and higher spatial resolution, than using full FOV DWI (22,33,34). The high similarity between the synthesized and actual zoomed FOV ACQ_{b1500} sets in terms of structural and textural features is the principal reason why image synthesis resulted in better image quality than both acquired and calculated DWI in these multicenter datasets. Another important reason is that we also used noise-reduction technology and edge-enhancement

Table 3: Comparison of Subjective Image Quality among Different DWI Sets

Parameter	ACQ _{b1500}	CAL _{b1500}	SYN _{b1500}	Opt-SYN _{b1500}
Suppression of benign prostate				
Reader 1	2.3 ± 0.8	3.0 ± 0.8	3.8 ± 0.7	4.2 ± 0.7
Reader 2	2.8 ± 1.0	3.2 ± 0.8	3.7 ± 0.8	4.1 ± 0.6
Absence of distortion				
Reader 1	3.1 ± 0.8	3.4 ± 0.8	3.4 ± 0.8	3.5 ± 0.7
Reader 2	3.3 ± 0.8	3.9 ± 0.8	3.9 ± 0.7	4.0 ± 0.8
Absence of ghosting				
Reader 1	2.3 ± 0.7	2.7 ± 0.9	2.9 ± 0.9	2.9 ± 0.9
Reader 2	2.5 ± 0.6	2.8 ± 0.7	3.1 ± 0.7	3.2 ± 0.6
Overall image quality				
Reader 1	2.3 ± 0.7	2.7 ± 0.9	2.9 ± 0.8	4.3 ± 0.7
Reader 2	2.4 ± 0.7	3.0 ± 0.8	3.3 ± 0.9	4.4 ± 0.6

Note.—Data are mean ± standard deviation. Image quality features were rated by each reader by using a five-point Likert scale (1 = poor, 2 = below average, 3 = average, 4 = above average, and 5 = excellent). ACQ_{b1500} = acquired high- b -value (b = 1500 sec/mm²) DWI, CAL_{b1500} = calculated (b = 1500 sec/mm²) DWI, DWI = diffusion-weighted imaging, Opt-SYN_{b1500} = SYN_{b1500} set generated by the optimized deep learning model based on the generative adversarial network, SYN_{b1500} = synthesized high- b -value (b = 1500 sec/mm²) DWI.

technology to further optimize the M_0 so that the synthesized images had a better signal-to-noise ratio and contrast ratio of prostate lesions.

For tumor detection, the mean AUC of the Opt-SYN_{b1500} set was higher than AUC of ACQ_{b1500} and CAL_{b1500} sets, and the ADC value derived from Opt-SYN_{b1500} set also had a higher AUC than those derived from the ACQ_{b1500}, CAL_{b1500}, and SYN_{b1500} sets. These results suggest that Opt-SYN_{b1500} sets can improve the tumor detection capability of radiologists. By comparing the DWI set AUCs from each individual reader, it was found that the Opt-SYN_{b1500} set did improve the tumor detection capability of young radiology residents; however, for the two radiologists with more than 20 years of diagnostic experience, the difference in the AUCs among the DWI sets was not significant. This implies that the image quality of DWI sets may have a greater impact on the diagnostic accuracy of young radiologists who are relatively inexperienced with diagnosis.

In addition to better image quality and better tumor detection, the deep learning model we proposed also has the following advantages. First, unlike acquired DWI technology, our framework does not deeply depend on the gradient performance, radiofrequency design, and software platform. Second, the use of our framework can effectively reduce the DWI time. In general, acquired DWI with high b values requires the use of more averages to achieve image quality similar to that of actual acquired DWI with a b value of 1000 sec/mm². Although the calculated DWI technique can also generate high- b -value DWI sets using lower- b -value images, it requires DWI sets acquired with at least two lower b values (7); in contrast, with our method, only one lower- b -value DWI set is needed. Moreover, the image quality

Figure 6: Left: Receiver operating characteristic (ROC) curve comparison of diagnostic accuracy of diffusion-weighted imaging (DWI) sets based on a five-point Likert score. Right: Area under the ROC curve (AUC) comparison of DWI sets. Values within boxes are z scores for comparison between the two AUC values, with P values also shown. ACQ_{b1500} = acquired high-b-value ($b = 1500 \text{ sec/mm}^2$) DWI, CAL_{b1500} = calculated ($b = 1500 \text{ sec/mm}^2$) DWI, M₀ = deep learning model based on the generative adversarial network, SYN_{b1500} = synthesized high-b-value ($b = 1500 \text{ sec/mm}^2$) DWI set generated by the M₀, Opt-SYN_{b1500} = SYN_{b1500} set generated by the optimized M₀.

Table 4: AUC for Prostate Cancer Detection for DWI Sets

Reader	ACQ _{b1500}	CAL _{b1500}	SYN _{b1500}	Opt-SYN _{b1500}
1	0.71 (0.64, 0.83)	0.80 (0.72, 0.89)	0.85 (0.78, 0.93)	0.87 (0.81, 0.94)
2	0.77 (0.69, 0.88)	0.84 (0.78, 0.93)	0.88 (0.82, 0.95)	0.89 (0.83, 0.96)
3	0.85 (0.79, 0.93)	0.87 (0.83, 0.96)	0.89 (0.84, 0.97)	0.91 (0.85, 0.97)
4	0.84 (0.80, 0.94)	0.90 (0.84, 0.96)	0.91 (0.85, 0.97)	0.91 (0.88, 0.98)
Mean	0.79 (0.76, 0.83)	0.85 (0.82, 0.87)	0.88 (0.85, 0.90)	0.89 (0.87, 0.92)

Note.—The 95% CIs are shown in parentheses. A comparison of the AUC values is shown in Figure 6. ACQ_{b1500} = acquired high-b-value ($b = 1500 \text{ sec/mm}^2$) DWI, AUC = area under the curve, CAL_{b1500} = calculated ($b = 1500 \text{ sec/mm}^2$) DWI, DWI = diffusion-weighted imaging, Opt-SYN_{b1500} = SYN_{b1500} set generated by the optimized deep learning model based on the generative adversarial network, SYN_{b1500} = synthesized high-b-value ($b = 1500 \text{ sec/mm}^2$) DWI.

from calculated DWI with high b values depends on the image quality of lower- b -value images, whereas with our model, acquired full FOV DWI can be used to generate high- b -value DWI sets in which the image quality is better than that of zoomed FOV DWI sets.

Our study had several limitations. First, in this study, we only used DWI sets with b values ranging from 800 to 1000 sec/mm^2 acquired by using 3-T imagers to generate high- b -value DWI sets. Whether our model can generate higher- b -value DWI sets ($b \geq 1500 \text{ sec/mm}^2$), use b values lower than 800 sec/mm^2 , or use DWI acquired by using a low-field-strength MRI scanner for image synthesis must be studied further. Second, in our study, targeted biopsy was used as the reference standard instead of prostatectomy. The use of whole-mount serial slices may improve the accuracy of the MR images in relation to histopathologic findings. Third,

because no moderate or severe cases of distortion were encountered in our study, whether this technique is still effective in cases of severe distortion requires further research and demonstration.

Deep learning based on the GAN may be a promising method to generate high- b -value DWI sets with equal or better image quality and higher accuracy of prostatic cancer detection than acquired DWI sets.

Author contributions: Guarantors of integrity of entire study, L.H., D.W.Z., Y.F.Z., L.L., H. He, W.H.X., L.Q., C.X.F., H. Hu, J.G.Z.; study concepts/study design or data acquisition or data analysis/interpretation, all authors; manuscript drafting or manuscript revision for important intellectual content, all authors; approval of final version of submitted manuscript, all authors; agrees to ensure any questions related to the work are appropriately resolved, all authors; literature research, L.H., D.W.Z., Y.F.Z., L.L., C.X.F., H. Hu, J.G.Z.; clinical studies, L.H., D.W.Z., Y.F.Z., L.L., H. He, W.H.X., Y.K.Z., H. Hu, J.G.Z.; experimental studies, L.H., D.W.Z., Y.F.Z., L.L., L.Q., Y.K.Z., C.X.F., H. Hu, J.G.Z.; statistical analysis, L.H., D.W.Z., L.L., H. Hu, J.G.Z.; and manuscript editing, L.H., D.W.Z., Y.F.Z., L.L., C.X.F., H. Hu, J.G.Z.

Disclosures of Conflicts of Interest: L.H. Activities related to the present article: disclosed no relevant relationships. Activities not related to the present article: disclosed no relevant relationships. Other relationships: author has pending patent involving a diffusion-weighted image generation system and method. D.W.Z. disclosed no relevant relationships. Y.F.Z. disclosed no relevant relationships. L.L. disclosed no relevant relationships. H. He disclosed no relevant relationships. W.H.X. disclosed no relevant relationships. L.Q. disclosed no relevant relationships.

Y.K.Z. disclosed no relevant relationships. **C.X.F.** Activities related to the present article: disclosed no relevant relationships. Activities not related to the present article: disclosed no relevant relationships. Other relationships: employee of Siemens Healthcare. **H. Hu** disclosed no relevant relationships. **J.G.Z.** disclosed no relevant relationships.

References

- Kasivisvanathan V, Rannikko AS, Borghi M, et al. MRI-targeted or standard biopsy for prostate-cancer diagnosis. *N Engl J Med* 2018;378(19):1767–1777.
- Koh DM, Blackledge M, Padhani AR, et al. Whole-body diffusion-weighted MRI: tips, tricks, and pitfalls. *AJR Am J Roentgenol* 2012;199(2):252–262.
- Weinreb JC, Barentsz JO, Choyke PL, et al. PI-RADS Prostate Imaging–Reporting and Data System: 2015, version 2. *Eur Urol* 2016;69(1):16–40.
- Jendoubi S, Wagner M, Montagne S, et al. MRI for prostate cancer: can computed high b-value DWI replace native acquisitions? *Eur Radiol* 2019;29(10):5197–5204.
- Bittencourt LK, Attenberger UI, Lima D, et al. Feasibility study of computed vs measured high b-value (1400 sec/mm²) diffusion-weighted MR images of the prostate. *World J Radiol* 2014;6(6):374–380.
- Ben-Amitay S, Jones DK, Assaf Y. Motion correction and registration of high b-value diffusion weighted images. *Magn Reson Med* 2012;67(6):1694–1702.
- Maas MC, Fütterer JJ, Scheenen TW. Quantitative evaluation of computed high b value diffusion-weighted magnetic resonance imaging of the prostate. *Invest Radiol* 2013;48(11):779–786.
- Brendle C, Martirosian P, Schwenzer NF, et al. Diffusion-weighted imaging in the assessment of prostate cancer: comparison of zoomed imaging and conventional technique. *Eur J Radiol* 2016;85(5):893–900.
- Tamada T, Ream JM, Doshi AM, Taneja SS, Rosenkrantz AB. Reduced field-of-view diffusion-weighted magnetic resonance imaging of the prostate at 3 Tesla: comparison with standard echo-planar imaging technique for image quality and tumor assessment. *J Comput Assist Tomogr* 2017;41(6):949–956.
- Rosenkrantz AB, Hindman N, Lim RP, et al. Diffusion-weighted imaging of the prostate: comparison of b1000 and b2000 image sets for index lesion detection. *J Magn Reson Imaging* 2013;38(3):694–700.
- Zhu Y, Man C, Gong L, et al. A deep learning radiomics model for preoperative grading in meningioma. *Eur J Radiol* 2019;116:128–134.
- Glaister J, Cameron A, Wong A, Haider MA. Quantitative investigative analysis of tumour separability in the prostate gland using ultra-high b-value computed diffusion imaging. *Annu Int Conf IEEE Eng Med Biol Soc* 2012;420–423.
- Rosenkrantz AB, Chandarana H, Hindman N, et al. Computed diffusion-weighted imaging of the prostate at 3 T: impact on image quality and tumour detection. *Eur Radiol* 2013;23(11):3170–3177.
- Ueno Y, Takahashi S, Ohno Y, et al. Computed diffusion-weighted MRI for prostate cancer detection: the influence of the combinations of b-values. *Br J Radiol* 2015;88(1048):20140738.
- Armanious K, Jiang C, Fischer M, et al. MedGAN: medical image translation using GANs. *Comput Med Imaging Graph* 2020;79:101684.
- Goodfellow IJ, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets. In: Ghahramani Z, Welling M, Cortes C, Lawrence ND, Weinberger KQ, eds. *Proceedings of the 27th International Conference on Neural Information Processing*. Vol 2. Cambridge, Mass: MIT Press, 2014; 2672–2680.
- Lee JH, Han IH, Kim DH, et al. Spine computed tomography to magnetic resonance image synthesis using generative adversarial networks: a preliminary study. *J Korean Neurosurg Soc* 2020;63(3):386–396.
- Lei Y, Harms J, Wang T, et al. MRI-only based synthetic CT generation using dense cycle consistent generative adversarial networks. *Med Phys* 2019;46(8):3565–3581.
- Sara U, Akter M, Uddin MS. Image quality assessment through FSIM, SSIM, MSE and PSNR—a comparative study. *J Comput Comm* 2019;7(3):8–18.
- Zhang K, Zuo W, Chen Y, Meng D, Zhang L. Beyond a Gaussian denoiser: residual learning of deep CNN for image denoising. *IEEE Trans Image Process* 2017;26(7):3142–3155.
- Hectors SJ, Semaan S, Song C, et al. Advanced diffusion-weighted imaging modeling for prostate cancer characterization: correlation with quantitative histopathologic tumor tissue composition—a hypothesis-generating study. *Radiology* 2018;286(3):918–928.
- Rosenkrantz AB, Chandarana H, Pfeuffer J, et al. Zoomed echo-planar imaging using parallel transmission: impact on image quality of diffusion-weighted imaging of the prostate at 3T. *Abdom Imaging* 2015;40(1):120–126.
- Katahira K, Takahara T, Kwee TC, et al. Ultra-high-b-value diffusion-weighted MR imaging for the detection of prostate cancer: evaluation in 201 cases with histopathological correlation. *Eur Radiol* 2011;21(1):188–196.
- Ohgiya Y, Suyama J, Seino N, et al. Diagnostic accuracy of ultra-high-b-value 3.0-T diffusion-weighted MR imaging for detection of prostate cancer. *Clin Imaging* 2012;36(5):526–531.
- Kitchen A, Seah J. Deep generative adversarial neural networks for realistic prostate lesion MRI synthesis. *ArXiv 1708.00129 [preprint]* <https://arxiv.org/abs/1708.00129>. Posted August 1, 2017. Accessed July 1, 2021.
- Chuquicusma MJM, Hussein S, Burt J, Bacgi U. How to fool radiologists with generative adversarial networks? A visual Turing test for lung cancer diagnosis. In: *Proceedings of the 2018 IEEE 15th international symposium on biomedical imaging*. Piscataway, NJ: Institute of Electrical and Electronics Engineers, 2018; 240–244.
- Wolterink JM, Leiner T, Viergever MA, Išgum I. Generative adversarial networks for noise reduction in low-dose CT. *IEEE Trans Med Imaging* 2017;36(12):2536–2545.
- Chartsias A, Joyce T, Papanastasiou G, et al. Factorised spatial representation learning: application in semi-supervised myocardial segmentation. In: Frangi A, Schnabel J, Davatzikos C, Alberola-López C, Fichtinger G, eds. *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018*. Vol 11071, Lecture Notes in Computer Science. Cham, Switzerland: Springer, 2018; 490–498.
- Cohen JP, Luck M, Honari S. Distribution Matching Losses Can Hallucinate Features in Medical Image Translation. In: Frangi A, Schnabel J, Davatzikos C, Alberola-López C, Fichtinger G, eds. *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018*. Vol 11070, Lecture Notes in Computer Science. Cham, Switzerland: Springer, 2018; 529–536.
- Li C, Wand M. Combining Markov random fields and convolutional neural networks for image synthesis. In: *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Piscataway, NJ: Institute of Electrical and Electronics Engineers, 2016; 2479–2486.
- Chen C, Liu W, Tan X, Wong KYK. Semi-supervised Learning for Face Sketch Synthesis in the Wild. In: Jawahar C, Li H, Mori G, Schindler K, eds. *Computer Vision – ACCV 2018*. Vol 11361, Lecture Notes in Computer Science. Cham, Switzerland: Springer, 2019; 216–231.
- Yu Z, Xiang Q, Meng J, Kou C, Ren Q, Lu Y. Retinal image synthesis from multiple-landmarks input with generative adversarial networks. *Biomed Eng Online* 2019;18(1):62.
- Kim H, Lee JM, Yoon JH, et al. Reduced field-of-view diffusion-weighted magnetic resonance imaging of the pancreas: comparison with conventional single-shot echo-planar imaging. *Korean J Radiol* 2015;16(6):1216–1225.
- Thierfelder KM, Scherr MK, Notohamiprodjo M, et al. Diffusion-weighted MRI of the prostate: advantages of zoomed EPI with parallel-transmit-accelerated 2D-selective excitation imaging. *Eur Radiol* 2014;24(12):3233–3241.