THE UNIVERSITY of EDINBURGH

Edinburgh Research Explorer

Fine-Grained MRI Reconstruction Using Attentive Selection Generative Adversarial Networks

OPEN ACCESS

FINE-GRAINED MRI RECONSTRUCTION USING ATTENTIVE SELECTION GENERATIVE ADVERSARIAL NETWORKS

Jingshuai Liu   *Mehrdad Yaghoobi*

IDCOM, School of Engineering, University of Edinburgh, UK, EH9 3JE
email: s1678560@ed.ac.uk, m.yaghoobi-vaighan@ed.ac.uk

ABSTRACT

Compressed sensing (CS) leverages the sparsity prior to provide the foundation for fast magnetic resonance imaging (fastMRI). However, iterative solvers for ill-posed problems hinder their adaption to time-critical applications. Moreover, such a prior can be neither rich to capture complicated anatomical structures nor applicable to meet the demand of high-fidelity reconstructions in modern MRI.

Inspired by the state-of-the-art methods in image generation, we propose a novel attention-based deep learning framework to provide high-quality MRI reconstruction. We incorporate large-field contextual feature integration and attention selection in a generative adversarial network (GAN) framework. We demonstrate that the proposed model can produce superior results compared to other deep learning-based methods in terms of image quality, and relevance to the MRI reconstruction in an extremely low sampling rate diet.

Index Terms— MRI Reconstruction, GAN-based Framework, Attention Selection

1. INTRODUCTION

Magnetic resonance imaging (MRI) utilizes a strong magnetic field and radio waves to primarily generate 2-dimensional (2-D) slices of cross-sections, and provides a radiation-free diagnosis tool. However, the time-consuming sampling and reconstruction steps impede its applications in time-critical diagnosis. The long acquisition time can be reduced via parallel imaging (PI). One successful PI method is to estimate the unobserved data in k-space from their neighboring points, by using a GRAPPA kernel [1] which is estimated using the fully collected data in the central region. However, such methods require expensive equipments and it is difficult to remove strong aliasing artifacts, using traditional PI methods.

Model-based MRI reconstruction leverages compressive sampling methods based on the data sparsity assumption. They model the image prior in the form of signal sparsity in some domain, e.g. Fourier space, and achieve accurate reconstruction by solving nonlinear optimizations, which are successful if some conditions, e.g. restricted isometry property, are met. The challenge in meeting such hypotheses in real-world scenarios hinders the development of CS methods in fast MRI reconstructions. In a contrast to the request for sparsity of signals in image domain, or a transform domain [2], [3] proposes to learn a sparse basis via dictionary learning which enables more parsimonious representations. However, the limited capacity of the sparsity prior puts restrictions on fast MR imaging and leaves rooms for improvements using deep learning-based methods.

Deep neural networks show great effectiveness in feature representation and have been leveraged to recover the under-sampled MRI observations. One parallel MR imaging method is proposed in [4] using two U-shaped networks (U-net) to predict the missing pixels. Generative adversarial networks (GAN) [5] show success in image generation and can be potentially employed in MRI reconstructions. The method in [6] performs MRI reconstruction by jointly optimizing the latent space of a pre-trained GAN to enforce it being in agreement with the measurements and tuning the parameters of the model to leverage the deep image prior [7]. Motivated by the previous works in domain translation [8], concurrent researches using cyclic adversarial frameworks include: Refine-GAN in [9] which applies the cyclic consistency to learn the mapping of the under-sampled measurements, and [10] which utilizes the theory of optimal transport to provide a foundation for cycle-consistent GAN (CycleGAN).

We propose to learn the mapping from the under-sampled data to the alias-free images by leveraging the generative prior and cyclic data consistency. We introduce a novel deep de-aliasing module to capture large-field spatial dependencies in feature spaces and form a coarse to fine MR image mapping in two stages. Our work is the first endeavour to present a channel-wise and spatial attention selection for MRI reconstruction. By qualitative and quantitative evaluations, we demonstrate that the proposed framework outperforms other data-driven methods in terms of reconstruction quality.

2. METHOD

2.1. Problem formulation

Let s be the fully-sampled MR image, i.e. sampled in k-space with Nyquist–Shannon rate. MRI reconstruction methods tra-
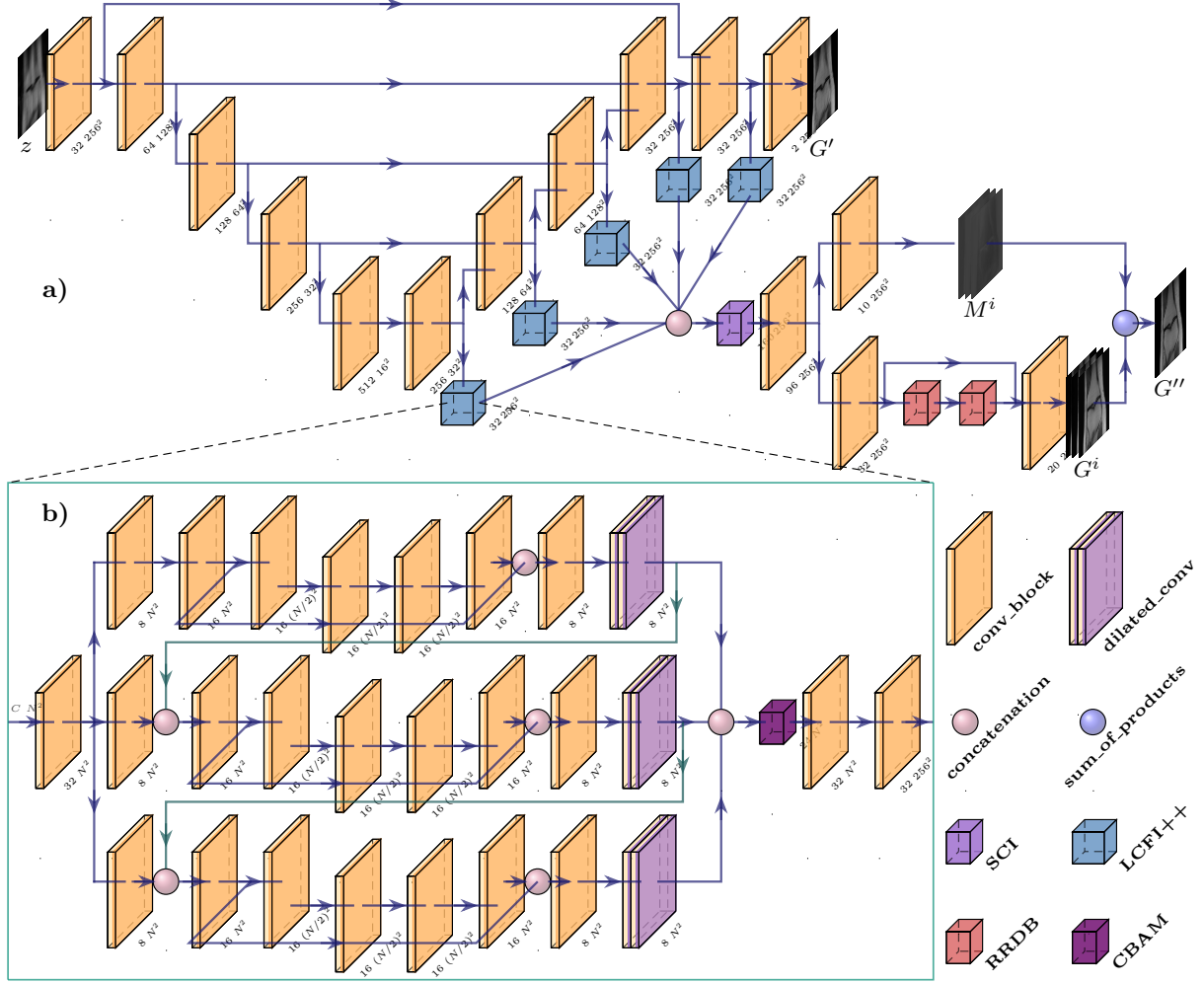
Fig. 1. Overview of the model architecture, a) zero-filled is preliminarily recovered by the U-net and the features at different scales are used to give the final output by performing the temporal and spatial attentive selection, and b) the structure of LCFI++.

ditionally recover the measurement $y = \mathcal{H} \circ \mathcal{F}(s)$, where $\mathcal{H}$ and $\mathcal{F}$ respectively denote the under-sampling operation and Fourier transform, by the following equation,

$$x = arg\min_{x} \|y - \mathcal{H} \circ \mathcal{F}(x)\|^2 + \lambda R(x), \qquad (1)$$

where $R(x)$ refers to a regularization penalty. However, existing optimization methods are computationally complex due to the iterative calculation of gradients, which involves manipulations of large matrices. We instead introduce a data-driven reconstruction framework which replaces the optimization process with a trained neural network, achieving high-quality reconstruction at a very low sampling rate.

2.2. Model architecture

We describe the architecture of proposed framework for MRI reconstruction in this section.

2.2.1. LCFI++

Motivated by the work in [11], which introduces a novel large-field contextual feature integration (LCFI) module, to capture long-range dependencies, we propose a deep dealiasing LCFI (LCFI++) block displayed in Figure 1. We replace the spatially separable convolution with shallow U-nets in the parallel structure of LCFI. The outputs of dilated convolution layers are fused together by a convolutional block attention module (CBAM) [12]. We discovered that our network equipped with LCFI++ behaves more stably than LCFI-integrated networks.

2.2.2. Generator

We propose to recover highly aliased observations in two stages, as illustrated in Figure 1. Initially, we adopt a U-net as the backbone to produce the coarse reconstruction G' from the aliased input. In order to circumvent an information bot-

tleneck, we extract features from the predicted syntheses at different levels of the decoder by using the proposed LCFI++.

We exploit the multi-channel attention selection mechanism which performs a more sophisticated synthesis method as introduced in [13]. A self-channel interaction (SCI) block [13] is exploited to incorporate the channel-wise interdependencies among the outputs from LCFI++ modules. The resulting features are then used to construct multiple intermediate reconstructions G^i and corresponding attention maps M^i. We adopt residual-in-residual dense blocks (RRDB) [14] in the generation of image maps to further enhance the reconstruction quality. M^i and G^i are later combined to produce the final output G'' as follows,

$$G'' = (M^1 \otimes G^1) \oplus \cdots \oplus (M^N \otimes G^N), \qquad (2)$$

where N is the number of attention pairs and $\otimes$ and $\oplus$ denote the element-wise multiplication and addition.

2.2.3. Discriminator

Conventional discriminators are trained to differentiate the synthetic date from real ones, which can fail to model local textures [15]. The Markovian discriminator is introduced in [8] to encourage high-frequency components by convolutionally focusing on pixels in a fixed perceptual field, i.e. at the scale of patches. However, it is hard to select a suitable patch size in practice. Hence, we conform to the strategy in [8] and utilise a multi-scale patch-based discriminator comprising 3 sub-networks with a shared structure to distinguish the patches at different scales. The discriminator is jointly trained with the generator in an adversarial learning diet.

2.3. Objective function

We describe the loss functions used in the training phase in the following. We arithmetically combine the pixel-wise metric and structure-based index with the GAN-based objective. We encourage visually appealing results via a perceptual metric.

2.3.1. Reconstruction loss

We leverage the L_1-norm and multi-scale structural similarity index metric (MS-SSIM) [18] to enforce the generation to be favorable to the fully-sampled reference s. MS-SSIM is proved to be effective to preserve the contrast of high-frequency components, while L_1 can maintain luminance. The mixed loss function is then given as follows,

$$\begin{aligned} L_{rec} =& \lambda_{rec}((1 - \alpha)(L_1(G'', s) + \omega \sum L_1(M^i \otimes G^i, \\ & M^i \otimes s)) + \alpha L^{SSIM}(G'', s) + \beta L_1(G', s)). \end{aligned} \qquad (3)$$

2.3.2. Adversarial loss

Generative adversarial networks (GAN) [5] are proved to be effective to synthesize photo-realistic images, by leveraging

a discriminator to differentiate real data from their synthetic counterparts and a generator to fool the discriminator. We use a least squares GAN (LSGAN) [19] to prevent the saturation during training. Empirically, LSGAN shows more stable reconstructions and faster convergences compared to Vallina GAN. The loss function is shown below,

$$\begin{aligned} L_{adv}^D &= E[\|D(s) - b\|_2^2] + E[\|D(G'') - a\|_2^2] \\ L_{adv}^G &= E[\|D(G'') - c\|_2^2], \end{aligned} \qquad (4)$$

where D denotes the discriminator, and the hyper-parameters b, a, and c are selected to be $b = c = 1$ and $a = 0$.

2.3.3. Data consistency

Data consistency aims to enforce the outputs of the trainable model to agree with the observed measurements. We compute a consistency loss in k-space as it was also suggested in [9]. An alternative is to apply it in image domain as in [16]. We found that the two methods are not fundamentally different and in simulations they behave very similarly. The consistency loss is given as follows,

$$L_{cyc} = \lambda_{cyc}\|y - \mathcal{H} \circ \mathcal{F}(G'')\|_1. \qquad (5)$$

2.3.4. Perceptual loss

In addition to the aforesaid losses, we introduce the perceptual loss to improve the visual quality of the generated images. We utilize a pre-trained network, often a VGG pipeline [20], to map the images into the feature spaces which are more consistent with the human visual judgement. The perceptual difference is given by,

$$\begin{aligned} L_{vgg} =& \lambda_{vgg} \sum \|f_{vgg}^i(G'') - f_{vgg}^i(s)\|_2^2 \\ & + \gamma\|f_{gram}^i(G'') - f_{gram}^i(s)\|_2^2, \end{aligned} \qquad (6)$$

where f_{vgg}^i denotes the pre-activations of the i-th layer in VGG and f_{gram}^i represents their Gram matrix.

3. SIMULATIONS

We extract 2800 images from the NYU knee MRI database [21] for training and 164 samples for test. All images are resized to 256×256. We use two channels to represent the real and imaginary parts of complex-valued images. We adopt a fixed random sampling mask at a rate of 12.5%. We apply the inverse Fourier transform to y to produce the zero-filled z, which is the input to the generator. Our code will be provided at https://github.com/JingshuaiLiu/ASGAN.

3.1. Qualitative analysis

We compare the results of the proposed framework with other two state-of-the-art deep learning-based methods. A deep
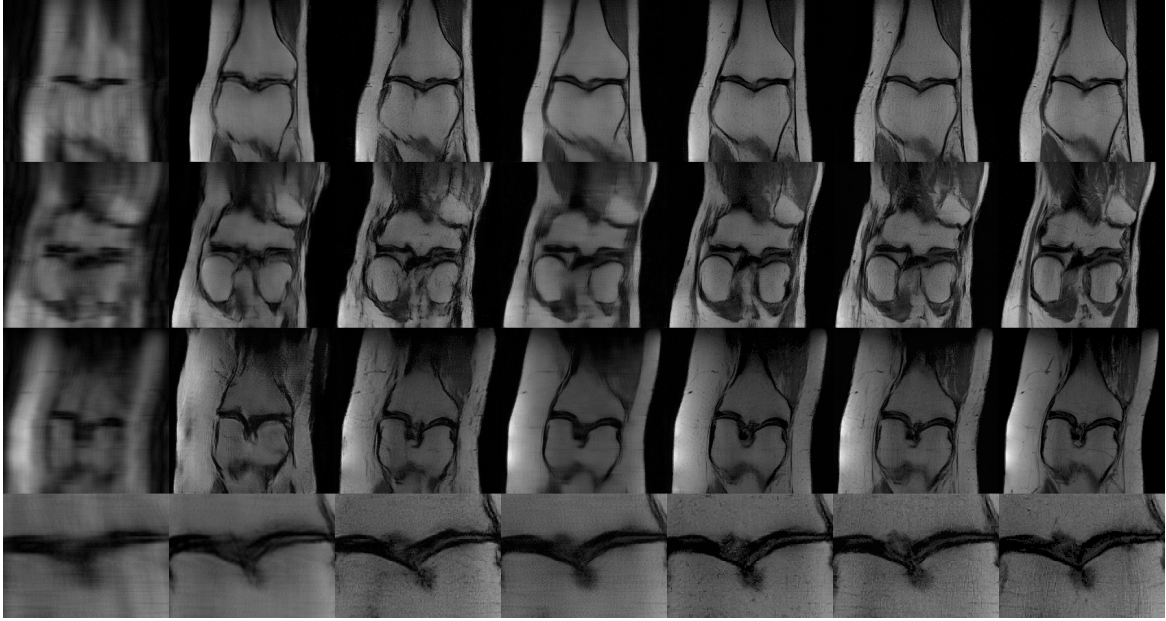
Fig. 2. Comparison of $8\times$ accelerated MRI reconstructions; first) zero-filled, second) CycleGAN-unsup [16], third) LCFI++ (ours) with half of the dataset, fourth) MICCAN [17], fifth) CycleGAN-sup [16], sixth) LCFI++ (ours), and last) ground truth.

Table 1. Quantitative Evaluation

method	PSNR↑	SSIM↑	FID↓	KID↓
proposed	25.45	0.638	104.34	0.036
CycleGAN-sup	25.62	0.655	143.36	0.093
MICCAN	26.61	0.642	180.66	0.146
CycleGAN-unsup	22.03	0.589	209.31	0.184

neural network with channel-wise attention modules (MIC-CAN) for MRI reconstruction is introduced in [17]. We implement the method proposed in [16], denoted by CycleGAN-unsup, which trains a CycleGAN [8] to recover the measurements in an "unsupervised" manner. We also present the outputs of the CycleGAN-based framework trained under supervision (CycleGAN-sup), i.e. using the whole fully-sampled dataset, which is presented in [16] as a reference method.

To further testify the performance of the proposed model, we train the pipeline only with half the training data and compare it with CycleGAN-unsup which uses the same number of high-quality samples. As CycleGAN-unsup has access to the other half of database in low-resolution, a comparison with the proposed method using only half of the dataset is fair or in the favour of CycleGAN. We display all comparison results in Figure 2. We can observe that the proposed model produces sharper and more detailed reconstructions than CycleGAN-unsup and MICCAN, and the generated images are more natural and relevant to full reconstructions than CycleGAN-sup. In the zoomed view, i.e. bottom row, we can observe that the proposed framework recovers finer textures, which confirms

that our method can achieve superior fine-grained reconstruction with more complicated local details in those examples, while preserving the global and structural information.

3.2. Quantitative analysis

We draw a quantitative analysis using PSNR and SSIM as evaluation metrics. We adopt the Fréchet inception distance (FID) and kernel inception distance (KID) [22] to measure the visual quality of reconstructions. Table 1 shows their average scores over all test samples. Overall, our model is competitive with CycleGAN-sup and MICCAN in terms of PSNR and SSIM, in which higher figures are better, and produces superior results than all the other methods with remarkably better FID and KID, in which lower figures are better.

4. CONCLUSIONS AND DISCUSSION

In this paper, we introduce a novel deep learning-based model to achieve fine-grained MRI reconstruction. We leverage a long-range contextual feature integration module with attention selection mechanism in a GAN-based framework to recover the observed measurements. We demonstrate that the proposed framework outperforms other deep neural network-based methods in terms of image quality. The qualitative and quantitative analyses prove its usefulness in providing accurate and fine-grained reconstruction. The future works include incorporating the proposed method with model-based algorithms to provide the convergence guarantee and extending the model to dynamic and parallel imaging.

5. REFERENCES

[1] M.A. Griswold, P.M. Jakob, Robin M. Heidemann, M. Nittka, V. Jellus, J. Wang, B. Kiefer, and A. Haase, "Generalized autocalibrating partially parallel acquisitions (GRAPPA)," *Magnetic resonance in medicine*, vol. 47, no. 6, pp. 1202—1210, June 2002.

[2] S. G. Lingala and M. Jacob, "Blind compressive sensing dynamic MRI," *IEEE Transactions on Medical Imaging*, vol. 32, no. 6, pp. 1132–1145, 2013.

[3] S. Ravishankar and Y. Bresler, "MR image reconstruction from highly undersampled k-space data by dictionary learning," *IEEE Transactions on Medical Imaging*, vol. 30, no. 5, pp. 1028–1041, 2011.

[4] S. Anuroop, Z. Jure, M. Tullie, Z. Lawrence, D. Aaron, and Daniel K.S., "GrappaNet: combining parallel imaging with deep learning for multi-coil MRI reconstruction," *ArXiv*, vol. abs/1910.12325, 2019.

[5] I.J. Goodfellow, A.J. Pouget, M. Mirza, B. Xu, F.D. Warde, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," *arXiv e-prints*, p. arXiv:1406.2661, June 2014.

[6] D. Narnhofer, K. Hammernik, F. Knoll, and T. Pock, "Inverse GANs for accelerated MRI reconstruction," *Wavelets and Sparsity XVIII*, vol. 11138, pp. 111381A, Sept. 2019.

[7] D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Deep image prior," *arXiv e-prints*, p. arXiv:1711.10925, Nov. 2017.

[8] J.Y. Zhu, T. Park, P. Isola, and A.A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," *arXiv e-prints*, p. arXiv:1703.10593, Mar. 2017.

[9] T.M. Quan, D.T. Nguyen, and W.K. Jeong, "Compressed sensing MRI reconstruction using a generative adversarial network with a cyclic loss," *arXiv e-prints*, p. arXiv:1709.00753, Sept. 2017.

[10] B. Sim, G. Oh, and J. C. Ye, "Optimal transport structure of CycleGAN for unsupervised learning for inverse problems," *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 8644–8647, 2020.

[11] H. Mei, X. Yang, Y. Wang, Y. Liu, S. He, Q. Zhang, X. Wei, and R.W.H. Lau, "Don't hit me! Glass detection in real-world scenes," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020.

[12] S. Woo, J. Park, J.Y. Lee, and I.S. Kweon, "CBAM: convolutional block attention module," *arXiv e-prints*, p. arXiv:1807.06521, July 2018.

[13] H. Tang, D. Xu, Y. Yan, J.J. Corso, P.H.S. Torr, and N. Sebe, "Multi-channel attention selection GANs for guided image-to-image translation," *arXiv e-prints*, p. arXiv:2002.01048, Feb. 2020.

[14] K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, C.C. Loy, Y. Qiao, and X. Tang, "ESRGAN: enhanced super-resolution generative adversarial networks," *CoRR*, vol. abs/1809.00219, 2018.

[15] C. Li and M. Wand, "Precomputed real-time texture synthesis with Markovian generative adversarial networks," *arXiv e-prints*, p. arXiv:1604.04382, Apr. 2016.

[16] B. Sim, G. Oh, J. Kim, C. Jung, and J.C. Ye, "Optimal transport, CycleGAN, and penalized LS for unsupervised learning in inverse problems," *arXiv e-prints*, p. arXiv:1909.12116, Sept. 2019.

[17] Q. Huang, D. Yang, P. Wu, H. Qu, J. Yi, and D. Metaxas, "MRI reconstruction via cascaded channel-wise attention network," *arXiv e-prints*, p. arXiv:1810.08229, Oct. 2018.

[18] H. Zhao, O. Gallo, I. Frosio, and J. Kautz, "Loss functions for image restoration with neural networks," *IEEE Transactions on Computational Imaging*, vol. 3, no. 1, pp. 47–57, 2017.

[19] X. Mao, Q. Li, H. Xie, R.Y.K. Lau, Z. Wang, and S.P. Smolley, "Least squares generative adversarial networks," *arXiv e-prints*, p. arXiv:1611.04076, Nov. 2016.

[20] L. A. Gatys, A. S. Ecker, and M. Bethge, "Image style transfer using convolutional neural networks," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2414–2423, June 2016.

[21] J. Zbontar, F. Knoll, A. Sriram, M.J. Muckley, M. Bruno, A. Defazio, M. Parente, K.J. Geras, J. Katsnelson, H. Chandarana, Z. Zhang, M. Drozdzal, A. Romero, M. Rabbat, P. Vincent, J. Pinkerton, D. Wang, N. Yakubova, E. Owens, C.L. Zitnick, M.P. Recht, D.K. Sodickson, and Y.W. Lui, "FastMRI: an open dataset and benchmarks for accelerated MRI," *CoRR*, vol. abs/1811.08839, 2018.

[22] M. Bińkowski, D.J. Sutherland, M. Arbel, and A. Gretton, "Demystifying MMD GANs," in *International Conference on Learning Representations*, 2018.