Adaptive LPD Radar Waveform Design with Generative Deep Learning

Matthew R. Ziemann*† and Christopher A. Metzler*
*University of Maryland, College Park, MD 20742
†DEVCOM Army Research Laboratory, Adelphi, MD 20783
Email: matthew.r.ziemann.civ@army.mil, metzler@umd.edu

Abstract—We propose a novel, learning-based method for adaptively generating low probability of detection (LPD) radar waveforms that blend into their operating environment. Our waveforms are designed to follow a distribution that is indistinguishable from the ambient radio frequency (RF) background—while still being effective at ranging and sensing. To do so, we use an unsupervised, adversarial learning framework; our generator network produces waveforms designed to confuse a critic network, which is optimized to differentiate generated waveforms from the background. To ensure our generated waveforms are still effective for sensing, we introduce and minimize an ambiguity function-based loss on the generated waveforms. We evaluate the performance of our method by comparing the single-pulse detectability of our generated waveforms with traditional LPD waveforms using a separately trained detection neural network. We find that our method can generate LPD waveforms that reduce detectability by up to 90% with improved ambiguity characteristics. The generator can be tuned to further improve ambiguity characteristics at the cost of increased detectability.

Index Terms—radar, deep learning, low probability of detection, waveform design, generative adversarial networks

Fig. 1. **LPD Radar Waveform Generation Framework.** The generative model optimizes range/velocity resolution while minimizing detectability.

I. INTRODUCTION

Radar systems are critical for military operations. They allow for surveillance, reconnaissance, navigation, fire control, air defense, and a host of other tasks on platforms that range from small unmanned systems to massive aircraft carriers. However, radar systems are vulnerable to detection and counter-action by adversary electronic support systems, including radar warning receivers and anti-radiation missiles [1]. Low probability of detection (LPD) radar addresses this vulnerability through the use of specialized waveforms and emission patterns that lower their likelihood of detection [2].

LPD waveforms are designed to minimize detectability while maintaining necessary radar performance metrics such as range and velocity resolution. Traditional methods for LPD radar waveform design include low peak power waveforms, wideband waveforms, frequency agile or hopping waveforms, coded or modulated waveforms, and irregular scan patterns [3]. These methods aim to reduce the peak power, bandwidth, or duration of the radar signal—or to introduce randomness or complexity in the signal structure—so intercept receivers cannot easily detect or identify the radar signal. However, these waveforms may still be vulnerable to advanced intercept systems that can exploit their statistical features, such as cyclostationarity [4], [5]. Modern software defined rada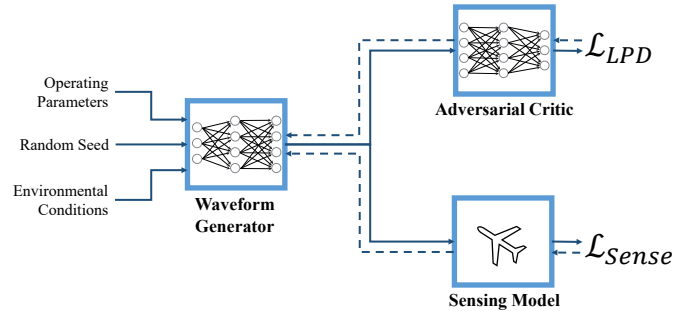r (SDRadar) has opened the door to increased waveform diversity—especially for cognitive radar applications—but the development of algorithms for adaptive SDRadar systems is still in its infancy [6]–[8].

In recent years, deep learning has enabled significant advances in radar signal processing, including advanced modulation classification, waveform design, and synthetic data generation [9]–[13]. We propose a novel application of deep learning for LPD radar waveform design. Generative machine learning techniques, such as generative adversarial networks (GANs), can learn the underlying distribution of a dataset and generate new examples that closely match the original distribution [14], [15]. This makes them well-suited for LPD waveform design tasks, as they can learn the statistical characteristics of the environment and generate radar waveforms that mimic background radio frequency (RF) signals.

In this work, we employ unsupervised GANs to generate unique and adaptive radar waveforms that match the statistical distribution of the background RF environment. In particular, we train a conditional Wasserstein GAN to generate LPD radar waveforms based on the distribution of common RF communication modulations. We utilize an objective based on a differentiable ambiguity function to optimize its outputs for range and Doppler ambiguity requirements. An overview of this framework is shown in Fig. 1. We find that our generated waveforms exhibit up to a 90% reduction in detectability compared to other LPD waveforms while maintaining useful range and velocity resolutions. In addition, we can increase the strength of the ambiguity function-based objective to further improve ambiguity characteristics at the cost of increased detectability.

II. RELATED WORK

Radar waveforms determine how well a radar system can detect, locate, and classify objects; how much interference a radar system causes and how sensitive a system is to interference and jamming; and how easy it is for a third party to detect a radar system. The primary objective of radar waveform design is to optimize these trade-offs so as to ensure a system can sense, coexist with, and survive in its environment.

Waveform Design. Over the years, the field of radar waveform design has evolved significantly, driven by both technological advancements and changing operational requirements. Initial designs were largely constrained by hardware limitations, focusing on frequency-modulated and phase-coded pulses [16]. As technology progressed, more complex designs emerged, such as pulse-Doppler, frequency-modulated continuous-wave (FMCW), and multiple-input, multiple-output (MIMO) systems [17]. These advancements helped to expand the operational scope of radar systems, allowing for more accurate and reliable surveillance in a variety of conditions. However, these waveform designs are primarily based on static or predefined scenarios and have limited adaptability in dynamic environments.

Low Probability of Detection. Waveform design for Low Probability of Detection (LPD) and Low Probability of Intercept (LPI) has been an area of significant interest, particularly for military applications. The primary goal of these designs is to minimize the radar's detectability while maintaining sufficient operational performance. Early LPI/LPD waveforms were characterized by low temporal, spectral, or spatial energy densities, which made them difficult to discern from background noise [18]. Techniques such as FMCW, frequency hopping, and non-traditional modulation methods like polyphase coding were common [3]. However, these methods often came with trade-offs in terms of range resolution and target identification capabilities. Noise radar emerged as an additional LPD approach that utilized random noise-like signals to reduce detectability. These waveforms have desirable ambiguity functions and are resistant to electronic countermeasures (i.e., jamming), but their random nature requires longer dwell times and more expensive processing [19].

Cognitive Radar. The concept of cognitive radar emerged in parallel with advancements in waveform design as a means of rapidly adapting radar to changing conditions [20]. Early cognitive radar, defined by the *sensor management* paradigm, used decision algorithms to adjust radar emission parameters such as power, direction, and beamwidth [21]. This quickly evolved to incorporate adaptive signal processing (e.g., space-time adaptive processing) [22]—and, eventually, waveform selection and intra-pulse adaptation [23], [24]. This growing field of *waveform diversity* [6] aims to optimize the radar waveform itself to maximize performance according to particular scenarios and tasks [25].

Modern waveform diversity incorporates the perception-action cycle (PAC) to optimize emissions based on the radar's environment [26]. In this framework, the cognitive system is provided with measurements from its environment in order to intelligently select emission parameters (e.g., waveform class, pulse repetition interval, and bandwidth [24]). While a host of techniques have been explored to accomplish this in applications varying from target detection to spectrum sharing [27], there are still many open research questions regarding optimization objective, technique selection, algorithm response time, and more [7], [8]. For example, the extension of these techniques to LPD objectives—such as adaptive spectrum shaping of noise radar [28]—and the incorporation of emerging deep learning and reinforcement learning methodologies [29] are two emerging areas of interest.

In this work, we introduce a novel application of unsupervised generative adversarial neural networks (GANs) to adaptively generate optimal LPD radar waveforms from measurements of the RF background. By using a conditional GAN, we simplify the perception-action cycle by generating waveforms directly from measurements. Unlike supervised waveform design methods, we do not train our GAN on a dataset of radar waveforms [30], [31]. Instead, we train our GAN on unlabeled data that is representative of (or taken directly from) the target environment, and then fine-tune it with a user-defined ambiguity function to produce LPD waveforms with desirable ambiguity characteristics. These LPD waveforms match the statistical behavior of the RF background distribution, rather than relying on assumptions about the environment or detector to reduce detectability (e.g., white noise, frequency hopping). This flexible framework reduces the data burden associated with deep learning-based techniques while enabling the model to adapt to changing RF environments or operational requirements.

III. PROBLEM FORMULATION

We define a simple radar detection problem:

$$x(t) = \alpha s(t) + n(t) + b(t), \tag{1}$$

$$\alpha = \sqrt{\frac{G_t G_i \lambda^2 L}{(4\pi)^2 R^2}}, \tag{2}$$

where $x(t)$ is the detector's received signal, $s(t)$ is the emitted radar signal, and α is the energy attenuation coefficient defined by the transmitter and detector gain $G_{t,i}$, wavelength λ, path loss L, and distance between the the emitter and detector R. Noise is denoted with $n(t)$, and $b(t)$ denotes the RF background signal(s) present in the spectrum. We assume the background $b(t)$ follows a distribution $\mathbb{P}_b$ and that we can observe many realizations $b_1(t), b_2(t), ..., b_m(t)$ of background.

A. Waveform Design

Given this problem definition, our goal is to find a signal $s(t)$ that has good radar performance and is difficult to
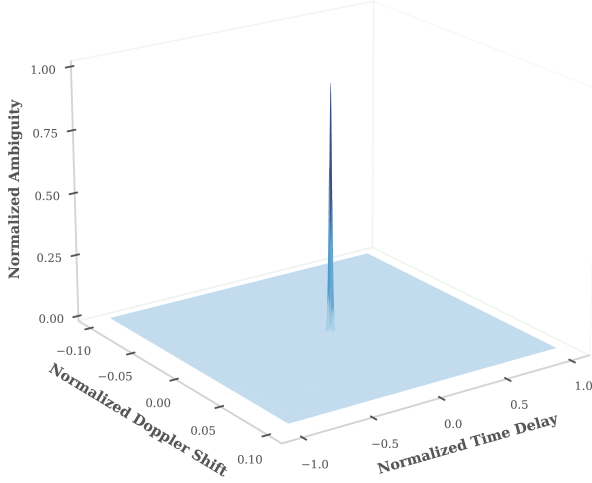
A good LPD waveform is therefore one that is similar to the RF background distribution and has a desirable ambiguity function. We can set up this optimization problem by minimizing the KLD between generated and background signals, and by minimizing the distance between the generated ambiguity function and some desired target ambiguity function. In this work, we accomplish this optimization with deep neural networks.

B. Generative Adversarial Networks

Deep generative models (DGMs) are a type of machine learning model that aim to learn the underlying probability distribution of complex and high-dimensional data [15]. They can then generate new and representative samples from the learned distribution. For example, a DGM trained on an image dataset of human faces can generate realistic faces that do not actually exist [34]. Common DGMs include variational autoencoders (VAEs) [35], generative adversarial networks (GANs) [14], normalizing flows [36], and diffusion models [34]. Although each method has strengths, we utilize GANs in this work as the adversarial training framework is particularly well suited to our LPD waveform design problem.

GANs are composed of two separate neural networks—a generator and a discriminator—trained in a zero-sum game. The generator produces fake examples from a random noise vector, and the discriminator distinguishes between fake examples and real examples from the training data distribution (sometimes called true or "target" examples). The networks are pitted against one another, and, as they train, the generator learns to produce examples that are indistinguishable from the real ones. Together, these networks optimize the following minimax objective:

$$\min_G \max_D \mathop{\mathbb{E}}_{x \sim \mathbb{P}_r} [\log D(x)] + \mathop{\mathbb{E}}_{\tilde{x} \sim \mathbb{P}_g} [\log(1 - D(\tilde{x}))], \quad (6)$$

where G is the generator, D is the discriminator, $\mathbb{P}_r$ is the real data distribution, and $\mathbb{P}_g$ is the generated data distribution defined by generated data $\tilde{x} = G(z)$. Here, z is the input to the generator that is randomly sampled from some distribution (e.g., Gaussian).

As these networks train, they effectively minimize the Jensen–Shannon divergence (JSD) between the real and generated distributions [37]:

$$JS(\mathbb{P}_r \| \mathbb{P}_g) = \frac{1}{2} KL(\mathbb{P}_r \| \mathbb{P}_A) + \frac{1}{2} KL(\mathbb{P}_g \| \mathbb{P}_A), \quad (7)$$

where $\mathbb{P}_A$ is the average distribution $\frac{P_r + P_g}{2}$. JSD is based on KLD, though it is generally easier to optimize because it is symmetric and always finite. However, this objective is not always continuous and leads to inconsistent training. The Wasserstein distance is a more suitable alternative that is continuous everywhere, and so we use the Wasserstein GAN (WGAN) [38]. Rather than minimizing the JSD, WGAN minimizes the Wasserstein-1 distance (also called the Earth-Mover distance):

$$W(\mathbb{P}_r \| \mathbb{P}_g) = \inf_{\gamma \in \Pi(\mathbb{P}_r, \mathbb{P}_g)} \mathbb{E}_{(x,y) \sim \gamma}[\|x - y\|], \quad (8)$$

Fig. 2. **Thumbtack Ambiguity Function.** Example of an optimal "thumbtack" ambiguity function with low delay (range) and Doppler (velocity) ambiguities.

detect. Radar performance is well described by the *ambiguity function*, which computes the matched filter response for a signal under both time and Doppler mismatches [32]. The complex ambiguity function $\hat{A}$ for time shift τ and Doppler shift F_D is

$$\hat{A}(\tau, F_D) = \int_{-\infty}^{\infty} s(t) \exp(j2\pi F_D t) s^*(t - \tau) dt. \quad (3)$$

This can be written in terms of the signal spectrum as

$$\hat{A}(\tau, F_D) = \int_{-\infty}^{\infty} S^*(F) S(F - F_D) \exp(j2\pi F\tau) dF, \quad (4)$$

where $S(F)$ represents the Fourier transform of the time domain signal $s(t)$. The ideal ambiguity function is dependent on the radar task, but desirable characteristics often include a narrow mainlobe and low ambiguity in both delay (range) and Doppler (velocity). This is often described as the "thumbtack" ambiguity function depicted in Fig. 2.

Optimizing a waveform for low detectability is more difficult, as there is not a singular definition of detectability. Peak-to-average power ratio, instantaneous bandwidth, and variable transmit period are just a few of many variables affecting detectability [3]. Chen *et al.* [33] introduce a novel approach for assessing LPD by utilizing the Kullback–Leibler divergence (KLD) between a radar waveform and the RF background. The KLD quantifies the difference between two distributions, and reducing the KLD between the radar waveform and RF background increases their similarity and reduces detectability. For some measurable space χ with probability measures $\text{Prob}(\chi)$, we can define the KLD as

$$KL(\mathbb{P}_b \| \mathbb{P}_g) = \int \log \left(\frac{P_b(x)}{P_g(x)} \right) P_b(x) d\mu(x), \quad (5)$$

where $\mathbb{P}_b$ and $\mathbb{P}_g$ are the distributions of RF background and generated signals such that $\mathbb{P}_b, \mathbb{P}_g \in \text{Prob}(\chi)$, μ is some measure defined on χ, and P_b and P_g are densities with respect to μ.
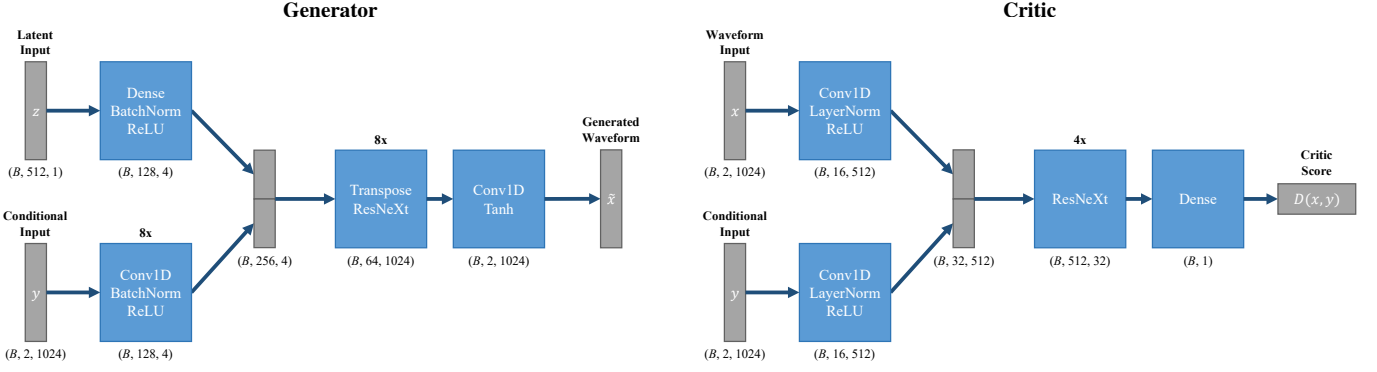
Fig. 3. **Network Architectures.** Overview diagrams of the generator network (left) and critic network (right) with 1024 signal length and 512 latent vector length. Output shapes are shown below layers with batch size B.

where $\Pi(\mathbb{P}_r, \mathbb{P}_g)$ is the set of all joint distributions $\gamma(x, y)$ with marginals $\mathbb{P}_r$ and $\mathbb{P}_g$. Using the Kantorovich–Rubinstein duality, the GAN objective (6) can then be rewritten to:

$$\min_G \max_{D \in \mathcal{D}} \mathbb{E}_{x \sim \mathbb{P}_r} [D(x)] - \mathbb{E}_{\tilde{x} \sim \mathbb{P}_g} [D(\tilde{x})], \quad (9)$$

where $\mathcal{D}$ is the set of 1-Lipschitz functions. In WGANs, the discriminator is instead referred to as a critic. It no longer outputs a probability of an example being "real" or "fake," but provides a continuous valued score used to compute the Wasserstein distance between distributions. This improves training stability and allows the critic to be a more meaningful measure of output quality. The 1-Lipschitz constraint is enforced with a gradient penalty term to constrain the norm of the network's gradients to at most 1 for a random subset of examples $\hat{x} \sim \mathbb{P}_{\hat{x}}$ [39]. This new objective for WGAN with gradient penalty (WGAN-GP) is

$$\min_G \max_D \mathbb{E}_{x \sim \mathbb{P}_r} [D(x)] - \mathbb{E}_{\tilde{x} \sim \mathbb{P}_g} [D(\tilde{x})] + \lambda \mathbb{E}_{\hat{x} \sim \mathbb{P}_{\hat{x}}} [(\|\nabla_{\hat{x}} D(\hat{x})\|_2 - 1)^2]. \quad (10)$$

By optimizing (10), we force the generator's distribution $\mathbb{P}_g$ to match the real distribution $\mathbb{P}_r$. We can use this Wasserstein-1 minimization in place of KLD for LPD waveform generation by setting the real distribution equal to the RF background distribution ($\mathbb{P}_r = \mathbb{P}_b$) for a target environment. This will force our generator to learn to generate waveforms that are statistically similar—have small Wasserstein-1 distance—to the RF background.

In the context of RF background signals, it is crucial to note that the statistical properties of the distribution $\mathbb{P}_b$ are time-dependent. Over long timescales—i.e., minutes to hours—the background exhibits consistent, cyclical behaviors. However, over shorter timescales—i.e., microseconds to milliseconds—the background is inherently nonstationary, with rapidly fluctuating characteristics. To reconcile these disparate timescales and to adapt to the nonstationary nature of short-term RF background signals, we employ a conditional WGAN-GP (cWGAN-GP) model. By doing so, we enable the generator to produce waveforms from the conditional distribution $\mathbb{P}_{g|y}$, where y is an instantaneous RF background measurement taken at or near the time of inference. This conditional

approach allows us to train the WGAN on a long-term, quasi-stationary distribution while generating examples that are instantaneously conditioned on the most recent sample, thereby providing a more accurate Wasserstein-1 approximation in the nonstationary regime. This gives us the cWGAN-GP objective:

$$L_W = \min_G \max_D \mathbb{E}_{x \sim \mathbb{P}_b} [D(x|y)] - \mathbb{E}_{\tilde{x} \sim \mathbb{P}_{g|y}} [D(\tilde{x}|y)] + \lambda \mathbb{E}_{\hat{x} \sim \mathbb{P}_{\hat{x}}} [(\|\nabla_{\hat{x}} D(\hat{x}|y)\|_2 - 1)^2], \quad (11)$$

where $\mathbb{P}_{g|y}$ is the conditional generated data distribution defined by $\tilde{x} = G(z|y)$. In terms of our generator and critic networks, this conditional term is simply an additional input to the networks. This enables the generator to have information of the instantaneous RF background at inference time to improve its output.

IV. METHODOLOGY

We formulate the LPD waveform design problem as jointly maximizing similarity to the RF background and similarity to a desired ambiguity function. We accomplish this with two distinct objectives in our WGAN. First, we use the cWGAN-GP loss function to force the distribution of generated waveforms to be similar to a defined RF background. Second, we use an ambiguity loss term to optimize the ambiguity of generated waveforms toward a target ambiguity function. The resulting generator neural network is then capable of producing a diverse set of LPD radar waveforms, conditioned on the instantaneous RF background, with desirable ambiguity characteristics.

A. Wasserstein GAN

We optimize our LPD and ambiguity objectives with a conditional Wasserstein GAN with gradient penalty (cWGAN-GP) implemented in PyTorch [40]. The generator and critic in our cWGAN-GP are based on a 1D variant of the ResNeXt architecture [41]. ResNeXt is a modular architecture built from *residual blocks* that utilize parallel convolution pathways and a residual ("skip") connection. The *cardinality* of the residual layer defines the number of parallel pathways. Importantly, each pathway has a lower dimensionality than a standard ResNet layer. This enables a greater degree of expressivity
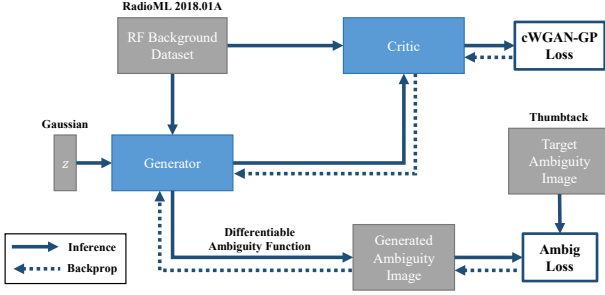
Fig. 4. **Adversarial Training Framework.** Training process for our cWGAN-GP, where our generator learns to produce waveforms with desirable ambiguity functions that mimic the RF background. Training is initially performed without the ambiguity loss, and alternates between training the critic and the generator. Once converged, the generator is fine-tuned for a small number of epochs with both the cWGAN-GP loss and ambiguity loss.

without increasing computational complexity. Each convolution is followed by batch normalization and a rectified linear unit (ReLU) activation function. We replace batch normalization with layer normalization in the critic to prevent correlation between examples, which is incompatible with our gradient penalty [39].

An overview of our network architectures can be seen in Fig. 3. We utilize IQ data with real and imaginary components separated into two channels. The generator outputs a 1024-sample IQ output from its latent (z) and conditional (y) inputs, and it uses a $2\tanh(x/2)$ activation function to ensure its outputs are close to the range of the training data. The critic outputs a critic score from its waveform and conditional (y) inputs, where the waveform can be either a real (x) or generated signal ($\tilde{x}$). We select conditioning waveforms (y) that are the same class and SNR as their x counterparts. In a more realistic scenario, y could be chosen to be any examples that are temporally close to x (e.g., two signals captured in quick succession). The generator and critic have 12.2 and 11.7 million trainable parameters, respectively.

We train the generator and critic with the method shown in Fig. 4. We alternate training the generator and critic within each epoch with five critic iterations to every generator iteration. This ensures the critic is providing high-quality information to the generator. We initially train the generator to convergence without the ambiguity loss, and then fine tune for a smaller number of epochs with both loss terms. We train the generator using the Adam optimizer with a learning rate of 1×10^{-5}, and train the critic using root mean square propagation (RMSProp) with a learning rate of 5×10^{-5}. We use a global batch size of 512 and train on two NVIDIA Tesla V100 GPUs.

B. Ambiguity Loss

The WGAN critic enforces LPD performance by minimizing the distance between the RF background and generated distributions. We then include an *ambiguity loss* to simultaneously optimize the generated waveforms for desirable ambiguity characteristics. We define these desired ambiguity

characteristics with a target ambiguity function

$$\hat{A}_T(\tau, F_D) = \begin{cases} \chi & \text{if } (\tau, F_D) = (0,0) \\ 0 & \text{otherwise} \end{cases}, \quad (12)$$

which is an ideal "thumbtack" ambiguity (Fig. 2) with a variable target peak χ.

Our ambiguity loss is composed of two parts. First, we enforce the condition that the generated mainlobe amplitude must be *at least* as large as the target mainlobe:

$$L_{\text{main}} = \frac{1}{M}\|\text{ReLU}(\hat{A}_T(0,0) - \hat{A}_G(0,0))\|_2^2, \quad (13)$$

where M is batch size of the generated waveforms and $\hat{A}_G$ is a batch of ambiguity images computed from the generated waveforms $G(z_m|y_m)$. This does not penalize the network for a larger mainlobe amplitude. We then enforce the condition that all sidelobes should be pushed to zero, with increased weight for the zero-Doppler slice:

$$L_{\text{side}}(\tau, F_D) = \frac{1}{M}\|\hat{A}_T(\tau, F_D) - \Gamma \circ \hat{A}_G(\tau, F_D)\|_2^2, \quad (14)$$

evaluated for all $(\tau, F_D) \neq (0,0)$. Γ, or the *zero-Doppler weight*, is a weight matrix that increases the loss for the zero-Doppler slice by a factor of γ:

$$\Gamma = \begin{bmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \\ \gamma & \cdots & \gamma \\ 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{bmatrix},$$

which prevents optimizing to minima with low Doppler sidelobes and high zero-Doppler sidelobes. The mainlobe loss and sidelobe loss terms are then combined for our full ambiguity loss:

$$L_{\text{ambig}} = L_{\text{main}} + L_{\text{side}}(\tau, F_D)\big|_{(\tau, F_D) \neq (0,0)}. \quad (15)$$

We compute the ambiguity loss for 256 Doppler slices over a $\pm 10\%$ Doppler shift, yielding ambiguity images with 1024×257 resolution. Combining the ambiguity loss with the cWGAN-GP loss gives our total objective function

$$L_{\text{total}} = L_W + \eta L_{\text{ambig}}, \quad (16)$$

where η is a scaling parameter to set the strength of the ambiguity loss.

V. EXPERIMENTAL SETUP

We evaluate our generator network by testing its output diversity, detectability, and ambiguity characteristics. We first train the network on a simple, well-defined background distribution to test its diversity. We then train the network on a complex RF background and compare detectability and ambiguity characteristics of the generated waveforms against traditional LPD waveforms.

TABLE I
OVERVIEW OF SIDLE RADAR DATASET [42]

Class	Modulation Type	Code Length	Pulse Width (μs)
1	Barker	7	[0.875, 7]
2	Barker	11	[1.375, 11]
3	Barker	13	[1.625, 13]
4	Combined Barker	16	[1, 8]
5	Combined Barker	49	[3.08, 21.1]
6*	Combined Barker	169	[10.58, 84.6]
7†	Max. Len. Pseudo Rand.	15	[1, 4.5]
8†	Max. Len. Pseudo Rand.	31	[0.235, 10.5]
9†	Max. Len. Pseudo Rand.	63	[4.221, 18.9]
10	Min. Peak Sidelobe	7	[1.05, 4.2]
11	Min. Peak Sidelobe	25	[1.25, 10]
12	Min. Peak Sidelobe	48	[2.4, 19.2]
13†	T1	N/A	[2, 16]
14†	T2	N/A	[1.5, 12]
15†	T3	N/A	[1, 8]
16†	Polyphase Barker	7	[0.875, 7]
17†	Polyphase Barker	20	[1, 8]
18†	Polyphase Barker	40	[2, 16]
19†	P1	N/A	[5, 20]
20†	P2	N/A	[3.2, 25.6]
21†	P3	N/A	[3.2, 25.6]
22†	P4	N/A	[5, 20]
23	Minimum Shift Key	63	[2, 18.9]

*Class omitted from experiments   †LPD modulation

A. RF Background

Let $\mathbb{P}_b$ define the RF background distribution. In realistic scenarios, this distribution can be quite complex, encompassing a variety of RF emitters and their environmental interactions. A complex background distribution provides our generator with a high number of degrees of freedom for its generated distribution, potentially leading to more effective LPD performance. However, this complex distribution can be difficult to quantitatively evaluate. Because of this, we test our generator with two representative cases: (1) a simple toy dataset of linear frequency modulated (LFM) waveforms, and (2) a complex dataset of realistic communication waveforms.

We initially define the RF background using a simple toy dataset of simulated LFM waveforms, or linear "chirps," so that we can understand the generated distribution's characteristics in a controlled setting. A linear chirp is defined as

$$s_{\text{chirp}}(t) = A \exp(j\pi\beta t^2/T), \quad (17)$$

where A is the amplitude, β is the sweep bandwidth, and T is the pulse duration. We generate our chirp dataset with examples defined as

$$x_{\text{chirp}}(t) = \exp(j\pi B t^2) + n(t), \quad (18)$$

where B, the sweep bandwidth, is a discrete random variable drawn from a uniform distribution of integer values $B \sim U(5, 19)$. We use a normalized pulse duration $T = 1$, such that the time-bandwidth product BT of our generated dataset lies in the range of $[5, 19]$. We introduce additive white Gaussian noise as $n(t) \sim \mathcal{N}(0, \sigma^2)$. Our generated dataset contains 50,000 waveforms with 1024 IQ time samples at 30 dB SNR. We use this simple chirp dataset to demonstrate the distribution-matching capabilities of our WGAN.
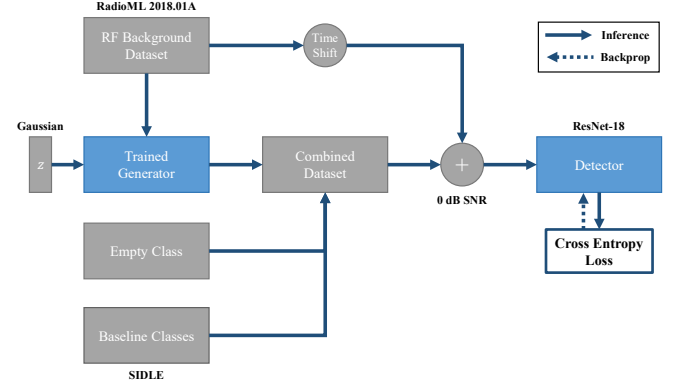

Fig. 5. **Evaluation Framework.** Training process for the detector neural network used to evaluate LPD performance. We add two additional classes to SIDLE—generated and empty—to create the *combined dataset*. We add randomly time-shifted RF background signals to these examples using waveforms from RadioML 2018.01A. We train a ResNet-18 detection network to identify if a radar signal is present or not.

We then formulate a more complex problem by defining the RF background distribution with common communication waveforms from the DeepSig RadioML 2018.01A dataset [43]. This dataset contains over-the-air recordings of the following 24 digital and analog modulations with synthetic simulated channel effects:

> OOK, ASK4, ASK8, BPSK, QPSK, PSK8, PSK16, PSK32, APSK16, APSK32, APSK64, APSK128, QAM16, QAM32, QAM64, QAM128, QAM256, AM_SSB_WC, AM_SSB_SC, AM_DSB_WC, AM_DSB_SC, FM, GMSK, and OQPS.

Each class contains 106,496 waveforms with variable SNRs from –20 to 30 dB in increments of 2 dB. Each waveform consists of 1024 IQ time samples.

B. LPD Effectiveness

We evaluate the success of our LPD waveform design method by comparing the detectability of our generated waveforms to common radar waveforms from the Air Force Research Laboratory's Signal Instantiations for Deep Learning Experiments (SIDLE) dataset [42], [44], [45]. This dataset contains simulations of 23 radar modulations, standard and LPD, with variable pulse widths shown in Table I. Each class contains 11,000 clean IQ waveforms, and we randomly select a 1024-sample segment from each waveform. For waveforms with less than 1024 samples, we randomly delay the waveform within a 1024 sample frame. We omit class 6 from this dataset as its longer pulse widths required too much memory to use effectively.

We compare the detectability of generated and SIDLE waveforms by training a separate detection neural network and evaluating its performance on the binary detection problem. Neural networks have been shown to be excellent modulation classifiers [46]–[48], and generating waveforms that can deceive them (without explicit adversarial attack) is a difficult task.

The detector network's goal is to determine if any radar signal is present within a given input signal in the presence
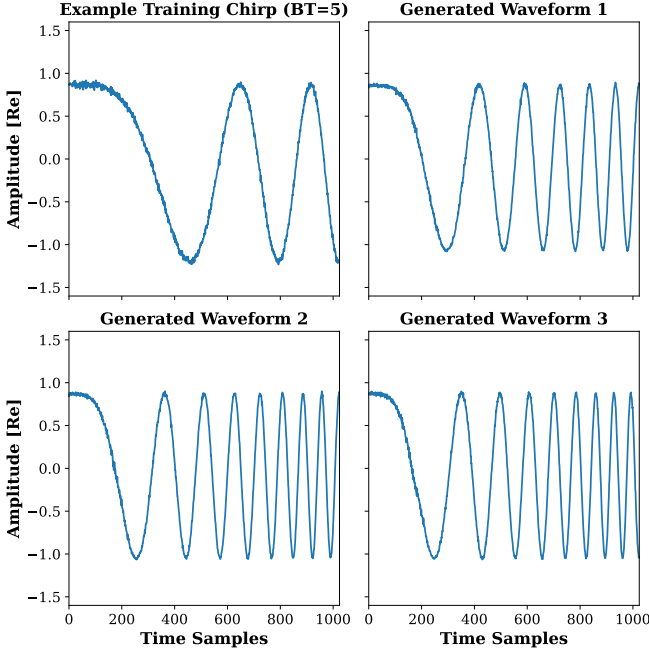
Fig. 6. **Generated Chirp Toy Examples.** We compare a linear chirp from the simulated training dataset (top left) to generated chirps with variable time-bandwidth products from our trained WGAN-GP. Generated examples closely match the distribution of the training dataset.

of RF background noise (i.e., it is performing detection, not classification). We train a ResNet-18 neural network [49] on a *combined dataset* equally composed of radar waveforms (SIDLE and generated) and an empty class with no signals present. Examples from the *combined dataset* are combined with randomly time-shifted RF background signals from the RadioML dataset at 0 dB SNR, and are then standardized to 0 mean, 1 standard deviation. This process is summarized in Fig. 5.

We train the detector with an 80%/20% train/test split—sampled equally across SNRs and classes—and we generate new waveforms from our generator with every training and testing epoch. This exposes the detector to the full distribution of generated waveforms $\mathbb{P}_g$, rather than a subset of fixed waveforms. We randomly time shift the RF background signals to ensure our generator does not have oracle time information. We also split the RadioML dataset into two equal parts: one for training our generator and a second for training the detector. This ensures our generator is not trained on any waveforms used to evaluate its LPD effectiveness.

C. Sensing Metrics

We evaluate the sensing performance of our generated waveforms by comparing their ambiguity characteristics to the SIDLE waveforms. The goal of our ambiguity loss is to optimize our generated waveforms toward the ideal "thumb-tack" ambiguity. The resulting waveforms should have low ambiguity in delay (range) and Doppler (velocity) with low sidelobes. We use three ambiguity function metrics to evaluate sensing performance: mainlobe width, peak sidelobe level (PSL), and integrated Doppler ambiguity. The mainlobe width
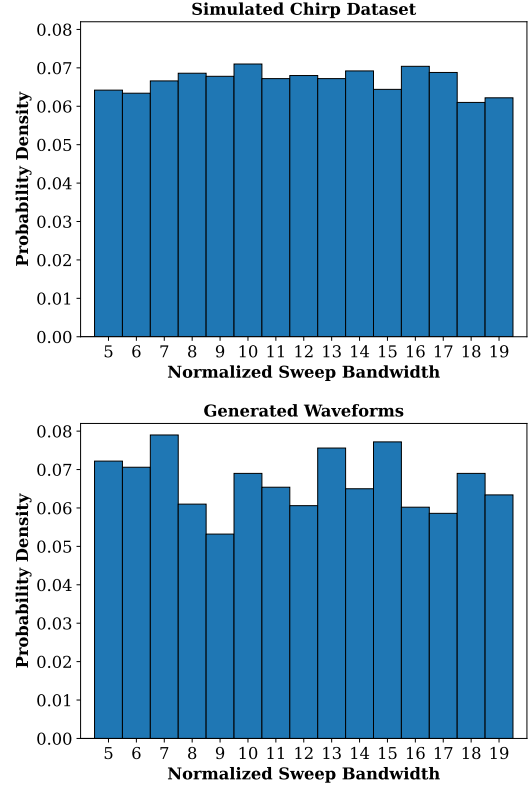

Fig. 7. **Generator Distribution Matching.** Comparison of the sweep bandwidths present within the chirp training dataset (top) and the sweep bandwidths of 5000 waveforms produced by the trained generator (bottom). The training dataset is created with a uniform distribution, and the generated distribution closely matches it.

and PSL evaluate performance in the zero-Doppler regime, and the integrated Doppler evaluates the overall Doppler ambiguity response of the waveforms.

Mainlobe width (Φ_{MW}) describes the half-power width of the ambiguity mainlobe, and it is an important metric for radar applications [50]. A narrower mainlobe (central peak) allows for better resolution in range and velocity, thus making it easier to distinguish between closely spaced targets. We define mainlobe width as

$$\Phi_{MW}\left[\hat{A}(\tau, F_D)\big|_{F_D=0}\right] = \tau_2 - \tau_1, \tag{19}$$

where τ_1 and τ_2 are given by

$$\hat{A}(\tau_1, 0) = \hat{A}(\tau_2, 0) = \frac{1}{2} \cdot \hat{A}(0, 0) \tag{20}$$

for $\tau_1 \in \left[-\tau_{\text{main}}, 0\right]$ and $\tau_2 \in \left[0, \tau_{\text{main}}\right]$. Here, the interval $\left[-\tau_{\text{main}}, \tau_{\text{main}}\right]$ contains the full mainlobe centered on $\tau = 0$.

PSL (Φ_{PSL}) is the ratio of the highest amplitude sidelobe to the amplitude of the mainlobe [6]. High sidelobes can cause unwanted interference and mask weaker signals near strong reflectors, so a high PSL can lead to decreased detection capability for objects of interest. PSL is defined as

$$\Phi_{PSL}\left[\hat{A}(\tau, F_D)\big|_{F_D=0}\right] = \max_{\tau}\left|\frac{\hat{A}(\tau, 0)}{\hat{A}(0, 0)}\right| \tag{21}$$

for $\tau \in \left[\tau_{\text{main}}, T\right]$, where the interval $\left[\tau_{\text{main}}, T\right]$ contains all sidelobes.
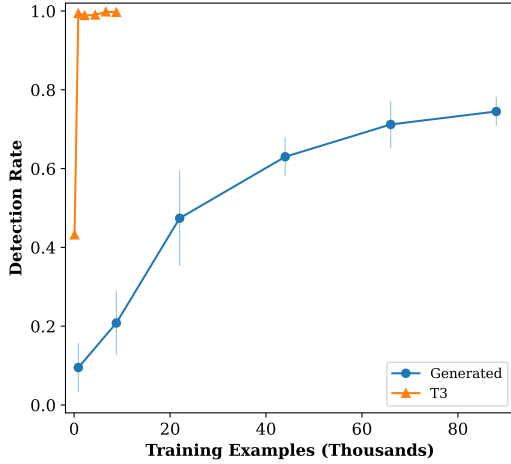
Fig. 8. **Detection Rate vs Data Quantity.** Detection rate comparison for T3 and generated waveforms with variable training examples seen by the detection network. Standard deviation over 10 runs shown as vertical lines. Generated waveforms significantly outperform best SIDLE waveform (T3) even with $10\times$ more training data.
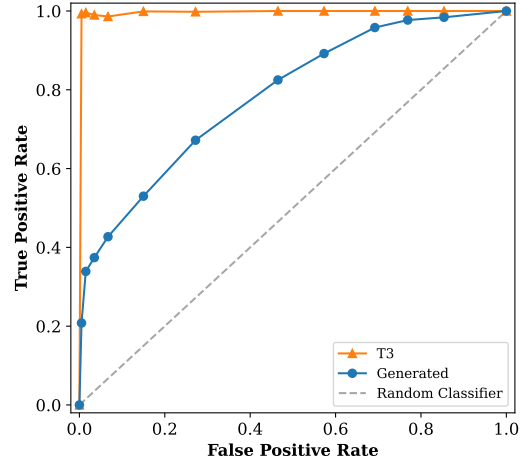

Fig. 9. **ROC Curve.** Receiver operating characteristic (ROC) curves for trained binary detection network comparing performance on T3 and generated waveforms. Generated waveforms significantly degrade detector performance.

Integrated Doppler ambiguity (Φ_{ID}) is the normalized sum of all Doppler terms of the ambiguity function. While not a standard metric, we use it to compare the global Doppler ambiguity response of our waveforms. A larger integrated Doppler value means a waveform has greater Doppler ambiguity. We define integrated Doppler ambiguity as

$$\Phi_{ID}\left[\hat{A}(\tau, F_D)\right] = \sum_{\tau=-T}^{T} \sum_{\substack{F_D=-F \\ F_D \neq 0}}^{F} \left|\frac{\hat{A}(\tau, F_D)}{\hat{A}(0, 0)}\right|^2 \qquad (22)$$

where the ambiguity function is computed over the delay and Doppler ranges of $\tau \in \left[-T, T\right]$ and $F_D \in \left[-F, F\right]$, respectively.

VI. EXPERIMENTAL RESULTS

A. Qualitative Example

We first explore the qualitative behavior of our generative network to ensure it can accurately learn RF distributions. We train the WGAN-GP on our linear chirp dataset with the non-conditional WGAN-GP objective (Eq. 10). We remove the ambiguity loss term from training (Eq. 15) to isolate performance under the WGAN-GP objective.

Once trained, the WGAN-GP accurately reproduces examples from the linear chirp distribution. Examples of generated waveforms can be seen in Fig. 6. To verify the generated distribution, we extract the time-bandwidth product from 10,000 generated chirps and compare them in Fig. 7. The simulated chirp dataset was generated with a uniform distribution of integer time-bandwidth products in the range $[5, 19]$, and the generated distribution closely matches that uniform distribution.

B. Detectability

We then extend our framework to the complex RF background dataset, RadioML 2018.01A, to evaluate LPD performance. We have shown the network can effectively minimize the distance between training and generated RF distributions. Since this distance is an effective surrogate for detectability, our generated waveforms should be more difficult to detect.

We test this hypothesis by training our conditional WGAN-GP on the 50% training split of RadioML with the cWGAN-GP loss (Eq. 11). We omit the ambiguity loss to isolate the effects of the cWGAN-GP loss term. We compare the performance of a separately trained detection network on detecting our generated waveforms and common radar waveforms from the SIDLE dataset. These waveforms are summed to the 50% test split of RadioML at 0 dB SNR.

We find that our generator is able to produce effective LPD waveforms. The generator produces new waveforms with every training epoch, so its effectiveness is dependent on the amount of training examples seen by the detector, as shown in Fig. 8. For an equal number of training examples as the SIDLE classes (8800), our generated waveforms maintain an average detection rate of $20.8\% \pm 8.1\%$. All SIDLE examples are detected with nearly 100% accuracy—the T3 waveforms are the most difficult to detect with a detection rate of 99.7%.

When the detector is exposed to $10\times$ more generated data (88,000 training examples), detection rate is increased to 74.5%. This is expected, as prolonged exposure to generated waveforms will allow the detector to more accurately learn the generated distribution. However, this behavior plateaus, with $100\times$ more generated data yielding an 80.8% detection rate. This shows that the generated waveforms maintain improved LPD performance even with significantly more training data.

We compare the receiver operating characteristic (ROC) curves for the generated and T3 waveforms in Fig. 9 using a detector trained on an equal number of examples for SIDLE and generated waveforms (8800 per class). To increase the false positive rate for these curves, we retrain the detector multiple times with a decreasing ratio of empty labels in the training dataset. We find that the T3 curve reflects the behavior of a perfect detector, and the generated waveforms significantly reduce the performance of the detector.
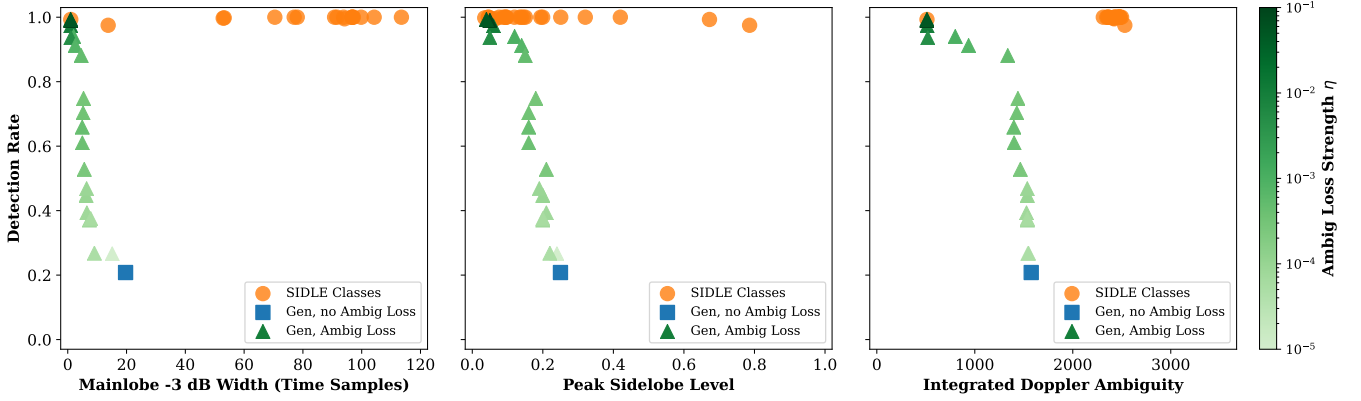
Fig. 10. **Tunable Sensing Performance.** Shown are detectability vs sensing metrics for LPD generators that have been fine-tuned with variable ambiguity loss strengths (shades of green). They are compared to the baseline SIDLE classes (orange), and a baseline generator without ambiguity loss fine-tuning (blue). Leftmost plot's x-axis represents mainlobe width—lower implies better range resolution. Center plot's x-axis represents peak sidelobe level—lower implies better target separation. Rightmost plot's x-axis represents integrated Doppler ambiguity—lower implies less velocity ambiguity. Adjusting the ambiguity loss strength (η) allows one to control the trade-off between sensing ability and detectability. Increasing the ambiguity loss strength yields narrower mainlobes, improved peak sidelobe level, and lower Doppler ambiguity at the cost of increased detectability. All values of η yield waveforms that offer comparable or improved sensing metrics compared to the SIDLE classes while being harder to detect.

C. Sensing Performance

We fine tune our generator for 3,000 iterations with both the cWGAN-GP loss and ambiguity loss (Eq. 15) to improve the ambiguity characteristics of the generated waveforms. We evaluate the effectiveness of the ambiguity loss term by computing the mainlobe width (MW), peak sidelobe level (PSL), and integrated Doppler ambiguity (IDA) of the generated waveforms conditioned on all RF background classes at $\{-10, 0, 10, 30\}$ dB SNR. We compare these to the same metrics computed for the SIDLE waveforms. Because of the variable signal length of the SIDLE waveforms, we zero pad all waveforms to 11,000 time samples prior to computing their ambiguity function.

We find that the ambiguity loss improves the MW, PSL, and IDA of the generated waveforms. To understand this in the context of LPD, we evaluate each model's detectability with a detection network trained on an equal number of SIDLE and generated examples (8800 per class). The resulting detectability is shown as a function of MW, PSL, and IDA in Fig. 10. The generators used in these figures were fine-tuned with a variable ambiguity loss strength η, target peak $\chi = 1000$, and zero-Doppler weight $\gamma = 1000$. Larger η leads to improved MW, PSL, and IDA at the cost of greater detectability. However, generated waveforms with no ambiguity loss retain MW, PSL, and IDA values comparable or better than those of the SIDLE waveforms. A comparison of ambiguity examples from SIDLE and the LPD generator can be seen in Fig. 11.

VII. DISCUSSION

Comparison to White Noise. A natural comparison to our method is the use of Gaussian noise-based waveforms as in noise radar. Gaussian noise waveforms are a special case of our method when the RF background is dominated by negative SNR signals. In this case, our framework converges to white noise as the optimal LPD waveform. When we compare the detectability of generated waveforms to white noise for –20–0 dB SNR RF background signals, generated waveform detectability is 13% while white noise is below the false positive rate, effectively 0%. In contrast to noise radar, our generator generalizes to cases where the RF background contains positive SNR signals. For 6–30 dB SNR RF background signals, our generated waveforms achieve a 38.6% detection rate while white noise is detected at a rate of 92.7%.

Bandwidth Requirements. When the RF background contains positive SNR signals, our generated waveforms have equivalent bandwidth to those background signals. However, when the RF background contains negative SNRs, our generated waveforms incorporate significantly more white noise. This increases the bandwidth requirements for our generated signals up to the sample rate of our inputs. The end-to-end nature of our generation framework allows for a third loss term to constrain the generated bandwidth—however, we've found this to be nontrivial and requiring additional investigation. Future work should explore input and output bandwidth considerations for hardware integration.

Conditional Class Performance. The distribution-matching nature of our framework necessitates that its outputs are limited by the signals present in the RF background. While we have shown that this is effective for a large number of background signals, not all signals are equally effective. Specifically, we have found that generated LPD waveforms conditioned on AM and FM modulations have lower sensing performance than other classes. Omitting AM and FM conditional classes from our generator yields an overall 35% improvement in mainlobe width, 12% improvement in PSL, and no change in IDA.

VIII. CONCLUSION

We have successfully developed an unsupervised technique for generating radar waveforms with low probability of detection (LPD). We utilize a generative adversarial framework to design waveforms that mimic the ambient radio frequency (RF) environment, making them difficult to distinguish from
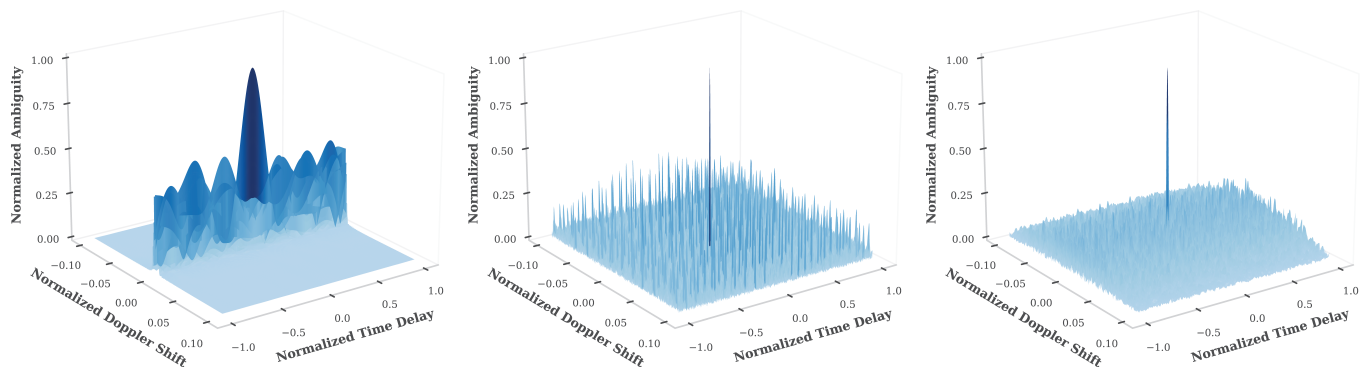
Fig. 11. **Ambiguity Comparison.** Example ambiguity images from SIDLE Barker 7 (left), SIDLE T3 (center), and our LPD waveform generator (right). The generated waveform is created with ambiguity loss strength $\eta = 0.0001$ conditioned on a BPSK background signal at 10 dB SNR. The generated waveform retains a narrow mainlobe with low sidelobes for improved range/velocity resolution.

the RF background. To maintain the sensing capabilities of these waveforms, we introduce and optimize an ambiguity function-based loss metric. Our evaluation demonstrates that the generated waveforms achieve up to a 90% reduction in detectability when compared to traditional LPD waveforms and retain desirable ambiguity characteristics. Furthermore, we are able to tune the generated waveforms for a desired trade-off between detectability and sensing accuracy.

This generative framework offers substantial versatility in waveform design beyond just LPD. By adjusting both the WGAN's target distribution and the target shape of the ambiguity loss, we can tailor the design to meet specific operational requirements. The inclusion of other objectives—e.g., a bandwidth constraint—may enable further applications such as anti-interference. Given the dynamic landscape of technological advancements and evolving RF conditions, these generative techniques offer a new tool to ensure radar systems remain both effective and adaptive across a wide array of requirements and scenarios.

ACKNOWLEDGMENT

REFERENCES

[1] A. De Martino, *Introduction to Modern EW Systems*, 2nd ed., ser. Electronic Warfare Library. Artech House, 2018.

[2] G. Liu, H. Gu, W. Su, and H. Sun, "The analysis and design of modern low probability of intercept radar," in *CIE International Conference on Radar Proceedings*. IEEE, Oct. 2001, pp. 120–124.

[3] D. Lynch, Jr, *An Introduction to RF Stealth*, 2nd ed. Scitech Publishing, Aug. 2021.

[4] A. F. Lime, Jr, "Analysis of low probability of intercept (LPI) radar signals using cyclostationary processing," Master's thesis, Naval Postgraduate School, Monterey, CA, 2002.

[5] X. Liu, T. Zhang, X. Yu, Q. Shi, G. Cui, and L. Kong, "LPI waveform design for radar system against cyclostationary analysis intercept processing," *Signal Processing*, vol. 201, Dec. 2022.

[6] S. D. Blunt and E. L. Mokole, "Overview of radar waveform diversity," *IEEE Aerospace and Electronic Systems Magazine*, vol. 31, no. 11, pp. 2–42, Nov. 2016.

[7] A. F. Martone, K. D. Sherbondy, J. A. Kovarskiy, B. H. Kirk, J. W. Owen, B. Ravenscroft, A. Egbert, A. Goad, A. Dockendorf, C. E. Thornton, R. M. Buehrer, R. M. Narayanan, S. Blunt, and C. Baylis, "Practical aspects of cognitive radar," in *2020 IEEE Radar Conference (RadarConf20)*, Sep. 2020, pp. 1–6.

[8] A. F. Martone and A. Charlish, "Cognitive radar for waveform diversity utilization," in *2021 IEEE Radar Conference (RadarConf21)*. IEEE, May 2021, pp. 1–6.

[9] A. J. Uppal, M. Hegarty, W. Haftel, P. A. Sallee, H. Brown Cribbs, and H. H. Huang, "High-Performance deep learning classification for radio signals," in *2019 53rd Asilomar Conference on Signals, Systems, and Computers*. IEEE, Nov. 2019, pp. 1026–1029.

[10] Z. Pan, B. Wang, R. Zhang, S. Wang, Y. Li, and Y. Li, "MIML-GAN: A GAN-Based algorithm for Multi-Instance Multi-Label learning on overlapping signal waveform recognition," *IEEE Transactions on Signal Processing*, vol. 71, pp. 859–872, 2023.

[11] G. E. Smith and T. J. Reininger, "Reinforcement learning for waveform design," in *2021 IEEE Radar Conference (RadarConf21)*, May 2021, pp. 1–6.

[12] A. Charlish and C. Schwalm, "Generating NLFM radar waveforms using variational autoencoders," in *2022 IEEE Radar Conference (RadarConf22)*, Mar. 2022, pp. 1–6.

[13] M. S. Lee, "Generating noise-like communication codes with deep learning," DEVCOM Army Research Laboratory, Tech. Rep. ARL-TR-9749, Aug. 2023.

[14] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Advances in Neural Information Processing Systems*, Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, and K. Q. Weinberger, Eds., vol. 27. Curran Associates, Inc., 2014.

[15] L. Ruthotto and E. Haber, "An introduction to deep generative modeling," *GAMM-Mitt.*, vol. 44, no. 2, Jun. 2021.

[16] N. Levanon and E. Mozeson, *Radar Signals*. John Wiley & Sons, Inc., Jul. 2004.

[17] W. L. Melvin and J. A. Scheer, Eds., *Principles of Modern Radar: Advanced Techniques*. Edison, NJ: SciTech Publishing, 2013, vol. 2.

[18] A. G. Stove, A. L. Hume, and C. J. Baker, "Low probability of intercept radar strategies," *IEE Proceedings - Radar, Sonar and Navigation*, vol. 151, no. 5, pp. 249–260, Oct. 2004.

[19] R. M. Narayanan, "Chapter 9 - noise radar techniques and progress," in *Advanced ultrawideband radar: Signals, targets, and applications*, J. D. Taylor, Ed. Boca Raton, FL: CRC Press, Dec. 2016.

[20] S. Z. Gurbuz, H. D. Griffiths, A. Charlish, M. Rangaswamy, M. S. Greco, and K. Bell, "An overview of cognitive radar: Past, present, and future," *IEEE Aerospace and Electronic Systems Magazine*, vol. 34, no. 12, pp. 6–18, Dec. 2019.

[21] A. O. Hero and D. Cochran, "Sensor management: Past, present, and future," *IEEE Sens. J.*, vol. 11, no. 12, pp. 3064–3075, Dec. 2011.

[22] J. Ward, "Space-time adaptive processing for airborne radar," Lincoln Laboratory, Massachusetts Institute of Technology, Tech. Rep. ESC-TR-94-109, Dec. 1998.

[23] D. J. Kershaw and R. J. Evans, "Optimal waveform selection for tracking systems," *IEEE Trans. Inf. Theory*, vol. 40, no. 5, pp. 1536–1550, Sep. 1994.

[24] H.-S. Kim and N. A. Goodman, "Waveform design by Task-Specific information," in *2010 IEEE Radar Conference*. IEEE, May 2010, pp. 848–852.

[25] IEEE Aerospace and Electronic Systems Society, "IEEE standard for radar definitions," IEEE Standards Association, Tech. Rep. 686-2017, Sep. 2017.

[26] S. Haykin, Y. Xue, and P. Setoodeh, "Cognitive radar: Step toward bridging the gap between neuroscience and engineering," *Proc. IEEE*, vol. 100, no. 11, pp. 3102–3130, Nov. 2012.

[27] NATO Task Group SET-227, "Cognitive radar," NATO Science and Technology Organization, Tech. Rep. AC/323(SET-227)TP/947, Oct. 2020.

[28] C. J. Pici and R. M. Narayanan, "Design of spectrally adaptive noise radar waveforms," in *Radar Sensor Technology XXIII*, vol. 11003. SPIE, May 2019, pp. 161–169.

[29] C. E. Thornton, M. A. Kozy, R. Michael Buehrer, A. F. Martone, and K. D. Sherbondy, "Deep reinforcement learning control for radar detection and tracking in congested spectral environments," *IEEE Transactions on Cognitive Communications and Networking*, vol. 6, no. 4, pp. 1335–1349, Dec. 2020.

[30] V. Saarinen and V. Koivunen, "MIMO radar waveform synthesis using generative adversarial networks," in *2023 IEEE 33rd International Workshop on Machine Learning for Signal Processing (MLSP)*. IEEE, Sep. 2023, pp. 1–6.

[31] A. Charlish and C. Schwalm, "Generating NLFM radar waveforms using variational autoencoders," in *2022 IEEE Radar Conference (RadarConf22)*, Mar. 2022, pp. 1–6.

[32] M. A. Richards, *Fundamentals of Radar Signal Processing*, 3rd ed. McGraw-Hill Education, Apr. 2022.

[33] J. Chen, J. Wang, Y. Zhang, F. Wang, and J. Zhou, "Spatial Information-Theoretic optimal LPI radar waveform design," *Entropy*, vol. 24, no. 11, Oct. 2022.

[34] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," in *Advances in Neural Information Processing Systems*, H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin, Eds., vol. 33. Curran Associates, Inc., 2020, pp. 6840–6851.

[35] D. P. Kingma and M. Welling, "Auto-Encoding Variational Bayes," in *International Conference on Learning Representations (ICLR)*, 2014.

[36] L. Dinh, J. Sohl-Dickstein, and S. Bengio, "Density estimation using real NVP," in *International Conference on Learning Representations*, 2017.

[37] M. Arjovsky and L. Bottou, "Towards principled methods for training generative adversarial networks," in *International Conference on Learning Representations*, 2017.

[38] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein generative adversarial networks," in *Proceedings of the 34th International Conference on Machine Learning - Volume 70*, Aug. 2017, pp. 214–223.

[39] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. C. Courville, "Improved training of wasserstein GANs," in *Advances in Neural Information Processing Systems*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds., vol. 30. Curran Associates, Inc., 2017.

[40] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, "PyTorch: An imperative style, high-performance deep learning library," *Adv. Neural Inf. Process. Syst.*, vol. 32, 2019.

[41] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 1492–1500.

[42] B. D. Rigling and C. Roush, "ACF-based classification of phase modulated waveforms," in *2010 IEEE Radar Conference*, May 2010, pp. 287–291.

[43] T. J. O'Shea, T. Roy, and T. C. Clancy, "Over-the-Air deep learning based radio signal classification," *IEEE J. Sel. Top. Signal Process.*, vol. 12, no. 1, pp. 168–179, Feb. 2018.

[44] A. Pavy and B. Rigling, "SV-Means: A fast SVM-Based level set estimator for Phase-Modulated radar waveform classification," *IEEE J. Sel. Top. Signal Process.*, vol. 12, no. 1, pp. 191–201, Feb. 2018.

[45] R. V. Chakravarthy, H. Liu, and A. M. Pavy, "Open-Set radar waveform classification: Comparison of different features and classifiers," in *2020 IEEE International Radar Conference (RADAR)*, Apr. 2020, pp. 542–547.

[46] M. Zhang, M. Diao, L. Gao, and L. Liu, "Neural networks for radar waveform recognition," *Symmetry*, vol. 9, no. 5, p. 75, May 2017.

[47] M. R. Ziemann and C. A. Metzler, "Transformers for robust radar waveform classification," in *2022 56th Asilomar Conference on Signals, Systems, and Computers*, Oct. 2022, pp. 759–763.

[48] Z. Ma, W. Yu, P. Zhang, Z. Huang, A. Lin, and Y. Xia, "LPI radar waveform recognition based on neural architecture search," *Comput. Intell. Neurosci.*, vol. 2022, p. 4628481, Jan. 2022.

[49] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*. openaccess.thecvf.com, 2016, pp. 770–778.

[50] M. A. Richards, J. Scheer, and W. A. Holm, *Principles of Modern Radar*. SciTech Publishing, 2010.