Abstract—In order to objectively assess new medical imaging technologies via computer-simulations, it is important to account for all sources of variability that contribute to image data. One important source of variability that can significantly limit observer performance is associated with the variability in the ensemble of objects to-be-imaged. This source of variability can be described by stochastic object models (SOMs), which are generative models that can be employed to sample from a distribution of to-be-virtually-imaged objects. It is generally desirable to establish SOMs from experimental imaging measurements acquired by use of a well-characterized imaging system, but this task has remained challenging. Deep generative neural networks, such as generative adversarial networks (GANs) hold potential for such tasks. To establish SOMs from imaging measurements, an AmbientGAN has been proposed that augments a GAN with a measurement operator. However, the original AmbientGAN could not immediately benefit from modern training procedures and GAN architectures, which limited its ability to be applied to realistically sized medical image data. To circumvent this, in this work, a modified AmbientGAN training strategy is proposed that is suitable for modern progressive or multi-resolution training approaches such as employed in the Progressive Growing of GANs and Style-based GANs. AmbientGANs established by use of the proposed training procedure are systematically validated in a controlled way by use of computer-simulated measurement data corresponding to a stylized imaging system. Finally, emulated single-coil experimental magnetic resonance imaging data are employed to demonstrate the methods under less stylized conditions.

Index Terms—Objective assessment of image quality, stochastic object models, generative adversarial networks.

I. INTRODUCTION

COMPUTER-simulation remains an important approach for the design and optimization of imaging systems. Such approaches can permit the exploration, refinement, and assessment of a variety of system designs that would be infeasible through experimental studies alone [1]–[3]. In the field of medical imaging, it has been advocated that imaging systems and reconstruction algorithms should be assessed and optimized by use of objective measures of image quality (IQ) that quantify the performance of an observer at specific diagnostic tasks [4]–[8]. To accomplish this, all sources of variability in the measured data should be accounted for. One important source of variability that can significantly limit observer performance is variation in the objects to-be-imaged [9]. This source of variability can be described by stochastic object models (SOMs) [10]. A SOM is a generative model that can be employed to produce an ensemble of to-be-imaged objects that possess prescribed statistical properties.

Available SOMs include texture models of mammographic images with clustered lumpy backgrounds [11], simple lumpy background models [9], and more realistic anatomical phantoms that can be randomly perturbed [12]. A variety of other computational phantoms [12]–[19], either voxelized or mathematical, have been proposed for medical imaging simulation, aiming to provide a practical solution to characterize object variability. However, the majority of these were established by use of image data corresponding to only a few subjects. Therefore, they may not accurately describe the statistical properties of the ensemble of objects that is relevant to an imaging system optimization task. A variety of anatomical shape models have also been proposed to describe both the common geometric features and the geometric variability among instances of the population for shape analysis applications [20]–[27]. To date, these have not been systematically explored for the purpose of constructing SOMs that capture realistic anatomical variations for use in imaging system optimization.

In order to establish SOMs that capture realistic textures and anatomical variations, it is desirable to utilize experimental imaging data. By definition, however, SOMs should be independent of the imaging system, measurement noise and any reconstruction method employed. In other words, they should provide an in silico representation of the ensemble of objects to-be-imaged and not estimates of them that would be indirectly measured or computed by imaging systems. To address this need, Kupinski et al. [10] proposed an explicit generative model for describing object statistics that was trained by use of noisy imaging measurements and a computational model of a well-characterized imaging system [10]. However, applications of this method have been limited to situations where the characteristic functional of the random object can be analytically determined [28], such as with lumpy and clustered lumpy object models [11], [29]. As such, there
remains an important need to generalize the method.

Deep generative neural networks, such as generative adversarial networks (GANs) [30], hold great potential for establishing SOMs that describe finite-dimensional approximations of objects. However, conventional GANs are typically trained by use of reconstructed images that are influenced by the effects of measurement noise and the reconstruction process. To circumvent this, an AmbientGAN has been proposed [31] that augments a GAN with a measurement operator. This permits a generative model that describes object randomness to be learned from indirect and noisy measurements of the objects themselves. In a preliminary study, the AmbientGAN was explored for establishing SOMs from imaging measurements for use in optimizing imaging systems [32]. However, similar to conventional GANs, the process of training AmbientGANs is inherently unstable. Moreover, the original AmbientGAN cannot immediately benefit from robust GAN training procedures, such as progressive growing [33], which limits its ability to synthesize high-dimensional images that depict accurate approximations of objects that are relevant to medical imaging studies.

In this work, modern multi-resolution training approaches, such as employed in the Progressive Growing of GANs (ProGANs) [33] and Style-based GANs (StyGANs) [34], [35], are modified for use in establishing AmbientGANs with high-dimensional medical imaging measurements. The resulting models will be referred to as Progressive Growing AmbientGANs (ProAmGANs) and Style-AmbientGANs (StyAmGANs). Numerical studies corresponding to a stylized imaging system are conducted to systematically investigate the proposed advanced AmbientGAN methods for establishing SOMs. The effects of noise levels and the imaging operator null space characteristics on model performance are assessed by use of both standard and objective measures. Emulated single-coil experimental magnetic resonance imaging data are also employed to demonstrate the method under less stylized conditions.

The remainder of this paper is organized as follows. In Sec. II, previous works on learning SOMs by employing characteristic functions and AmbientGANs are summarized. The progressive growing training strategy for GANs is also reviewed. The proposed ProAmGAN and StyAmGAN for learning SOMs from noisy imaging measurements are described in Sec. III. Sections IV and V describe the numerical studies and results that demonstrate the ability of the advanced AmbientGANs to learn SOMs from noisy imaging measurements. Finally, a discussion and summary of the work is presented in Sec. VI.

II. BACKGROUND

Object properties that are imaged by medical imaging systems are inherently described by continuous functions. However, it is common practice when performing computer-simulation studies of imaging systems to approximate the object by use of a finite-dimensional representation [36], [37]. In such cases, a discrete-to-discrete (D-D) description of a linear imaging system can be described as [7]:

\[ g = Hf + n, \]

where \( g \in \mathbb{R}^M \) is a vector that describes the measured image data, \( f \in \mathbb{R}^N \) denotes the finite-dimensional representation of the object being imaged, \( H \in \mathbb{R}^{M \times N} \) denotes a D-D imaging operator \( \mathbb{R}^N \to \mathbb{R}^M \) that maps an object in the Hilbert space \( \mathbb{U} \) to the measured discrete data in the Hilbert space \( \mathbb{V} \), and the random vector \( n \in \mathbb{R}^M \) denotes the measurement noise. Below, the imaging process described in Eq. (1) is denoted as: \( g = H_n(f) \). In this work, it will be assumed that the D-D imaging model is a sufficiently accurate representation of the true continuous-to-discrete (C-D) imaging model that describes a digital imaging system and the impact of model error will be neglected. Accordingly, as described below, the objective of this work will be to establish SOMs that describe the finite-dimensional vector \( f \).

When optimizing imaging system performance by use of objective measures of IQ, all sources of randomness in \( g \) should be considered. In diagnostic imaging applications, object variability is an important factor that limits observer performance. In such applications, the object \( f \) should be described as a random vector that is characterized by a multivariate probability density function (PDF) \( p(f) \) that specifies the statistical properties of the ensemble of objects to-be-imaged.

Direct estimation of \( p(f) \) is rarely tractable in medical imaging applications due to the high dimensionality of \( f \). To circumvent this difficulty, a parameterized generative model, referred to throughout this work as a SOM, can be introduced and established by use of an ensemble of experimental measurements. The generative model can be explicit or implicit. Explicit generative models seek to approximate \( p(f) \), or equivalently, its characteristic function, from which samples \( f \) can subsequently be drawn. On the other hand, implicit generative models do not seek to estimate \( p(f) \) directly, but rather define a stochastic process that can draw samples from \( p(f) \) without having to explicitly specify it. Variational autoencoders and GANs are examples of explicit and implicit generative models, respectively, that have been actively explored [38]. Two previous works that sought to learn SOMs from noisy and indirect imaging measurements by use of explicit and implicit generative models are presented below.

A. Establishing SOMs by use of explicit generative modeling: Propagation of characteristic functionals

The first method to learn SOMs from imaging measurements was introduced by Kupinski et al. [10]. In that seminal work, a C-D imaging model was considered in which a function that describes the object is mapped to a finite-dimensional image vector \( g \). For C-D operators, it has been demonstrated that the characteristic functional (CF) describing the object can be readily related to the characteristic function (CF) of the measured data vector \( g \) [39]. This provides a relationship between the PDFs of the object and measured image data. In their method, an object that was parameterized by the vector \( \Theta \)
was considered and analytic expressions for the CFl were utilized. Subsequently, by use of the known imaging operator and noise model, the corresponding CF was computed. The vector $\Theta$ was estimated by minimizing the discrepancy between this model-based CF and an empirical estimate of the CF computed from an ensemble of noisy imaging measurements. From the estimated CFl, an ensemble of objects could be generated. This method was applied to establish SOMs where the CFl of the object can be analytically determined. Such cases include the lumpy object model [29] and clustered lumpy object model [11]. The applicability of the method to more complicated object models remains unexplored.

B. Establishing finite-dimensional SOMs by use of implicit generative modeling: GANs and AmbientGANs

Generative adversarial networks (GANs) [30], [40]–[49] are implicit generative models that have been actively explored to learn the statistical properties of ensembles of images (i.e., finite-dimensional approximations of object properties) and generate new images that are consistent with them. A traditional GAN consists of two deep neural networks - a generator and a discriminator. The generator is jointly trained with the discriminator through an adversarial process. During its training process, the generator is trained to map random low-dimensional latent vectors to higher dimensional images that represent samples from the distribution of training images. The discriminator is trained to distinguish the generated, or synthesized, images from the actual training images. These are often referred to as the “fake” and “real” images in the GAN literature. Subsequent to training, the discriminator is discarded and the generator and associated latent vector probability distribution form as an implicit generative model that can sample from the data distribution to produce new images. However, images produced by imaging systems are contaminated by measurement noise and potentially an image reconstruction process. Therefore, GANs trained directly on images do not generally represent SOMs because they do not characterize object variability alone.

An augmented GAN architecture named AmbientGAN has been proposed [31] that enables learning a SOM that describes the statistical properties of finite-dimensional approximations of objects from noisy indirect measurements of the objects. As shown in Fig. 1, the AmbientGAN architecture incorporates the measurement operator $M_n$, defined in Eq. (1), into the traditional GAN framework. During the AmbientGAN training process, the generator is trained to map a random vector $z \in \mathbb{R}^k$ described by a latent probability distribution to a generated object $\hat{f} = G(z; \Theta_G)$, where $G : \mathbb{R}^k \rightarrow \mathbb{R}^N$ represents the generator network that is parameterized by a vector of trainable parameters $\Theta_G$. Subsequently, the corresponding simulated imaging measurements are computed as $\hat{g} = M_n(\hat{f})$. The discriminator neural network $D : \mathbb{R}^M \rightarrow \mathbb{R}$, which is parameterized by the vector $\Theta_D$, is trained to distinguish the real and simulated imaging measurements by mapping them to a real-valued scalar $s$. The adversarial training process can be represented by the following two-player minimax game [30]:

$$\min_{\Theta_G} \max_{\Theta_D} V(D, G) = E_{\hat{g} \sim p(\hat{g})} [l(D(\hat{g}; \Theta_D))] + E_{g \sim p(g)} [l(1 - D(g; \Theta_D))], \quad (2)$$

where $l(\cdot)$ represents a loss function. When the distribution of objects $p(f)$ uniquely induces the distribution of imaging measurements $p(g)$, i.e., when the imaging operator is injective, and the minimax game achieves the global optimum, the trained generator can be employed to produce object samples drawn from $p(f)$ [30], [31].

Zhou et al. have demonstrated the ability of the AmbientGAN to learn a simple SOM corresponding to a lumpy object model that could be employed to produce small ($64 \times 64$) object samples [32]. However, adversarial training is known to be unstable and the use of AmbientGANs to establish realistic and large-scale SOMs has, to-date, been limited.

C. Advanced GAN training strategies

A novel training strategy for GANs—progressive growing of GANs (ProGANs)—has been recently developed to improve the stability of the GAN training process [33] and hence the ability to learn generators that sample from distributions of high-resolution images. GANs are conventionally trained directly on full size images through the entire training process. In contrast, ProGANs adopt a multi-resolution approach to training. Initially, a generator and discriminator are trained by use of down-sampled (low resolution) training images. During each subsequent training stage, higher resolution versions of the original training images are employed to train progressively deeper discriminators and generators, continuing until a final version of the generator is trained by use of the original high-resolution images. A similar progressive training procedure is employed in the StyleGAN framework [34]. More recently, an advanced GAN training strategy—StyleGAN2—has been developed to further improve the image quality of the synthesized images [35]. Although, StyleGAN2 does not employ the progressive growing strategy, the generator does make use of multiple scales of image generation via skip connections between lower resolution generated images to the final generated image [35]. While these advanced
training strategies have found widespread success on training GANs, they cannot be directly used to train AmbientGANs for establishing SOMs from medical imaging measurements. This is because these GAN training procedures and architectures are designed to train the generator that produces images in the same Hilbert space as the training images. However, medical imaging measurements \( g \) that are used as training data of AmbientGANs are typically indirect representation of to-be-imaged objects \( f \) and generally reside in a different Hilbert space than the generator-produced objects \( \hat{f} \). For example, in magnetic resonance imaging (MRI), the to-be-imaged objects reside in a real Hilbert space while the k-space measurements reside in a complex Hilbert space. A solution to this problem that enables the use of advanced GAN training methods for training AmbientGANs is described next.

III. ESTABLISHING SOMs BY USE OF ADVANCED AMBIENTGANs

In order to train the AmbientGAN by use of advanced GAN training methods that employ the progressive growing approach, such as ProGAN and StyleGANs, an image reconstruction operator \( O: \mathbb{R}^M \rightarrow \mathbb{R}^N \) is included in the AmbientGAN architecture. The discriminator is trained to distinguish between the real reconstructed images \( f = O(g) \) and the fake reconstructed images \( \hat{f} = O(\hat{g}) \). In this way, the generator and the discriminator are associated with images in the same Hilbert space, which enables the use of advanced GAN training methods to train AmbientGANs. This advanced AmbientGAN training strategy is illustrated in Fig. 2.

![Fig. 2: An illustration of the proposed modified AmbientGAN architecture. Any advanced GAN architecture employing a progressive growing training procedure can be employed in this framework.](image_url)

Given a training dataset that comprises measured data \( g \), a set of reconstructed images \( f \) is computed by applying the reconstruction operator \( O \) to the measured data \( g \). The generator is trained with the discriminator through an adversarial process to generate objects \( f = G(z; \Theta_G) \) that result in (fake) reconstructed images \( \hat{f} \) that are indistinguishable, in distribution, from the real reconstructed images \( f \). This adversarial training process can be represented by a two-player minimax game:

\[
\min_{\Theta_G} \max_{\Theta_D} V(D, G) = \mathbb{E}_{\hat{f}_r \sim p(\hat{f}_r)} \left[ \mathbb{E}_{f \sim p(f)} \right] + \mathbb{E}_{f \sim p(f)} \left[ \mathbb{E}_{\hat{f}_r \sim p(\hat{f}_r)} \right],
\]

where \( \hat{f}_r = O(H_n(G(z; \Theta_G))) \). As with the original AmbientGAN, when the distribution of objects \( p(\hat{f}) \) uniquely induces the distribution of reconstructed objects \( p(f) \), and the generator and the discriminator achieve the global optimum, the trained generator can be employed to produce object samples drawn from the distribution \( p(f) \).

It should be noted that when the generator and the discriminator are established progressively by use of ProGAN or StyleGAN methods, the generator is initially trained to produce low resolution images that are subsequently upsampled to the original image dimension. The measurement operator \( H_n \) is subsequently applied to the upsampled images to simulate the measurement data and the reconstructed images are produced by use of the reconstruction operator \( O \). The reconstructed images are down-sampled and the discriminator is subsequently trained on the down-sampled (low resolution) reconstructed images. The generator and the discriminator are progressively trained until the original high-resolution images are achieved. Additional details about the ProAmGAN training procedure can be found in the Supplemental File.

While the progressive growing strategy has achieved many successes in stabilizing the GAN training for synthesizing high-resolution images, it can cause certain artifacts in the generated images [35]. As mentioned above, the StyleGAN2 that trains a redesigned generator without progressive growing was developed to further improve the synthesized image quality [35]. The new generator employs skip connections to form images that are summation of images with different resolutions. This enables the multi-resolution training of the generator without the explicit use of progressive growing strategy. The training of AmbientGANs can be potentially improved further by employing the StyleGAN2 generator and discriminator in the proposed AmbientGAN training framework that is illustrated in Fig. 2. Additional details about the implementation can be found in the Supplemental File. Below, the advanced AmbientGAN that employs the ProGAN was referred to as ProAmGAN and the one that employs the StyleGAN2 was referred to as Sty2AmGAN.

IV. NUMERICAL AND EXPERIMENTAL STUDIES

Computer-simulation and experimental studies were conducted to demonstrate the ability of the proposed advanced AmbientGAN methods to establish SOMs from imaging measurements. Details regarding the design of these studies are provided below.

A. Stylized imager that acquires fully-sampled data

A stylized imaging system that acquires fully-sampled 2D Fourier space (a.k.a., k-space) data was investigated first. This imaging system can be described as:

\[
g = F(f) + n, \tag{4}
\]

where \( F \) denotes a 2D discrete Fourier transform (DFT), \( f \) denotes the discretized object to-be-imaged, and \( n \) denotes the measurement noise. While Eq. (4) can be interpreted as a simplified model of MRI, it should be noted that here we do not attempt to model the real-world complexities of data-acquisition in MRI. A situation where modeling error is present is addressed later in Sec. IV-C. A collection of clinical
brain MR images from the Alzheimer’s Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu) [50] were employed to serve as ground truth objects \( f \). Fifteen thousand sagittal brain slices of dimension 256 \( \times \) 256 were selected from this dataset and were normalized to the range between 0 and 1. These images were employed to form the collection of ground-truth objects \( f \). Examples of \( f \) are shown in Fig. 3.

Fig. 3: Examples of ground-truth objects \( f \).

From the ensemble of objects \( f \), k-space measurement data were simulated according to Eq. (4). The measurement noise \( n \) was modeled by i.i.d. zero mean complex Gaussian distribution with a standard deviation of \( \sigma_n(g) \) for both the real and imaginary components. Different measurement noise levels corresponding to standard deviations \( \sigma_n(g) = 4 \) and 16 were considered.

From each ensemble of simulated k-space data, reconstructed images \( \hat{f} \), were produced by acting a 2D inverse discrete Fourier transform (IDFT) operator \( \mathcal{F}^{-1} \) to the measured image data \( g \) and taking the real component: \( \hat{f}_r = \text{Re}(\mathcal{F}^{-1}(g)) \). For each noise level, a ProAmGAN was trained to establish a SOM that characterizes the distribution of objects \( f \) by use of the ensemble of reconstructed noisy images \( f_r \).

For comparison, standard (i.e., non-ambient) ProGANs were trained directly by use of reconstructed images \( f_r \). In this case, because the reconstructed images are affected by measurement noise, the resulting generator will learn to sample from the distribution of noisy reconstructed images instead of the distribution of (noiseless) objects \( f \).

The Fréchet Inception Distance (FID) [51] score, a widely employed metric for assessing generative models, was computed to evaluate the performance of the original ProGAN and the proposed ProAmGAN. The FID score quantifies the distance between the features extracted by the Inception-v3 network [52] from the ground-truth (“real”) and generated (“fake”) objects. Lower FID score indicates better quality and diversity of the generated objects. The FID scores were computed by use of 15,000 ground-truth objects \( f \) and 15,000 ProAmGAN-generated objects \( \hat{f} \) for each data-acquisition design. To assess the ability of ProAmGANs to accurately learn the variation in the measurement components of the objects, the FID score was computed by use of the ground-truth measurement components \( f_{\text{meas}}^* = H^*Hf \) and ProAmGAN-generated measurement components \( \hat{f}_{\text{meas}} = H^*\hat{H}f \) for each data-acquisition design.

B. Stylized imager that acquires incomplete data

Imaging systems sometimes acquire under-sampled, or incomplete, measurement data to accelerate the data-acquisition process or for other purposes. In such cases, the imaging operator \( H \) has a non-trivial null space and only the measurement component \( f_{\text{meas}} = H^*Hf \) can be observed by the imaging system, where \( H^* \) denotes the Moore-Penrose pseudo-inverse of \( H \). Because of this, it is expected that the performance of an AmbientGAN trained by use of incomplete measurements will be adversely affected by this information loss. This topic is investigated below and the extent to which ProAmGANs can learn to sample from the distribution of measurement components of an object is demonstrated.

The ensemble of 15,000 clinical MR images that was described in Sec. IV-A was employed to serve as ground truth objects. Three accelerated data-acquisition designs that under-sample k-space by use of the Cartesian sampling pattern with an acceleration factor (also known as the reduction factor) \( R \) of 1.25, 2 and 4 were considered. The acceleration factor \( R \) is defined as the ratio of the amount of fully-sampled k-space data to the amount of k-space data collected in the accelerated data-acquisition process. For each considered design, a collection of 15,000 measured data \( g \) were simulated by computing and sampling the k-space data and adding i.i.d. zero mean Gaussian noise with a standard deviation of 4 to both the real and imaginary components.

A stylized imager was considered in which under-sampled k-space data were acquired and \( H^* \) could therefore be computed by applying a 2D IDFT to the zero-filled k-space data. For each data-acquisition design, reconstructed objects \( \hat{f}_r \), were produced by acting \( H^* \) on the given measured image data \( g \). A ProAmGAN was subsequently trained to establish a SOM for each data-acquisition design. In the training process, \( H \) and \( H^* \) were applied to the generator-produced objects as discussed in Sec. III. The FID score was computed by use of 15,000 ground-truth objects \( f \) and 15,000 ProAmGAN-generated objects \( \hat{f} \) for each data-acquisition design. To assess the ability of ProAmGANs to accurately learn the variation in the measurement components of the objects, the FID score was computed by use of the ground-truth measurement components \( f_{\text{meas}} = H^*Hf \) and ProAmGAN-generated measurement components \( \hat{f}_{\text{meas}} = H^*\hat{H}f \) for each data-acquisition design.

C. Experimental emulated single-coil MRI data

As a step towards transcending the stylized studies, an emulated set of single-coil knee MRI k-space measurements were also employed to train a ProAmGAN and Sty2AmGAN. These measurements were obtained from the NYU fastMRI Initiative database [54]. The central 256 \( \times \) 256 regions of the k-space were extracted and a total of 11,400 k-space acquisitions were employed for model training. The reconstructed images were formed as the magnitude of the IDFT of the k-space data. The magnitude MR images are commonly employed in MRI because they can avoid the phase artifacts that are commonly presented in experimental MR measurement data [55].

When training the ProAmGAN and Sty2AmGAN, the canonical measurement model was assumed: \( \hat{g} = H_n(f) = \)
\( F(f) + n \). However, when dealing with experimental measurements, the noise model that characterizes \( n \) is unknown and needs to be estimated. This was accomplished as follows. The noise in the (emulated) experimental k-space measurements was assumed to be described by i.i.d. complex-valued Gaussian random variables; accordingly, the noise in the reconstructed magnitude MR image was modeled by a Rayleigh distribution [55]. The standard deviation of the measurement noise was subsequently estimated by fitting a Rayleigh distribution to a set of patches, residing outside the support of the object, in the magnitude images that were reconstructed from the noisy k-space measurements. The estimated standard deviation specified the k-space noise model in the measurement model above.

In order to train the ProAmGAN and Sty2AmGAN by employing the magnitude MR images as the input to the discriminator, care must be taken when computing the simulated reconstructed image \( \hat{f}_r \). Specifically, if the modulus operator, which is denoted as \( \text{abs}(\cdot) \), is directly applied to the IDFT of the simulated k-space measurements \( \hat{g} \), i.e., \( \hat{f}_r = \text{abs}(F^{-1}(\hat{g})) \equiv \text{abs}(\hat{f} + F^{-1}(n)) \), the fake magnitude images \( \hat{f}_r \) can be indistinguishable from the real magnitude images \( \hat{f} \), despite the fact that the corresponding fake objects \( \hat{f} \) can be negative. This can prevent the generator from being properly trained for use as a SOM.

To address this issue, we computed the fake reconstructed image \( \hat{f}_r \) as:

\[
\hat{f}_r = \hat{f} + e,
\]

where \( e = \text{abs}(F^{-1}(F(\text{ReLU}(\hat{f})) + n)) - \text{ReLU}(\hat{f}) \). Here, \( \text{ReLU}(\cdot) \) is the component-wise Rectified Linear Unit (ReLU) operator that outputs the input value if the input value is positive; while if the input value is negative, it outputs 0. The quantity \( \hat{f}_r \) can be subsequently expressed as:

\[
\hat{f}_r = \begin{cases} 
\text{abs}(F^{-1}(F(\hat{f}) + n)), & \text{if } \hat{f} \geq 0 \\
\hat{f} + \text{abs}(F^{-1}(n)), & \text{if } \hat{f} < 0
\end{cases}
\]

In this way, fake reconstructed images \( \hat{f}_r \) that are produced by positive objects can represent the corresponding magnitude images while those that are produced by negative objects cannot represent magnitude images. Therefore, when the training is completed such that the fake reconstructed images \( \hat{f}_r \) are indistinguishable from the ground-truth magnitude MR images \( \hat{f} \), the generator would produce non-negative objects.

D. Task-based image quality assessment

The generative models established by use of the ProGANs and ProAmGANs in the stylized numerical studies described in Sec. IV-A and Sec. IV-B were further evaluated by use of objective measures of IQ. To accomplish this, a signal-known-exactly and background-known-statistically (SKE/BKS) binary classification task was considered. A Hotelling observer was employed to classify noisy images \( g \) as satisfying either a signal-absent hypothesis \( H_0 \) or signal-present hypothesis \( H_1 \):

\[
H_0: g = f + n, \quad (7a)
\]
\[
H_1: g = f + s + n, \quad (7b)
\]

where \( s \) denotes a signal image and \( n \) is i.i.d. zero-mean Gaussian noise having the standard deviation of 2%. Two studies were conducted in which the background objects \( f \) corresponded to ground truth brain images or synthetic images produced by use of an AmbientGAN. As such, this study sought to determine whether the GAN-generated objects could ‘fool’ the Hotelling observer on the specified detection task. An example of the “real” object \( f \), the corresponding noisy signal-absent image \( g \), and the considered signal are shown in Fig. 4.

The considered signal detection task was performed on a region of interest (ROI) of dimension of 64×64 pixels centered at the signal location. The signal-to-noise ratio of the Hotelling observer (HO) test statistic \( \text{SNR}_{HO} \) was employed as the figure-of-merit for assessing the image quality [7]:

\[
\text{SNR}_{HO} = \sqrt{s_{ROI}^TK^{-1}s_{ROI}},
\]

where \( s_{ROI} \in \mathbb{R}^{4096 \times 1} \) denotes the vectorized signal image in the ROI, and \( K \in \mathbb{R}^{4096 \times 4096} \) denotes the covariance matrix corresponding to the ROIs in the noisy MR images. When computing \( \text{SNR}_{HO} \), \( K^{-1} \) was calculated by use of a covariance matrix decomposition [7]. The values of \( \text{SNR}_{HO} \) were computed by use of 15,000 generated objects produced by each trained generative model. They were compared to the \( \text{SNR}_{HO} \) computed by use of 15,000 ground truth objects.

E. Training details

All ProGAN, ProAmGAN, and Sty2AmGAN models were trained by use of Tensorflow [56] on 2 NVIDIA Quadro RTX 8000 GPUs. The Adam algorithm [57], which is a stochastic gradient algorithm, was employed as the optimiser in the training process.

The ProAmGANs were implemented by modifying the ProGAN code (https://github.com/tkarras/progans) according to Fig. 2. Specifically, for each considered imaging system, the corresponding measurement operator was applied to the generator-produced images for simulating the measurement data and the reconstruction operator was applied to the measurement data for producing the reconstructed images used as the input to the discriminator. The default ProGAN architecture with the latent space having the dimensionality of 512 and the initial image resolution of 4 × 4 was employed to implement the ProAmGANs for the considered numerical studies. Additional details about the
ProGAN architecture and the progressive growing training method can be found in the literature [33]. The Sty2AmGAN was implemented by modifying the StyleGAN2 code (https://github.com/NVlabs/stylegan2) by augmenting the StyleGAN2 with the measurement operator $\mathcal{H}_n$ and the reconstruction operator $\mathcal{O}$ according to Fig. 2. For the considered experimental study, the default StyleGAN2 architecture (i.e., “config F” [35]) with the input latent space having the dimensionality of 512 was employed to implement the Sty2AmGAN. Additional details regarding the StyleGAN2 architecture and the corresponding training strategy can be found in the literature [35].

V. Results

A. Stylized imager that acquires fully-sampled data

Images that were synthesized by use of the ProAmGANs and ProGANs that were trained by use of fully-sampled noisy k-space data or images reconstructed from them, respectively, are shown in Figs. 5 (a) and (b). Subfigures (a) and (b) correspond to measurement noise levels of 4 and 16, respectively.

![Fig. 5: (a) ProGAN-generated (top row) and ProAmGAN-generated (bottom row) images corresponding to $\sigma_n(g) = 4$. (b) The corresponding images for $\sigma_n(g) = 16$.](image)

It is observed that the ProGAN-generated images contain significant noise when $\sigma_n(g) = 16$, while the ProAmGAN generated clean images that do not contain significant noise. This demonstrates the ability of the ProAmGAN to mitigate measurement noise when establishing SOMs.

The FID scores, estimated standard deviation of the noise in the generated images $\sigma_n(\hat{f})$, and SNR$_{HO}$ were evaluated for both the ProGANs and ProAmGANs. These metrics are shown in Table. I. The ProAmGANs produced FID scores that were smaller than those produced by the ProGANs, which indicates that the ProAmGANs outperformed the ProGANs. The estimated standard deviation of the noise in the ProGAN-generated images increased nearly linearly as the standard deviation of measurement noise was increased; while the estimated standard deviation of the noise in the ProAmGAN-generated images were almost unchanged. The SNR$_{HO}$ values corresponding to the ProGANs had negative biases to the reference value that were computed by use of ground-truth objects, and this negative bias became more significant as the measurement noise level increased. This is because the ProGANs capture both the object variability and the noise randomness, instead of object variability alone, which degrades the estimated observer performance. The SNR$_{HO}$ values corresponding to the ProAmGANs were closer to the reference value.

![Table I: The FID score of the objects, the estimated noise standard deviation, and the SNR$_{HO}$](table)

B. Stylized imager that acquires incomplete data

Images that were synthesized by use of the ProAmGANs that were trained by use of under-sampled k-space measurement data acquired with different acceleration factors are shown in Fig. 6. The images produced by the ProAmGANs corresponding to the acceleration factor $R = 1.25$ and 2 are visually plausible; while when the acceleration factor was increased to 4, the generated images were obviously contaminated by artifacts and some structures were distorted. This demonstrates that the ProAmGAN was adversely affected by the incompleteness of the measurement data acquired by imaging systems having non-trivial null-space.

The quantitative metrics that include FID scores and SNR$_{HO}$ are summarized in Table. II. The FID scores corresponding to the generated objects $\hat{f}$ were increased when the acceleration factor $R$ increased, which indicates that the ProAmGAN was detrimentally affected by the null space of the imaging operator. However, the FID scores corresponding to the measurement components of the generated objects $\hat{f}_{meas}$ were not significantly affected, which suggests that...
the ProAmGAN can reliably learn the distribution of the measurement components of the objects. The SNR$_{HO}$ values produced by the ProAmGANs had positive biases to the reference value that was computed by use of the ground-truth objects. This indicates that the ProAmGAN was unable to learn complete object variation from incomplete imaging measurements.

|       | $R = 1.25$ | $R = 2$ | $R = 4$ |
|-------|------------|---------|---------|
| FID ($\hat{f}$) | 20.64     | 39.25   | 118.27  |
| FID ($\hat{f}_\text{meas}$) | 12.83     | 13.25   | 8.96    |
| SNR$_{HO}$ | 1.75      | 1.80    | 1.84    |

TABLE II: The FID score of the objects, the FID score of the measurement components, and the SNR$_{HO}$ (reference value 1.72) corresponding to the objects produced by the ProAmGANs that were trained with under-sampled k-space data with different acceleration factors.

C. Experimental emulated single-coil MRI data

Images produced by the ProGAN, ProAmGAN, and Sty2AmGAN are shown in the top row, middle row, and bottom row of Fig. 7, respectively. The ProGAN-produced images were contaminated by noise because the ProGAN was trained directly by use of noisy reconstructed images. Both the ProAmGAN and Sty2AmGAN produced images that did not appear to be degraded by noise, which demonstrates the ability of advanced AmbientGAN strategies to mitigate the measurement noise when establishing an SOM. The Sty2AmGAN can further improve the training of the AmbientGAN for establishing the SOM. For example, the images produced by the ProAmGAN were more blurred than those produced by the Sty2AmGAN. Because the ground-truth objects corresponding to the synthesized images were not accessible in this experimental study, only a subjective visual assessment was performed.

![Fig. 6: Top row: ProAmGAN-generated images corresponding to the k-space sampling acceleration factor $R$ of 1.25. Middle row: The corresponding images for $R = 2$. Bottom row: The corresponding images for $R = 4$.](image1)

![Fig. 7: Results from emulated single-coil MRI data. Top row: ProGAN-generated images. Middle row: ProAmGAN-generated images. Bottom row: Sty2AmGAN-generated images.](image2)

VI. DISCUSSION AND CONCLUSION

It is known that it is important to address object variability when computing objective measures of image quality for use in imaging system characterization or optimization. When computer-simulation studies are employed, SOMs are the means by which this can be accomplished. However, establishing realistic SOMs from experimental image data has remained challenging and few methods are available to accomplish this.

Motivated by this need, a methodology for training AmbientGANs by use of medical image data was proposed in this study. The trained generator of the AmbientGAN represents the sought-after SOM. The proposed methodology enables the use of advanced GAN training methods and architectures in the AmbientGAN training and therefore permits the AmbientGAN to be applied to realistically sized medical image data. To demonstrate this, two specific advanced AmbientGANs were considered: ProAmGANs and Sty2AmGANs.
Stylized numerical studies were systematically conducted in which ProAmGANs were trained on simulated measurement data corresponding to an object ensemble and a variety of stylized imaging systems. Both visual examinations and quantitative analyses, including task-specific validations, demonstrated that the proposed ProAmGANs hold promise to establish realistic SOMs from imaging measurements. Additionally, an experimental study was conducted in which the ProAmGAN and Sty2AmGAN were both trained on emulated experimental MRI measurement data. This demonstrated the effectiveness of the methods under less stylized conditions in which modeling error was present.

In addition to objectively assessing imaging systems and data-acquisition designs, the SOMs established by the proposed advanced AmbientGAN methods can be employed to regularize image reconstruction problems. Recent methods have been developed for regularizing image reconstruction problems based on GANs such as Compressed Sensing using Generative Models (CSGM) [60] and image-adaptive GAN-based reconstruction methods (IAGAN) [61], [62]. Sty2AmGANS can also be used for prior image-constrained reconstruction [59]. The established SOMs can also be used to produce clean reference images for training deep neural networks for solving other image-processing problems such as image denoising [63], [64] and image super-resolution [65].

There remain additional topics for future investigation. We have conducted a preliminary objective assessment of the AmbientGANs by use of the Hotelling Observer [7], [66] and a binary signal detection task. It will be important to validate the SOMs established by use of the proposed methods when clinically relevant tasks are addressed by a variety of numerical observers such as the ideal observer [67]–[70] and anthropomorphic observers [71].

VII. ACKNOWLEDGEMENTS

Data collection and sharing for this project was funded by the Alzheimer’s Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012).

REFERENCES

[1] E. Abadi, W. P. Segars, B. M. Tsui, P. E. Kinahan, N. Bottenus, A. F. Frangi, A. Maidment, J. Lo, and E. Samei, “Virtual clinical trials in medical imaging: a review,” Journal of Medical Imaging, vol. 7, no. 4, p. 042805, 2020.
[2] A. Badano, A. Badal, S. Glick, C. G. Graff, F. Samuelson, D. Sharma, and R. Zeng, “In silico imaging clinical trials for regulatory evaluation: initial considerations for viciere, a demonstration study,” in Medical Imaging 2017: Physics of Medical Imaging, vol. 10132. International Society for Optics and Photonics, 2017, p. 1013220.
[3] D. Wilson, H. H. Barrett, and E. W. Clarkson, “Reconstruction of two- and three-dimensional images from synthetic-collimator data,” IEEE transactions on medical imaging, vol. 19, no. 5, pp. 412–422, 2000.
[4] K. J. Myers, R. F. Wagner, and K. M. Hanson, “Rayleigh task performance in tomographic reconstructions: Comparison of human and machine performance,” in Medical Imaging 1993: Image Processing, vol. 1898. International Society for Optics and Photonics, 1993, pp. 628–637.
[5] R. F. Wagner and D. G. Brown, “Unified SNR analysis of medical imaging systems,” Physics in Medicine & Biology, vol. 30, no. 6, p. 489, 1985.
[6] H. H. Barrett, J. Yao, J. P. Rolland, and K. J. Myers, “Model observers for assessment of image quality,” Proceedings of the National Academy of Sciences, vol. 90, no. 21, pp. 9758–9765, 1993.
[7] H. H. Barrett and K. J. Myers, Foundations of Image Science. John Wiley & Sons, 2013.
[8] M. A. Anastasios, C.-Y. Chou, A. M. Zysk, and J. G. Brankov, “Analysis of ideal observer signal detectability in phase-contrast imaging employing linear shift-invariant optical systems,” JOSA A, vol. 27, no. 12, pp. 2648–2659, 2010.
[9] J. P. Rolland and H. H. Barrett, “Effect of random background inhomogeneity on observer detection performance,” J. Opt. Soc. Am. A, vol. 9, no. 5, pp. 649–658, May 1992. [Online]. Available: http://josaa.osa.org/abstract.cfm?URI=josaa-9-5-649.
[10] M. A. Kupinski, E. Clarkson, J. W. Hoppin, L. Chen, and H. H. Barrett, “Experimental determination of object statistics from noisy images,” JOSA A, vol. 20, no. 3, pp. 421–429, 2003.
[11] F. O. Bochud, C. K. Abbey, and M. P. Eckstein, “Statistical texture synthesis of mammographic images with clustered lumpy backgrounds,” Optics Express, vol. 4, no. 1, p. 33–43, 1999.
[12] W. P. Segars, M. Mahesh, T. J. Beck, E. C. Frey, and B. M. Tsui, “Realistic ct simulation using the 4d xcat phantom,” Medical Physics, vol. 35, no. 8, pp. 3800–3808, 2008.
[13] W. P. Segars and B. M. Tsui, “Simulation of the efficacy of respiratory gating in myocardial spec using the new 4-d ct phantom,” IEEE Transactions on Nuclear Science, vol. 49, no. 3, pp. 675–679, 2002.
[14] X. G. Xu, “An exponential growth of computational phantom research in radiation protection, imaging, and radiotherapy: a review of the fifty-year history,” Physics in medicine and biology, vol. 59, no. 18, p. R233, 2014.
[15] M. Zankl and K. Eckerman, “The gsf voxel computational phantom family,” Handbook of Anatomical Models for Radiation Dosimetry, pp. 65–80, 2015.
[16] D. L. Collins, A. P. Zijdenbos, V. Kollokian, J. G. Sled, N. J. Kabani, C. J. Holmes, and A. C. Evans, “Design and construction of a realistic digital brain phantom,” IEEE Transactions on Medical Imaging, vol. 17, no. 3, pp. 463–468, 1998.
[17] J. Caon, “Voxel-based computational models of real human anatomy: a review,” Radiation and Environmental Biophysics, vol. 42, no. 4, pp. 229–235, 2004.
[18] X. G. Zu, “The VIP-Man Model-A Digital Human Testbed for Radiation Simulations,” SAE Transactions, pp. 779–787, 2005.
[19] C. M. Li, W. P. Segars, G. D. Tourassi, J. M. Boone, and J. T. Dobbins III, “Methodology for generating a 3d computerized breast phantom from empirical data,” Medical Physics, vol. 36, no. 7, pp. 3122–3131, 2009.
[20] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, “Active shape models-their training and application,” Computer vision and image understanding, vol. 61, no. 1, pp. 38–59, 1995.
[21] T. Cootes, M. Roberts, K. Babalola, and C. Taylor, “Active Shape and Appearance Models,” in Handbook of Biomedical Imaging. Springer, 2015, pp. 105–122.
[22] T. Heimann and H.-P. Meinzer, “Statistical shape models for 3d medical image segmentation: a review,” Medical Image Analysis, vol. 13, no. 4, pp. 543–563, 2009.
[23] V. Ferrari, F. Jurie, and C. Schmid, “From images to shape models for object detection,” International Journal of Computer Vision, vol. 87, no. 3, pp. 284–303, 2010.
[24] K.-k. Shen, J. Fripp, F. Mériaudeau, G. Chételat, O. Salvado, P. Bourget, A. D. N. Initiative et al., “Detecting global and local hippocampal shape changes in alzheimer’s disease using statistical shape models,” Neuroimage, vol. 59, no. 3, pp. 2155–2166, 2012.
[25] N. Gordillo, E. Montesino, and P. Sobrerro, “State of the art survey on MRI brain tumor segmentation,” Magnetic resonance imaging, vol. 31, no. 8, pp. 1426–1438, 2013.
[26] S. Tomoshige, E. Oost, A. Shimizu, H. Watanabe, and S. Nawano, “A conditional statistical shape model with integrated error estimation of the conditions; application to liver segmentation in non-contrast ct images,” Medical Image Analysis, vol. 18, no. 1, pp. 130–143, 2014.
[27] F. Ambellan, A. Tack, M. Ehlke, and S. Zachow, “Automated segmentation of knee bone and cartilage combining statistical shape knowledge and convolutional neural networks: Data from the osteoarthritis initiative,” Medical Image Analysis, vol. 52, pp. 109–118, 2019.
[28] E. Clarkson and H. H. Barrett, “Characteristic functions in imaging and image-quality assessment: tutorial,” JOSA A, vol. 33, no. 8, pp. 1464–1475, 2016.
[29] M. A. Kupinski and H. H. Barrett, Small-animal SPECT imaging. Springer, 2005, vol. 233.
Learning stochastic object models from medical imaging measurements by use of advanced AmbientGANs (Supplementary Material)

Weimin Zhou, Sayantan Bhadra, Frank J. Brooks, Hua Li, and Mark A. Anastasio

I. PROAmGAN TRAINING PROCEDURE

The training procedure of Progressive Growing AmbientGAN (ProAmGAN) is illustrated in Fig. S. 1. The generator and discriminator are initially trained on low-resolution images. Deeper generator and discriminator are progressively trained by use of higher-resolution versions of the original images until the final generator that produces images having the original-resolution is trained.

Fig. S. 1: ProAmGAN training procedure. Initially, the generator and discriminator are trained low-resolution images. More layers in the generator and discriminator are trained by use of higher-resolution versions of the original images when the training advances. More details about the progressive growing method can be found in the original ProGAN paper [T. Karras et al., 2018]

II. STY2AmGAN TRAINING PROCEDURE

The Style2-AmbientGAN (Sty2AmGAN) training procedure is illustrated in Fig. S. 2. The generator employs skip connections to produces images that are summation of images at different resolutions. The discriminator employs residual connections that can be beneficial for performing image classification tasks. The Sty2AmGAN is trained without progressive growing.

Fig. S. 2: Sty2AmGAN training procedure. The generator employs skip connections and forms the images by explicitly summing images at different resolutions. The discriminator employs residual connections that can be helpful for performing image classification tasks. The generator and discriminator are trained without progressive growing. More details about the generator and discriminator architectures can be found in the original StyGAN2 paper [T. Karras et al., 2019]