Eurographics Workshop on Visual Computing for Biology and Medicine (2019)
K. Lawonn and R. G. Raidou (Editors)

Short Paper

Multiparametric Magnetic Resonance Image Synthesis using Generative Adversarial Networks

Christoph Haarburger[1], Nicolas Horst[1], Daniel Truhn[2,1], Mirjam Broeckmann[2], Simone Schrading[2], Christiane Kuhl[2] and Dorit Merhof[1]

[1]Insitute of Imaging and Computer Vision, RWTH Aachen University, Germany
[2]Department of Diagnostic and Interventional Radiology, University Hospital Aachen, Germany

Abstract

Generative adversarial networks have been shown to alleviate the problem of limited training data for supervised learning problems in medical image computing. However, most generative models for medical images focus on image-to-image translation rather than de novo image synthesis. In many clinical applications, image acquisition is multiparametric, i.e. includes contrast-enchanced or diffusion-weighted imaging. We present a generative adversarial network that synthesizes a sequence of temporally consistent contrast-enhanced breast MR image patches. Performance is evaluated quantitatively using the Fréchet Inception Distance, achieving a minimum FID of 21.03. Moreover, a qualitative human reader test shows that even a radiologist cannot differentiate between real and fake images easily.

CCS Concepts

• ***Computing methodologies*** → *Modeling methodologies;*

1. Introduction

Contemporary methods of medical image segmentation and classification are mostly driven by machine learning algorithms that learn from data in a supervised fashion. This is problematic in several ways: Firstly, supervised learning algorithms rely on large amounts of annotated data. Especially with clinical data, requiring expert knowledge for annotation, the annotation process is costly and time-consuming. As a consequence, algorithms are often trained and evaluated on datasets that are too small to unfold the full potential of a neural network. Secondly, domain shift limits the applicability of many machine learning models. The implicit assumption of supervised learning that training and test data arise from the same distribution as real-world data often does not hold, which leads to poor inference performance.

Generative adversarial networks (GANs) [GPAM*14] have the potential to solve these problems by modelling the latent distribution of the training data effectively. In medical image computing, GANs have successfully been applied to image-to-image translation [WLVI17, OSDU18], segmentation [KBS*17], synthesis [WLI18] and many other subfields. Furthermore, it has been shown that supervised training of neural networks for segmentation [**?**, **?**] and classification [FAKA*18, MMKSM18] with additional synthetic training data improves prediction performance on real test data.

Clinical acquisition protocols often involve MR imaging with several contrasts or multiparametric imaging such as dynamic contrast-enhanced (DCE) or diffusion-weighted imaging (DWI). For this kind of data, there are very few works on image synthesis using data-driven generative models so far. Beers et al. [BBC*18] trained a Progressive Growing GAN [KALL17] to generate several contrasts of brain MR images. The vast majority of other works essentially performs image-to-image translation rather than *de novo* synthesis [WKLI18]. However, a model capable of synthesizing a set of images that is consistent across contrasts and parametric maps would be very helpful for image segmentation and classification. Potential applications may be breast or prostate MR-imaging that both notoriously suffer from small datasets and high biological variance.

Therefore, in this work we present a setup that is able to synthesize consistent DCE-MR image patches (64×64 crops) of the breast *de novo* that could be utilized to improve breast lesion malignancy classification [TSH*19] in the future. Our contributions is that we extend two GAN algorithms, deep convolutional GANs (DCGANs) and Wasserstein GANs (WGANs), to synthesize dynamic contrast-enhanced MR images conditioned on healthy tissue, benign lesions and malignant lesions, respectively.

2. Data

Our dataset consists of multiparametric contrast-enhanced bilateral breast MR images of 408 patients. All images were acquired according to a clinical routine standard protocol [KSB*17], which consists of a T2-weighted turbo spin-echo sequence (acquisition matrix 512×512) and a T1-weighted gradient echo dynamic series, which are both acquired in axial orientation. The dynamic series is

DOI: 10.2312/vcbm.20191226

comprised of one precontrast image and four images acquired every 70 seconds after administration of contrast agent.

All suspicious lesions were manually segmented on every slice by a radiologist with 13 years of experience using an in-house developed software. Malignancy was determined based on histology or 12 month follow-up if no histology was available. More details on the dataset are provided in Tab. 1.

Subjects	Lesions		Patches		
	benign	malignant	benign	malignant	healthy
408	478	664	1149	2206	401,525

Table 1: *Description of the dataset.*

3. Methods

We aimed to keep the preprocessing pipeline as simple as possible to ensure that the synthesized images closely resemble images in clinical practice. To this end, we did *not* perform a bias field correction but simply rescaled image intensities to a fixed range of [0, 1023]. Since all data originates from a highly standardized protocol, i.e. slice thickness and in-plane resolution are approximately the same for all training data, the training images were resampled to a fixed resolution of $512 \times 512 \times 32$.

To generate MR images that are useful for a future lesion classification, axial 64×64 patches were extracted as follows:

1. Patches of benign and malignant lesions were extracted by cropping based on the centers of mass of the segmentation masks of individual lesions.
2. Patches of healthy tissue were extracted by random sampling of centerpoints from locations inside the breast with no clinically relevant findings in a 32×32 neighborhood. Following that procedure we extracted 401,525 patches in total, which is equivalent to sampling 984 patches from each patient or 15 patches per breast and axial slice on average. The resulting patches partially overlap which can be interpreted as an augmentation of the dataset. A summary is provided in Tab. 1. An example set of patches is given in Fig. 1.

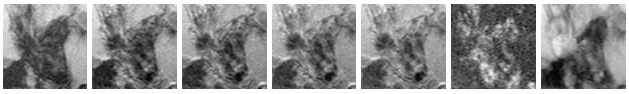
(a) $\mathrm{T1}_{\mathrm{pre}}$ (b) $\mathrm{T1}_{\mathrm{post1}}$ (c) $\mathrm{T1}_{\mathrm{post2}}$ (d) $\mathrm{T1}_{\mathrm{post3}}$ (e) $\mathrm{T1}_{\mathrm{post4}}$ (f) Sub (g) T2

Figure 1: *Contrasts of an example patch depicting an invasive carcinoma. The* $_{pre}$ *and* $_{post}$ *indices refer to images that were acquired prior or after administration of contrast agent. The subtraction image (Sub) is calculated by Sub* $= \mathrm{T1}_{\mathrm{post1}} - \mathrm{T1}_{\mathrm{pre}}$.

The pace of progress in generative models is fast. It has been shown that there is no clear consensus concerning which GAN algorithm is better than another and many of the recently proposed models achieve comparable performance [LKM*18]. For this reason we compare deep convolutional GAN (DCGAN) [RMC15] because of it's simplicity and Wasserstein GANs (WGANs) [ACB17] becaue of it's successfull prior application in medical image synthesis [WLI18].

3.1. Deep Convolutional GAN

The DCGAN framework as proposed in [RMC15] is a baseline model. The discriminator D is composed of four layers of 2D strided convolutions, batch normalization, LeakyReLU activation functions and a fully-connected layer. The generator G consists of four layers of upconvolutions, batch normalization, ReLU activation and finally an upconvolution with tanh activation function. In contrast to the original implementation, we modified the upconvolutions to the "better upconvolution" as proposed in [ODO16] to reduce checkerboard artifacts. Furthermore, we changed the generator output to have seven channels, i.e. one channel for each MR contrast or DCE time point. We denote the real and generated data distributions as $\mathbb{P}_r$ and $\mathbb{P}_g$. The generator G generates new samples $\tilde{\mathbf{x}} = G(\mathbf{z})$ from a random noise vector $\mathbf{z} \sim p(\mathbf{z})$ following a Gaussian distribution and the optimization objective is defined as

$$\min_G \max_D \mathop{\mathbb{E}}_{\mathbf{x}\sim\mathbb{P}_r}[\log D(\mathbf{x})] - \mathop{\mathbb{E}}_{\tilde{\mathbf{x}}\sim\mathbb{P}_g}[\log(D(\tilde{\mathbf{x}}))]. \tag{1}$$

Optimization was performed using a batch size of 64 and Adam optimizer with all hyperparameters set as suggested in [RMC15].

3.2. Wasserstein GANs

In Wasserstein GANs [ACB17], the generator is trained to minimize an approximation of the Earth Mover's (EM) distance between the distributions of real and synthetic data. Loosely speaking, the EM distance can be interpreted as moving a pile of dirt following a certain distribution into another distribution at minimum cost. In this setting, the discriminator no longer classifies samples as real or fake but approximates the EM distance. As compared to the original GAN loss in Eq. 1, the Wasserstein GAN framework propagates stronger gradients to the generator, which reduces the risk of mode collapse during training [ACB17]. The objective then becomes

$$\min_G \max_{D\in\mathcal{D}} \mathop{\mathbb{E}}_{\mathbf{x}\sim\mathbb{P}_r}[D(x)] - \mathop{\mathbb{E}}_{\tilde{\mathbf{x}}\sim\mathbb{P}_g}[D(\tilde{\mathbf{x}}))]. \tag{2}$$

However, in order to approximate the EM distance, the discriminator is required to be Lipschitz continuous. This was initially implemented by clipping the weights of the discriminator, which has several drawbacks: Firstly, clipping weights effectively prevents the network from learning and using it's full capacity. Secondly, a new hyperparameter, the clip value c, is introduced that is very sensitive to small changes. However, the authors claim that despite these drawbacks, the WGAN practically provides high quality results. As suggested by Arjovsky et al. [ACB17], we set the update ratio of discriminator updates per generator update to five.

To overcome the aforementioned drawbacks of WGANs, adding a gradient penalty to the Wasserstein loss was proposed in [GAA*17], which regularizes the loss from Eq. 2 to

$$\min_G \max_{D\in\mathcal{D}} \mathop{\mathbb{E}}_{\tilde{\mathbf{x}}\sim\mathbb{P}_g}[D(\tilde{\mathbf{x}}))] - \mathop{\mathbb{E}}_{\mathbf{x}\sim\mathbb{P}_r}[D(x)] + \lambda \mathop{\mathbb{E}}_{\hat{\mathbf{x}}\sim\mathbb{P}_{\hat{\mathbf{x}}}}[(\|\nabla_{\hat{\mathbf{x}}} D(\hat{\mathbf{x}})\|_2 - 1)^2], \tag{3}$$

where $\mathbb{P}_{\hat{\mathbf{x}}}$ is the distribution of linear interpolations between pairs of samples from the data and generator distribution, respectively. This enforces a local Lipschitz continuity and eliminates the need for weight clipping. However, in experiments that are beyond the scope of this work we found that the Wasserstein loss with gradient penalty as in Eq. 3 was extremely sensitive to the the weighting factor λ, which lead to unstable and diverging training. Therefore, all experiments in this work are based on the Wasserstein loss as introduced in Eq. 1 despite it's theoretical shortcomings.

In order to generate images that are coherent across MRI contrasts and time points, we adapt the channel dimensions of our generator networks to match the number of channels, i.e. contrasts and time points for 2D+t data. We trained the WGAN with 64 samples per batch using the RMSProp optimizer and set all hyperparameters as suggested in [GAA*17].

In order to generate patches of the three relevant patch classes, we conditioned the GANs with a condition $\mathbf{y} \in \{\text{benign}, \text{malignant}, \text{no findings}\}$ [MO14]. This is implemented by concatenating the one-hot-encoded vector $\mathbf{y}$ to the noise vector $\mathbf{z}$.

4. Experiments

For evaluating synthetic data, several aspects need to be considered: Obviously, image quality needs to be assessed. To this end, we evaluate Fréchet Inception Distance (FID) [HRU*17] and perform a human reader test. Moreover, we assess image diversity by means of calculating the 1-nearst-neighbor (1-NN) classification accuracy as suggested in [XHY*18].

4.1. Quantitative Evaluation

Quantitaive evaluation of image synthesis is challenging because the field still lacks a widely-accepted measure. Previously proposed distance measures such as the Inception Distance (ID) and the Fréchet Inception Distance (FID) are constructed for natural RGB images rather than medical images.

Fréchet Inception Distance [HRU*17] has been shown to be superior in terms of consistency to the previously proposed Inception score [SGZ*16]. In FID, the Inception-v3 network is utilized to extract activations from the `pool3` layer. With $X_r \sim \mathcal{N}(\mu_r, \Sigma_r)$ and $X_g \sim \mathcal{N}(\mu_g, \Sigma_g)$ denoting the distributions of those features for real and generated data, respectively, FID is defined as

$$\text{FID} = ||\mu_r - \mu_g||^2 + \text{Tr}(\Sigma_r + \Sigma_g - 2(\Sigma_r \Sigma_g)^{1/2}). \quad (4)$$

More similar activation distributions lead to lower values for FID. Hence, smaller values correspond to higher similarity. To make our data fit the Inception-v3 network we selected the T2, T1_{pre} and T1_{post1} images as R, G and B channel, respectively. As Kuhl et al. reported in [KSS*14], diagnostic accuracy based on this abbreviated protocol is equivalent to the full protocol. To calculate the distribution statistics we used 1000 randomly chosen real samples of the dataset and an equal amount of generated samples. We ran the calculation with 10 different sets of real and fake samples and averaged the scores.

FID scores for all models are depicted in Tab. 2. The best FID of 20.23 is achieved by training a DCGAN on malignant patches. Consistently, the DCGAN achieves better FIDs than the WGAN.

		FID_{min}	1-NN real	1-NN fake
benign	DCGAN	**27.14**	**0.272**	**0.221**
	WGAN	33.90	0.229	0.166
malignant	DCGAN	**20.23**	**0.306**	**0.268**
	WGAN	24.51	0.264	0.21
no findings	DCGAN	**21.03**	**0.428**	**0.223**
	WGAN	29.48	0.337	0.151

Table 2: *FID and 1-NN accuracies for DCGAN and WGAN for benign, malignant and healthy tissue. All results are based on the "best" training epoch determined by the lowest FID value (FID_{min}).*

1-Nearest-Neighbor accuracy has been proposed in [LPO16] to measure quality of generated images [XHY*18]. With two sets, $S_r \sim \mathbb{P}_r$ consisting of real samples and $S_g \sim \mathbb{P}_g$ of fake samples, we assign S_r positive labels and S_g negative labels, respectively. Then, the accuracy of a 1-nearest-neighbor classifier trained on S_r and S_g is computed for the two sets seperately. In theory, if the images are perfectly diverse, the 1-NN classifier achieves 50% accuracy on both sets. 1-NN accuracies for our experiments with $|S_r| = |S_g| = 1000$ are depicted in Tab. 2. The best 1-NN accuracies are achieved by the DCGAN.

4.2. Human Reader Test

For a systematic approach to visual human inspection, we asked a radiologist and a layperson with no technical or medical background to classify sets of images as real or fake. To this end, we presented a T1 dynamic series with corresponding subtraction images as well as the T2 image to the human raters as depticted in Fig. 4 and 5. For DCGAN and WGAN, respectively, we generated 40 random samples from the generator and drew 40 random samples of real data for each benign, malignant and healthy patches. In total this leads to 240 real and 240 fake images that were assessed by the human readers. The experiment was started with a warm up phase, in which the human readers were shown 24 real and 24 fake samples and their respective origin to get used to the task. Results of the human reader tests are listed in Tab. 3. The radiologist performed about 10 % better than the layperson.

	Real	DCGAN	WGAN
Layperson	64.6%	62.5%	70.0%
Radiologist	73.3%	75.0%	76.7%

Table 3: *Fraction of correctly classified samples from human reader tests.*

Example images for both DCGAN and WGAN are depicted in Fig. 2 and 3. Additional patches showing the temporal and structural consistency accross time points in the dynamic series and contrasts are provided in Fig. 4 and 5. The images are structurally highly consistent across time points and contrasts. As expected, some parts of the image show a contrast agent enhancement, which mostly occurs between $\mathrm{T1}_{\mathrm{pre}}$ and $\mathrm{T1}_{\mathrm{post1}}$.

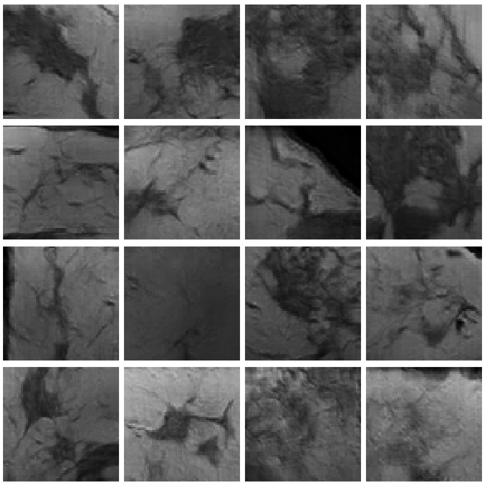

Figure 2: *Example patches for fake malignant native T1 images generated by a DCGAN.*

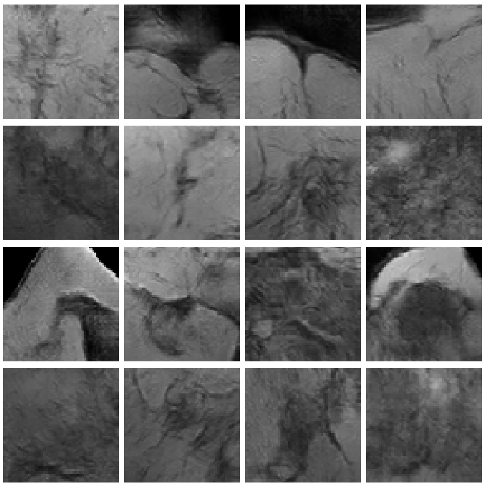

Figure 3: *Example patches for fake malignant native T1 images generated by a WGAN.*

5. Discussion

We present a method for synthesis of multiparametric breast MR image patches. Our results show that both DCGAN and WGAN are able to generate sequences of T2 and T1-DCE image patches that are spatially and temporally coherent, such that even a radiologist cannot differentiate between real and fake images easily.

FID values are in a range that is comparable to other works reporting FID for DCGAN and WGAN based on ImageNet and CIFAR-10 datasets [ODM18]. In terms of FID and 1-NN accuracies, the DCGAN is superior to WGAN. Interestingly, 1-NN accuracy and FID are strongly correlated which may indicate that FID actually captures image quality for multiparametric MR images despite the fact that it was designed for RGB images. Moreover, we observed that tuning WGAN-specific hyperparameters can be challengig and may lead to unstable training.

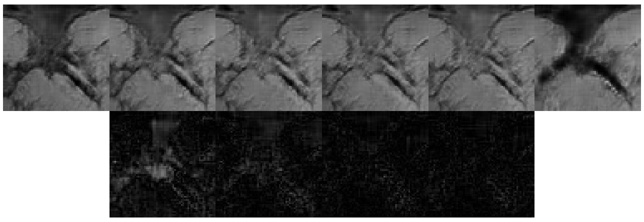

Figure 4: *Sequence of fake benign T1 dynamic series (top five left) and T2 image (top right) and subtraction images of T1 dynamic series (bottom) generated by DCGAN.*

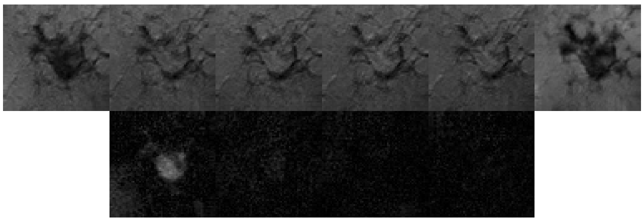

Figure 5: *Sequence of fake malignant T1 dynamic series (top seven left) and T2 image (top right) and subtraction images of T1 dynamic series (bottom) generated by WGAN.*

The human readers are able to differentiate between real and fake images in most cases. The radiolgist could identify real and fake images with comparable accuracies with only a minor difference between fake images from DCGAN and WGAN, respectively. In order to fool the radiologist, the images not only need to look morphologically realistic but also the enhancement pattern needs to appear highly coherent and realistic. We observed that the GANs even produced common artifacts such as fat-shift artifacts.

Our work has several limitations: Firstly, the FID is designed for natural RGB images rather than multiparametric MR images. Since there is no accepted quality measure for these type of images we decided to evaluate using the FID despite this limitation and add 1-NN accuracy as an additional measure. Moreover, the human reader test was performed by only two raters which limits it's validity.

For future work, other GAN frameworks such das Progressive Growing GANs [KALL17] could be used in the future to synthesize not only patches but whole axial slices of MR images. Lastly, the influence of the synthesized image patches on a breast lesion classifier as in [TSH*19] will be evaluated.

6. Conclusion

The evaluated GANs achieve similar performance in producing realistic multiparametric MR images of the breast. The synthesized images are diverse and realistic to a degree that they may help for training a classifier in the future.

References

[ACB17] ARJOVSKY M., CHINTALA S., BOTTOU L.: Wasserstein GAN, 2017. URL: https://arxiv.org/abs/1701.07875, arXiv:arXiv:1701.07875. 2

[BBC*18] BEERS A., BROWN J., CHANG K., CAMPBELL J. P., OSTMO S., CHIANG M. F., KALPATHY-CRAMER J.: High-resolution medical image synthesis using progressively grown generative adversarial networks, 2018. URL: http://arxiv.org/abs/1805.03144, arXiv:arXiv:1805.03144. 1

[FAKA*18] FRID-ADAR M., KLANG E., AMITAI M., GOLDBERGER J., GREENSPAN H.: Synthetic Data Augmentation using GAN for Improved Liver Lesion Classification, 2018. URL: https://arxiv.org/abs/1801.02385, arXiv:arXiv:1801.02385. 1

[GAA*17] GULRAJANI I., AHMED F., ARJOVSKY M., DUMOULIN V., COURVILLE A.: Improved Training of Wasserstein GANs, 2017. URL: http://arxiv.org/abs/1704.00028, arXiv:arXiv:1704.00028. 2, 3

[GPAM*14] GOODFELLOW I. J., POUGET-ABADIE J., MIRZA M., XU B., WARDE-FARLEY D., OZAIR S., COURVILLE A., BENGIO Y.: Generative Adversarial Networks. URL: https://arxiv.org/abs/1406.2661, arXiv:arXiv:1406.2661. 1

[HRU*17] HEUSEL M., RAMSAUER H., UNTERTHINER T., NESSLER B., HOCHREITER S.: Gans trained by a two time-scale update rule converge to a local nash equilibrium. In *Advances in Neural Information Processing Systems* (2017), pp. 6626–6637. 3

[KALL17] KARRAS T., AILA T., LAINE S., LEHTINEN J.: Progressive Growing of GANs for Improved Quality, Stability, and Variation, 2017. URL: https://arxiv.org/abs/1710.10196, arXiv:arXiv:1710.10196. 1, 4

[KBS*17] KOHL S., BONEKAMP D., SCHLEMMER H.-P., YAQUBI K., HOHENFELLNER M., HADASCHIK B., RADTKE J.-P., MAIER-HEIN K.: Adversarial Networks for the Detection of Aggressive Prostate Cancer, 2017. URL: http://arxiv.org/abs/1702.08014, arXiv:arXiv:1702.08014. 1

[KSB*17] KUHL C. K., STROBEL K., BIELING H., LEUTNER C., SCHILD H. H., SCHRADING S.: Supplemental breast mr imaging screening of women with average risk of breast cancer. *Radiology 283*, 2 (2017), 361–370. PMID: 28221097. doi:10.1148/radiol.2016161444. 1

[KSS*14] KUHL C. K., SCHRADING S., STROBEL K., SCHILD H. H., HILGERS R.-D., BIELING H. B.: Abbreviated breast magnetic resonance imaging (mri): First postcontrast subtracted images and maximum-intensity projection—a novel approach to breast cancer screening with mri. *Journal of Clinical Oncology 32*, 22 (2014), 2304–2310. doi:10.1200/JCO.2013.52.5386. 3

[LKM*18] LUCIC M., KURACH K., MICHALSKI M., GELLY S., BOUSQUET O.: Are gans created equal? a large-scale study. In *Advances in Neural Information Processing Systems 31*, Bengio S., Wallach H., Larochelle H., Grauman K., Cesa-Bianchi N., Garnett R., (Eds.). Curran Associates, Inc., 2018, pp. 700–709. URL: http://papers.nips.cc/paper/7350-are-gans-created-equal-a-large-scale-study.pdf. 2

[LPO16] LOPEZ-PAZ D., OQUAB M.: Revisiting Classifier Two-Sample Tests, 2016. URL: http://arxiv.org/abs/1610.06545, arXiv:arXiv:1610.06545. 3

[MMKSM18] MADANI A., MORADI M., KARARGYRIS A., SYEDA-MAHMOOD T.: Semi-supervised learning with generative adversarial networks for chest x-ray classification with ability of data domain adaptation. In *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)* (April 2018), pp. 1038–1042. doi:10.1109/ISBI.2018.8363749. 1

[MO14] MIRZA M., OSINDERO S.: Conditional Generative Adversarial Nets, 2014. URL: http://arxiv.org/abs/1411.1784, arXiv:arXiv:1411.1784. 3

[ODM18] OSTROVSKI G., DABNEY W., MUNOS R.: Autoregressive Quantile Networks for Generative Modeling, 2018. URL: http://arxiv.org/abs/1806.05575, arXiv:arXiv:1806.05575. 4

[ODO16] ODENA A., DUMOULIN V., OLAH C.: Deconvolution and checkerboard artifacts. *Distill* (2016). URL: http://distill.pub/2016/deconv-checkerboard, doi:10.23915/distill.00003. 2

[OSDU18] OLUT S., SAHIN Y. H., DEMIR U., UNAL G.: Generative Adversarial Training for MRA Image Synthesis Using Multi-Contrast MRI, 2018. URL: http://arxiv.org/abs/1804.04366, arXiv:arXiv:1804.04366. 1

[RMC15] RADFORD A., METZ L., CHINTALA S.: Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. URL: https://arxiv.org/abs/1511.06434, arXiv:arXiv:1511.06434. 2

[SGZ*16] SALIMANS T., GOODFELLOW I., ZAREMBA W., CHEUNG V., RADFORD A., CHEN X., CHEN X.: Improved techniques for training gans. In *Advances in Neural Information Processing Systems 29*, Lee D. D., Sugiyama M., Luxburg U. V., Guyon I., Garnett R., (Eds.). Curran Associates, Inc., 2016, pp. 2234–2242. URL: http://papers.nips.cc/paper/6125-improved-techniques-for-training-gans.pdf. 3

[TSH*19] TRUHN D., SCHRADING S., HAARBURGER C., SCHNEIDER H., MERHOF D., KUHL C.: Radiomic versus convolutional neural networks analysis for classification of contrast-enhancing lesions at multiparametric breast mri. *Radiology* (2019). doi:10.1148/radiol.2018181352. 1, 4

[WKLI18] WOLTERINK J. M., KAMNITSAS K., LEDIG C., IŠGUM I.: Generative adversarial networks and adversarial methods in biomedical image analysis, 2018. URL: http://arxiv.org/abs/1810.10352, arXiv:arXiv:1810.10352. 1

[WLI18] WOLTERINK J. M., LEINER T., ISGUM I.: Blood Vessel Geometry Synthesis using Generative Adversarial Networks, 2018. URL: http://arxiv.org/abs/1804.04381, arXiv:arXiv:1804.04381. 1, 2

[WLVI17] WOLTERINK J. M., LEINER T., VIERGEVER M. A., ISGUM I.: Generative adversarial networks for noise reduction in low-dose ct. *IEEE Transactions on Medical Imaging PP*, 99 (2017), 1–1. doi:10.1109/TMI.2017.2708987. 1

[XHY*18] XU Q., HUANG G., YUAN Y., GUO C., SUN Y., WU F., WEINBERGER K.: An empirical study on evaluation metrics of generative adversarial networks, 2018. URL: http://arxiv.org/abs/1806.07755, arXiv:arXiv:1806.07755. 3