StitchAD-GAN for Synthesizing Apparent Diffusion Coefficient Images of Clinically Significant Prostate Cancer

Zhiwei Wang[1]
zhiweiwang@hust.edu.cn
Yi Lin[1]
lin_yi@hust.edu.cn
Chunyuan Liao[2]
liaocy@hiscene.com
Kwang-Ting (Tim) Cheng[3]
timcheng@ust.hk
Xin Yang[1*]
xinyang2014@hust.edu.cn

[1] Shool of Electronic Information and Communications, Huazhong University of Science and Technology, Wuhan, China
(* Corresponding Author)

[2] HiScene Information Technology, Co., Ltd, China

[3] School of Engineering, Hong Kong University of Science and Technology, Hong Kong, China

Abstract

The high volume and quality of apparent diffusion coefficient (ADC) data containing clinically significant (CS) prostate cancer (PCa) are critical for automated PCa detection with a high accuracy. However, ADC data of CS PCa is scarce and costly to obtain in practice. This paper proposes a novel Generative Adversarial Network (GAN), named StitchAD-GAN, for synthesizing high-quality ADC images of CS PCa. Our StitchAD-GAN employs a StitchLayer in the generator to address the difficult-to-optimize problem in most GANs. Instead of directly optimizing a complex generation from a low dimensional noise to an ADC image, we optimize n easier generations of sub-images, which are then aggregated into a full size image in the StitchLayer. Our discriminative module approximates two distances: 1) the Wasserstein distance (W-distance) between the synthetic and real ADC data of CS PCa, and 2) an auxiliary distance (AD) Jensen-Shannon divergence (JSD) between the synthetic CS PCa and real nonCS PCa data. By minimizing the W-distance and maximizing the JSD simultaneously, our StitchAD-GAN can capture CS PCa features in addition to predominant prostate gland information, and in turn synthesize more clinically meaningful ADC data of CS PCa. Visual and quantitative results demonstrate greater quality of our synthetic CS PCa data than those of the state-of-the-art methods' and even real data.

1 Introduction

Prostate cancer (PCa) is the most commonly diagnosed cancer other than skin cancer, and also one of the leading causes of cancer death among men [20]. Men with clinically significant (CS) PCa whose Gleason Score (GS) is equal to or greater than 7 could experience high fatality rates [21]. Fortunately, it has been demonstrated recently that the ADC values

derived from diffusion-weighted magnetic resonance images (DWI-MRI) are useful non-invasive biomarkers for accurate detection of CS PCa [15]. A timely diagnosis of PCa via ADC data followed by a proper treatment can greatly decrease the mortality. However, ADC data of CS PCa is often scarce, limiting the usage of powerful yet data-hungry advances in deep learning, e.g. convolutional neural network (CNN), for PCa detection based on ADC images.

Several efforts [6, 12, 17, 18, 25, 26] have been made to address this problem. A common approach is data augmentation by applying rigid and/or non-rigid image transformations to the original data to increase the data volume [12, 25, 26]. However, neither rigid nor non-rigid transformations can address the problem of lack of data variety. Another category of approaches is to synthesize medical data to increase the data variety via creating 'unseen' data [6, 17, 18, 22]. Among these approaches, GAN based image synthesis has achieved the greatest success in multiple medical image analysis applications, e.g. retinal image synthesis [18] and cross-domain medical data transformation [17]. A typical GAN [8] consists of two modules: 1) a *generative module* to generate fake data based on given random noises, and 2) a *discriminative module* to approximate the distance between the distributions of fake and real data. Existing GAN-based methods differ from each other in the design of these two modules. Osokin *et al.* [18] proposed a dual path generative module, each path of which generates an individual channel of two-channel cell images. The discriminative module estimates the W-distance between the fake and real data, which is then minimized for training the dual-path generative module. Nie *et al.* [17] reconstructed CT images from MR images by utilizing a fully convolutional network (FCN) as a generative module to convert an MR image to CT. Their generative module was trained by minimizing both CT reconstruction error and the JSD between the reconstructed and real CT. Despite the successes, these GAN based methods directly mapped a noise vector with a very low dimension (e.g., 128-d) to a medical image with a higher dimension (e.g., 64×64, 128×128), yielding an unstable and fragile optimization process. Additionally, little effort has been made to address the problem of clinically significant medical data synthesis, for example ADC image of CS PCa in our case. Directly applying existing methods could hardly capture clinically meaningful information of CS PCa. As a result, generative modules of existing methods mostly rely on the predominant prostate gland information to synthesize ADC data while largely ignoring CS PCa-relevant visual patterns.

This paper presents a GAN-based method, named StitchAD-GAN, for synthesizing ADC data of CS PCa with an original size of 64×64. The quality of synthetic data by the GANs, e.g. DCGAN [19] and InfoGAN [3], usually degrades dramatically as the image size increases, implying that generation of low dimensional images is much easier than that of higher dimensions. Motivated by this, we divide a target image space into subspaces, each of which has a lower dimension, to reduce the complexity for modeling the manifold of data. Instead of directly generating the image in the target space, we first synthesize sub-images in divided subspaces, and then 'stitch' them into the full size target image by a non-parametric StitchLayer in interlaced manner as shown in Fig. 1. To capture more clinically meaningful CS PCa information, our discriminative module approximates two distances: 1) the W-distance to the real CS PCa data, and 2) the auxiliary distance (AD) of JSD to the real nonCS PCa data. Minimizing the W-distance enables visually realistic ADC image synthesis, and maximizing the JSD enforces synthesis of ADC data with clinically meaningful CS PCa features.

The proposed method is evaluated in terms of both visual quality and the applicability to CS vs. nonCS PCa classification. Experimental results demonstrate that the synthetic ADC

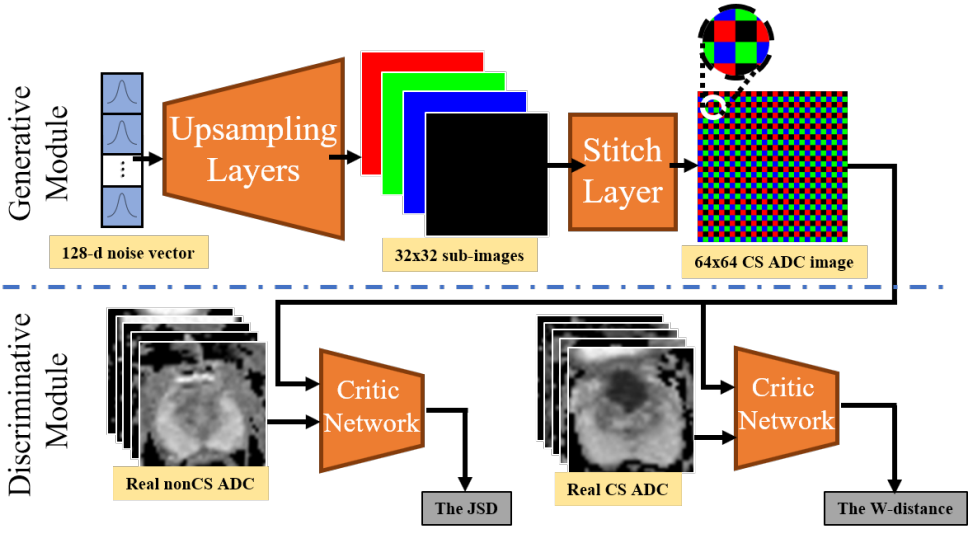

Figure 1: The framework of our proposed StitchAD-GAN for CS ADC data synthesis.

data by our method is of high quality and could significantly improve the CS vs. nonCS classification accuracy comparing with those synthesized by the state-of-the-art methods [9, 19] and real data.

2 Method

A typical matrix size of abdominal MRI scan is around 180×144, in which the prostate gland and its neighboring tissues roughly locate at the center of an ADC image and cover around $1/9$ area of the entire image. The goal of this work is to synthesize ADC images of prostate regions with the original resolution (i.e. 64×64). Fig. 1 shows the framework of the proposed StitchAD-GAN. Detailed network architecture is shown in the supplementary material. The generative module consists of a batch of upsampling layers to synthesize four 32×32 intermediate sub-images given a random 128-d noise, followed by a Stitch-Layer to interlace them into a 64×64 target image. The discriminative module consists of two critic networks: one approximates the W-distance between the synthetic data and the real ADC data of CS PCa, and the other approximates the JSD from the real nonCS PCa data as an auxiliary distance (AD). In each training round, we first train the discriminative module several times for a better approximation of the two distances. Then the generative module is trained to simultaneously minimize the W-distance and maximize the JSD with the parameters of discriminative module fixed. In the following, we detail each module.

2.1 The Generative Module with a StitchLayer

This work aims to synthesize ADC image of CS PCa with an original size of 64×64. Most GANs are limited to synthesize very small images such as images from the CIFAR-10 [10] and MNIST [13] datasets whose image sizes are 32×32 and 28×28 respectively. Before introducing our solution, we first visually compare the performance of the widely-used DC-

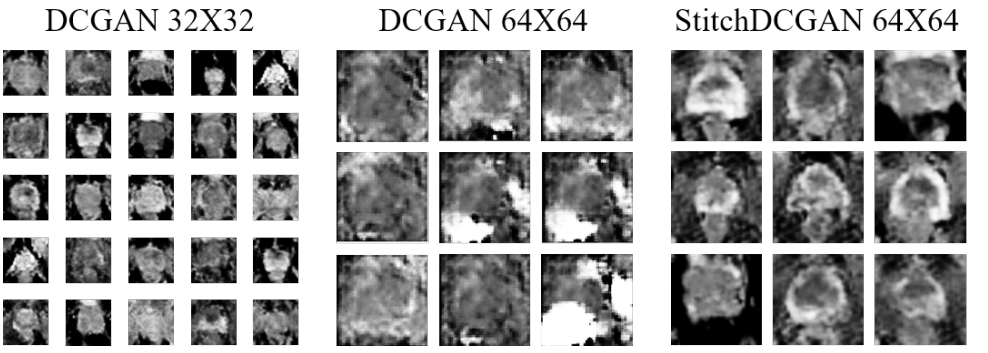

Figure 2: Exemplar outputs synthesized by three different DCGAN models.

GAN for synthesizing CS ADC data of two different sizes, i.e. 32×32 and 64×64 in Fig. 2. The 32×32 images for training the DCGAN are obtained by down-scaling the original 64×64 ADC images of CS PCa. We manually optimize the configurations of the two DCGANs (i.e. *DCGAN 32X32* and *DCGAN 64X64*) for generating visually reasonable results. As can be seen in Fig. 2, the *DCGAN 32X32* is able to synthesize somewhat acceptable ADC images where prostate glands have a visible nut-like shape, while the performance of *DCGAN 64X64*, whose the image dimension is increased by four times compared to *DCGAN 32X32*, is poor with an ambiguous shape for the prostate gland.

A potential solution to synthesize higher dimensional images is coarse-to-fine learning adopted in recent studies [6, 10, 23]. In these studies, customized generative networks and/or sophisticated training strategies were developed to synthesize data from low to high resolutions gradually. However, these enabling techniques are usually hard and time-consuming to tune in their training phase, and overqualified for synthesis of 64×64 images given that synthesis of 32×32 is within the power of most plain GANs.

Inspired by the similar idea of coarse-to-fine learning, we propose a StitchLayer, which is simple yet effective, and can be embedded in any generative networks to boost existing GANs to synthesize images with greater size. Specifically, given a noise vector z, instead of modeling a direct generation $G(z) \rightarrow X$, where X represents the 64×64 real ADC image and G is a generative network, our generative module produces four generations $\{G^n(z) \rightarrow x^n, n = 1,2,3,4\}$ of 32×32 sub-images. As shown in Fig. 1, x^1, x^2, x^3, and x^4 are denoted by the red, green, blue and black squared feature maps respectively. To 'stitch' these 4 sub-images $\{x^n\}$ into the full size ADC image X by the StitchLayer, we adopt a similar idea described in [16] called *shift-and-stitch*, which aims to obtain dense prediction from coarse outputs. We consider that X with image size of 64×64 consists of 32×32 non-overlapped 2×2 blocks, each of which is assembled by pixels from the same location in the four sub-images, which is formulated as:

$$\begin{bmatrix} X_{2i-1,2j-1} & X_{2i-1,2j} \\ X_{2i,2j-1} & X_{2i,2j} \end{bmatrix} = B_{i,j} = \begin{bmatrix} x^1_{i,j} & x^2_{i,j} \\ x^3_{i,j} & x^4_{i,j} \end{bmatrix}, i,j = 1,2,\ldots,32 \quad (1)$$

where $B_{i,j}$ indicates a 2×2 block at the i-th row and j-th column.

Rearranging pixels from $\{x^n\}$ according to Eq. (1), we can interlace the four sub-images into a 64×64 target ADC image. We design our generative networks $\{G^n, n = 1,2,3,4\}$ for producing the four sub-images as follows. The four networks share common features

in the up-sampling layers, and in the last fully-convolutional layer four convolutional kernels are applied to produce four feature maps $\{x^n, n = 1,2,3,4\}$ which are considered as the four sub-images. Sharing common features in the up-sampling layers among $\{G^n\}$ ensures the globally spatial consistency among the four sub-images and reduces many unnecessary computational costs. Meanwhile, by utilizing four different convolutional kernels in the last layer, each sub-image captures some unique and detailed information of the full size image and such information in each sub-image is complementary with each other. Accordingly, by 'stitching' the four sub-images together we can obtain a full size image with both global structure and detailed local information. The generative networks $\{G^n\}$ are trained concurrently by minimizing the distance between a synthesized full size image and real 64×64 ADC data, which is approximated by the following discriminative module. Minimizing such distance enforces our generative networks to not only output sub-images that are globally-consistent with the true full size image but also to encode complementary local information in each sub-image.

Comparing to directly synthesizing the full size image using a single generative network, the StitchLayer allows us to focus on the divided image subspaces with size of 32×32, in which each one of $\{G^n\}$ is capable of correctly mimicking the data distribution. In turn, we can obtain robust synthesis of higher dimensional ADC images via learning the correlations among sub-images (i.e. the complementary local information of sub-images). We demonstrate its impact by embedding a StitchLayer into a DCGAN (denoted by *StitchDCGAN 64X64*). Exemplar outputs in the 3^{rd} column of Fig. 2 show that with a StitchLayer, the DCGAN is able to synthesize reasonable 64×64 ADC images of CS PCa with visible nut-shape prostate glands.

The main difference between the *shift-and-stitch* and our proposed StitchLayer is that the *shift-and-stitch* obtains sub-images using the same network from the multiple shifted versions of input image while the StitchLayer enables learning of multiple generative networks $\{G^n\}$ to get sub-images, which could guarantee learnable and reasonable complementary local information of sub-images, yielding more robust full size target images with higher dimension.

2.2 The Discriminative Module with Auxiliary Distance

The discriminative module which approximates the distance between the synthetic and real data is indispensable for the optimization of a generative module in which the approximated distance is minimized.

In this work, we utilized the W-distance between synthetic and real data, which is approximated by the discriminative module as follows:

$$W(\theta_G) = \max_{\theta_D}\{E_{x\sim\mathbb{P}_{CS}}[D(x;\theta_D))] - E_{z\sim\mathbb{P}_z}[D(G(z;\theta_G);\theta_D)] - \lambda R(\theta_D)\} \tag{2}$$

where D is a critic network with tunable parameters θ_D, G is the generative module with tunable parameters θ_G, $\mathbb{P}_{CS}$ and $\mathbb{P}_z$ are distributions of real CS ADC data and noise, $R(\theta_D)$ is for enforcing the 1-Lipschitz constraint of D [9].

W-distance has been recognized as an optimal distance for training a generative module since it always guarantees meaningful gradients w.r.t θ_G regardless of how far away between real and synthetic data [2]. However, minimizing only the W-distance in a generative module achieves visually realistic images of a prostate gland with little CS PCa patterns. The problems are twofold: 1) normal prostate gland tissues typically take predominant regions in

a prostate ADC image comparing with a CS lesion, yielding great difficulties for the generative module to capture sufficient CS PCa-relevant information, and 2) real CS PCa data for training is quite scarce, leading to over-fitting with a high probability and in turn worsen the former problem. To address the two problems, we introduce another critic network to learn CS PCa features from a prostate gland by distinguishing between synthetic CS PCa data and real nonCS PCa data. In this critic network, an auxiliary distance JSD between synthetic CS PCa and real nonCS PCa data is approximated by by the discriminative module as follows:

$$L(\theta_G) = \max_{\theta_D}\{E_{x\sim\mathbb{P}_{nonCS}}[log(D(x;\theta_D))] + E_{z\sim\mathbb{P}_z}[1 - log(D(G(z;\theta_G);\theta_D))]\} \tag{3}$$

Accordingly, after approximation of the two distances, the overall objective of our generative module is to concurrently minimize the W-distance and maximize the JSD as expressed in Eq. (4):

$$\theta_G = \arg\min_{\theta_G}\{\alpha W(\theta_G)\} + \arg\max_{\theta_G}\{\beta L(\theta_G)\} \tag{4}$$

where α and β are weights tuning the contributions of the two distances, which are set to 1 and 0.1 respectively in our experiments.

We choose JSD as an auxiliary distance (AD) rather than W-distance is because JSD could better guide the generative module to increase the distance between CS and nonCS PCa data only if the synthetic data lacks CS PCa information since JSD derives no gradient unless the manifolds of synthetic CS and real nonCS PCa data align each other [1].

The over-fitting problem of the generator can be greatly alleviated, as there is far more nonCS data available than CS PCa data for training. The task of simultaneously minimizing the W-distance and maximizing the JSD forces the generator to capture CS PCa patterns besides the prostate gland.

3 Experiment Results

3.1 Dataset

The study was approved by our local institutional review board. The ADC images used in the study are collected from two datasets: 1) a locally collected dataset including 156 patients' data pathologically validated by a 12-core systematic TRUS-guided plus targeted prostate biopsy. Dataset details are listed in [24, 25], 2) a public dataset PROSTATEx (training) [4, 7, 14], including data of 204 MRI-targeted biopsy-proven patients. Among 360 patients' data, 226 patients are normal, with benign prostatic hyperplasia (BPH) or indolent lesions, which are collectively referred to as nonCS PCa, and 134 patients are with CS PCa. From the two datasets, a radiologist manually select 533 original ADC images containing CS PCa and 1992 ADC images containing nonCS PCa. The selection criterion was that both CS PCa lesions and prostate glands were clearly visible. We applied a prostate detector [25] to obtain prostate ADC images. The entire dataset was randomly divided into the training set (483 CS and 1942 nonCS) and the test set (50 CS and 50 nonCS).

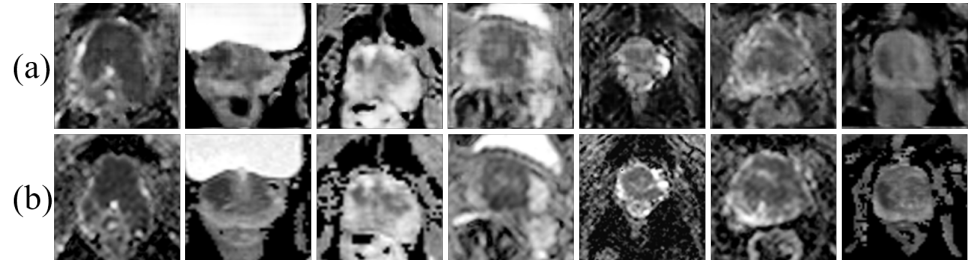

Figure 3: (a) synthetic images of our method and their (b) closest real images.

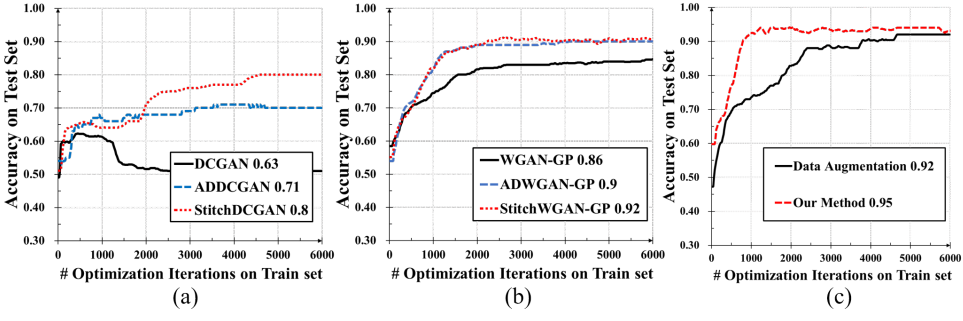

Figure 4: Curves of CS vs. nonCS PCa classification accuracy for the test set with respect to the optimization iterations. Data Augmentation indicates using a combination of the real and conventionally augmented data for training

3.2 Subjective Visual Quality Evaluation

We first qualitatively evaluate the performance of our StitchAD-GAN. To demonstrate that the StitchAD-GAN does not simply memorize the training data, for each synthetic image G_i we identified its most similar image R_i which has the maximal Mutual Information $MI(G_i, R_i)$. Fig. 3 shows that both nut-shape prostate gland and CS PCa lesions are clearly visible in our synthetic data with correct spatial relationships. In addition, the synthetic images differ from their closest real images in terms of gland shape, lesion distribution and surrounding tissues. These results well demonstrates the good generalization ability of our model.

3.3 CS vs. nonCS PCa Classification Using Synthetic Training Data

We quantitatively examine the quality of the synthetic data via a slice-level CS vs. nonCS PCa classification using the classification accuracy as the metric. First, we evaluate the accuracy improvement achieved by the proposed generative module with a StitchLayer and the discriminative module with an auxiliary distance respectively. For this evaluation, we implemented two state-of-the-art GANs, i.e., DCGAN [19] and WGAN-GP [9]. Fig. 4 summarizes the results. The prefix 'Stitch' and 'AD' before the model name respectively indicate embedding a StitchLayer and an auxiliary distance in the two methods. We trained all GANs for generating the 64×64 ADC data of CS PCa using the training set without any data augmentation. For each GAN, we combine the 1942 synthesized ADC images of CS PCa

and 1942 real ADC images of nonCS PCa from the training set to train an Artificial Neural Network (ANN), which consists of two fully connected layers, as the classifier. As the DCGAN and ADDCGAN failed to synthesize 64×64 data, we simply up-scaled the 32×32 synthetic images to 64×64 via bilinear interpolation. The results in Fig. 4(a) and (b) show that the proposed StitchLayer and auxiliary distance provide obvious improvements for both DCGAN and WGAN-GP. Even though WGAN-GP is already able to synthesize 64×64 images, StitchWGAN-GP still outperforms WGAN-GP, indicating that the StitchLayer is not limited to enabling higher dimensional image generation, but also can make the data be easier analyzed by GANs.

We further compare the real data with the synthetic data of our method (i.e. applying both StitchLayer and AD to WGAN-GP, denoted as StitchAD-GAN). For a fair comparison, we augmented real CS PCa data to 1942 as [25]. Fig. 4(c) shows that our method outperforms the real data with data augmentation, validating that the proposed method could be a better alternative for addressing the insufficiency of medical data need for data-hungry deep learning models.

4 Conclusion

This paper presents the StitchAD-GAN for synthesis of realistic and clinically meaningful ADC images of CS PCa. Two main contributions of our StitchAD-GAN are: 1) a StitchLayer to effectively address the difficult-to-optimize problem for high-dimensional image synthesis, and 2) an auxiliary distance based on JSD between the synthetic CS PCa data and real nonCS PCa data to ensure clinically meaningful CS PCa features to be presented in synthetic images. Extensive experimental results demonstrate the superiority of our method to the state-of-the-art methods [9, 19] and real data with conventional data augmentation.

Acknowledgment

This work was supported by the National Key Research and Development Program of China (2017YFA0700402), the National Natural Science Foundation of China (61502188), the Wuhan Science and Technology Bureau under Award (2017010201010111), and the Program for HUST Acadamic Frontier Youth Team.

References

[1] Martin Arjovsky and Léon Bottou. Towards principled methods for training generative adversarial networks. *arXiv preprint arXiv:1701.04862*, 2017.

[2] Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein gan. *arXiv preprint arXiv:1701.07875*, 2017.

[3] Xi Chen, Yan Duan, Rein Houthooft, John Schulman, Ilya Sutskever, and Pieter Abbeel. Infogan: Interpretable representation learning by information maximizing generative adversarial nets. In *Advances in Neural Information Processing Systems*, pages 2172–2180, 2016.

[4] Kenneth Clark, Bruce Vendt, Kirk Smith, John Freymann, Justin Kirby, Paul Koppel, Stephen Moore, Stanley Phillips, David Maffitt, and Michael Pringle. The cancer imaging archive (tcia): Maintaining and operating a public information repository. *Journal of Digital Imaging*, 26(6):1045, 2013.

[5] Pedro Costa, Adrian Galdran, Maria Ines Meyer, Meindert Niemeijer, Michael Abràmoff, Ana Maria Mendonça, and Aurélio Campilho. End-to-end adversarial retinal image synthesis. *IEEE transactions on medical imaging*, 2017.

[6] Emily L Denton, Soumith Chintala, Rob Fergus, et al. Deep generative image models using a laplacian pyramid of adversarial networks. In *Advances in neural information processing systems*, pages 1486–1494, 2015.

[7] Litjens Geert, Debats Oscar, Barentsz Jelle, Karssemeijer Nico, and Huisman Henkjan. Prostatex challenge data, 2017. URL https://doi.org/10.7937/K9TCIA.2017.MURS5CL. The Cancer Imaging Archive.

[8] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680, 2014.

[9] Ishaan Gulrajani, Faruk Ahmed, Martin Arjovsky, Vincent Dumoulin, and Aaron Courville. Improved training of wasserstein gans. *arXiv preprint arXiv:1704.00028*, 2017.

[10] Xun Huang, Yixuan Li, Omid Poursaeed, John Hopcroft, and Serge Belongie. Stacked generative adversarial networks. *arXiv preprint arXiv:1612.04357*, 2016.

[11] Alex Krizhevsky and Geoffrey Hinton. Learning multiple layers of features from tiny images. 2009.

[12] Minh Hung Le, Jingyu Chen, Liang Wang, Zhiwei Wang, Wenyu Liu, Kwang-Ting Tim Cheng, and Xin Yang. Automated diagnosis of prostate cancer in multi-parametric mri based on multimodal convolutional neural networks. *Physics in Medicine & Biology*, 62(16):6497, 2017.

[13] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

[14] Geert Litjens, Oscar Debats, Jelle Barentsz, Nico Karssemeijer, and Henkjan Huisman. Computer-aided detection of prostate cancer in mri. *IEEE transactions on medical imaging*, 33(5):1083–1092, 2014.

[15] Gerardus Johannes Silvester Litjens. *Computerized detection of cancer in multi-parametric prostate MRI*. [Sl: sn], 2015.

[16] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440, 2015.

[17] Dong Nie, Roger Trullo, Jun Lian, Caroline Petitjean, Su Ruan, Qian Wang, and Dinggang Shen. Medical image synthesis with context-aware generative adversarial networks. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 417–425. Springer, 2017.

[18] Anton Osokin, Anatole Chessel, Rafael E Carazo Salas, and Federico Vaggi. Gans for biological image synthesis. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 2252–2261. IEEE, 2017.

[19] Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*, 2015.

[20] Rebecca L Siegel, Kimberly D Miller, and Ahmedin Jemal. Cancer statistics, 2015. *CA: a cancer journal for clinicians*, 65(1):5–29, 2015.

[21] Anton Stangelberger, Matthias Waldert, and Bob Djavan. Prostate cancer in elderly men. *Reviews in urology*, 10(2):111, 2008.

[22] Gijs van Tulder and Marleen de Bruijne. Why does synthesized data improve multi-sequence classification? In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 531–538. Springer, 2015.

[23] Ting-Chun Wang, Ming-Yu Liu, Jun-Yan Zhu, Andrew Tao, Jan Kautz, and Bryan Catanzaro. High-resolution image synthesis and semantic manipulation with conditional gans. *arXiv preprint arXiv:1711.11585*, 2017.

[24] Zhiwei Wang, Chaoyue Liu, Danpeng Cheng, Liang Wanga, Xin Yang, and K-T Tim Chengb. Automated detection of clinically significant prostate cancer in mp-mri images based on an end-to-end deep neural network. *IEEE Transactions on Medical Imaging*, 2018.

[25] Xin Yang, Chaoyue Liu, Zhiwei Wang, Jun Yang, Hung Le Min, Liang Wang, and Kwang-Ting Tim Cheng. Co-trained convolutional neural networks for automated detection of prostate cancer in multi-parametric mri. *Medical image analysis*, 42:212–227, 2017.

[26] Xin Yang, Zhiwei Wang, Chaoyue Liu, Hung Minh Le, Jingyu Chen, Kwang-Ting Tim Cheng, and Liang Wang. Joint detection and diagnosis of prostate cancer in multi-parametric mri based on multimodal convolutional neural networks. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 426–434. Springer, 2017.