DOI:10.19652/j.cnki.femt.2305721

基于 AdvGAN 的无线通信信号反侦察方法*

陈奕功[1,2] 张 江[2] 乔晓强[2] 龙伟军[1]
(1.南京信息工程大学电子与信息工程学院 南京 210044;2.国防科技大学第六十三研究所 南京 210007)

摘 要:深度学习技术凭借其强大的特征提取能力,在信号认知方面取得广泛应用,这对有保密需求的无线通信系统的保密性安全带来极大威胁。针对上述问题,提出一种基于对抗生成网络(generating adversarial examples with adversarial networks,AdvGAN)的无线通信信号反侦察方法。首先实现两种不同的调制信号识别模型;再使用3种对抗样本生成方法构造伪装信号;最后叠加在原始信号上并在调制信号识别模型上进行测试。实验结果表明,所提方法能够使侦收方的智能调制识别模型的识别准确率大幅下降,在信噪比10 dB条件下,使侦收方未知模型识别准确率下降约66%,从而有效反制侦收方的智能识别模型。

关键词:调制信号识别;深度学习;通信反侦察;信号伪装;生成对抗网络
中图分类号:TN92;TP183 **文献标识码**:A **国家标准学科分类代码**:510.5015

Anti-reconnaissance method for wireless communication signals based on AdvGAN

Chen Yigong[1,2] Zhang Jiang[2] Qiao Xiaoqiang[2] Long Weijun[1]
(1. School of Electronic and Information Engineering, Nanjing University of Information Science and Technology, Nanjing 210044, China; 2. The 63rd Research Institute of National University of Defense Technology, Nanjing 210007, China)

Abstract: Deep learning technology has been widely used in signal recognition by virtue of its powerful feature extraction capability, which brings great threat to the confidentiality security of wireless communication systems with confidentiality needs. To address the above problems, this paper proposes a wireless communication signals anti-reconnaissance method based on generating adversarial examples with adversarial networks (AdvGAN). Firstly, two different modulated signal recognition models are realized. Then three antagonistic sample generation methods are used to construct camouflage signals. Finally, they are superimposed on the original signals and tested on the modulated signal recognition model. The experimental results show that the method proposed in this paper can make the recognition accuracy of the intelligent modulation recognition model of the reconnaissance side drop dramatically, and make the recognition accuracy of the unknown model of the reconnaissance side drop by about 66% under the condition of SNR is 10 dB, so as to effectively counteract the intelligent recognition model of the reconnaissance side.

Keywords: modulated signal recognition; deep learning; communication anti-reconnaissance; signal camouflage; generative adversarial networks

0 引 言

近年来,随着深度学习技术在图像和语音信号处理领域的成功应用,其在无线信号领域也越来越受到关注,并取得诸多研究成果,主要体现在信号认知领域,其典型应用有自动调制识别(automatic modulation recognition, AMR)和辐射源个体识别[1-2]等。

2016年,O'Shea等[3]使用GNU Radio生成无线电机器学习数据集[4],首次研究了卷积神经网络(convolutional neural networks,CNN)在复值时间无线电信号领域的自适应问题,证明了CNN对信号时间序列进行识别分类的可行性。王超等[5]使用VGG16模型并针对调制信号数据

收稿日期:2023-10-31
* **基金项目**:国家自然科学基金(62371463,62071440)项目资助

集进行模型调整，提高了高信噪比(signal-to-noise ratio，SNR)下的识别准确率并以此进行对抗攻击实验。杨小蒙等[6]对信号的时序特征和空间特征进行特征串联融合，充分利用不同模态特征之间的互补性，提高了不同信噪比下的识别准确率。除了在信号序列上可以有效的完成一系列信号识别分类问题，文献[7-8]分别利用信号的星座图和等势星球图，将调制分类问题转换为为图像分类问题，在一定程度上改善了低信噪比下的识别准确率。

上述技术的快速发展对有保密需求的无线通信系统的保密性安全带来极大挑战。提高无线通信系统保密性安全的方法可分为信息域方法和信号域方法两类。其中信息域技术主要包括各类加密方法，信号域方法主要包括时域、频域、码域、功率域以及波形域的各类反侦察、抗截获技术[9]，如跳频、扩频等技术。本文重点关注信号域方法，随着宽带数字接收、数字信号处理技术以及深度学习所驱动的智能侦察技术的快速发展，现已突破针对上述抗截获通信系统的电子侦察和截获技术，因此需要从新的角度考虑无线通信反侦察、抗截获问题。

“对抗样本”具备加入微小扰动使智能模型发生误判的能力，为对抗智能信号认知系统提供了新的思路，为反侦察提供了新的技术途径。2013 年，Szegedy 等[10]首次提出“对抗样本”这一概念，对输入图像添加微小扰动，可以导致分类模型产生错误的决策。2015 年，Goodfellow 等[11]提出了快速梯度符号算法(fast gradient sign method，FGSM)来构造对抗样本，能够欺骗图像分类模型。Kurakin 等[12]在 FGSM 算法基础上提出基本迭代法(basic iterative method，BIM)，通过对输入样本添加多次迭代的微小扰动从而优化攻击效果。Madry 等[13]在 BIM 算法基础上提出投影梯度下降法(projected gradient descent，PGD)，通过将每次迭代后的对抗样本投影回到一个预先确定的范围内，以提升攻击强度。

基于上述分析可见，对抗样本的研究在图像分类领域已经取得了丰硕的理论和应用成果，但在电磁空间领域的研究与应用还处于初步阶段。Sadeghi 等[14]首次将对抗样本应用于调制信号识别任务，通过向原始样本添加微小扰动，使智能调制识别模型识别错误。为了验证传统对抗攻击算法在无线通信领域的可行性，Lin 等[15]将 FGSM、BIM、PGD 等经典攻击算法应用到调制信号识别任务并对比攻击效果，验证了基于深度学习的调制信号识别模型容易受到来自对抗样本的威胁。

目前，对抗样本在无线信号领域的应用较少，主要是以第 3 方向频谱监测接收机注入对抗样本的方式使其识别率大幅下降[16]。本文从全新的应用场景出发，提出一种基于 AdvGAN 的无线通信信号反侦察方法，在基带利用生成对抗网络生成微小的伪装信号，并将其叠加到通信信号上共同发射。本文实现了基于 CNN 和 VGG16 的两种深度学习调制识别模型，取得了较高的识别准确率，将这两种模型作为侦收方的智能分类器。实现了一种基于 AdvGAN 的无线通信信号反侦察方法，分别在白盒识别模型和黑盒识别模型两种条件下，验证所提方法的有效性，并与使用 FGSM、PGD 生成伪装信号的反侦察效果进行了对比。

1 系统模型与方法描述

1.1 系统模型

本文所提出的反侦察系统模型如图 1 所示。

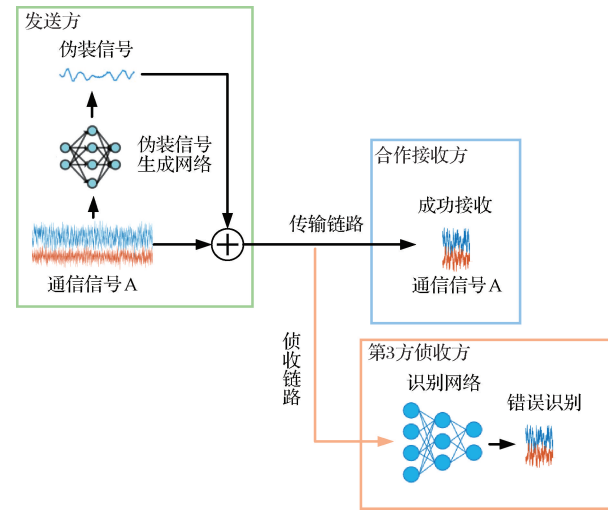


图 1 反侦察系统模型

Fig. 1 The model of anti-reconnaissance system

首先，发送方在完成信源编码、信道编码以及数字调制后生成基带调制信号 $x(t)$，然后针对敌方智能调制识别模型生成伪装信号 $\eta(t)$，并将其叠加到基带调制信号上形成伪装后的基带调制信号 $y(t)=x(t)+\eta(t)$，最后经上变频后通过天线发射出去，发送系统框图如图 2 所示。

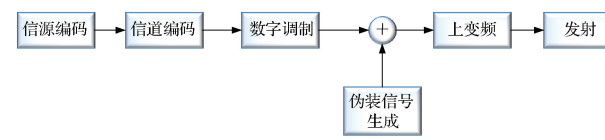


图 2 发送系统框图

Fig. 2 The block diagram of transmission system

为保证合法接收方接收性能，同时使侦收方智能调制识别模型识别发生错误，生成的伪装信号需满足如下两个条件。

1) 伪装信号功率足够小，不会对通信系统性能带来太大影响。

2) 伪装信号能够使侦收方智能调制识别模型发生误判。

基于上述约束，伪装信号的生成可借鉴对抗样本中对抗性扰动的构造方法。传统的对抗样本生成方法，如

FGSM 和 PGD，需要对扰动信号的功率进行约束，无法保证以最小的功率实现最优的对抗效果。虽然其在图像领域取得了成功应用，但在通信领域，基于上述方法产生的对抗样本会对合作通信方的接收性能带来较大影响，因此不适合用于生成伪装信号。

生成对抗网络[17]（generative adversarial networks，GAN）凭借生成器和判别器的对抗性训练，具备出色的生成能力，能够实现不同分布信号之间的映射，为伪装信号的生成提供了的思路。

1.2 方法描述

AdvGAN[18]是一种改进型生成对抗网络，对损失函数设计和应用场景进行了改进。本文方法基于 AdvGAN 的伪装信号生成网络结构如图 3 所示，通过加入的调制识别模型损失项，对生成器和判别器进行对抗性训练，学习从原始信号到伪装信号的映射，生成微小的伪装信号的同时能使目标模型识别准确率大幅下降，且不需要访问目标模型的内部架构和参数，仅使用目标模型的输出来指导生成器的训练。

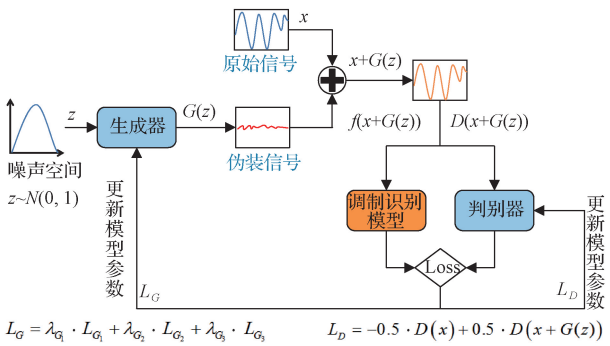


图 3 基于 AdvGAN 的伪装信号生成网络结构

Fig. 3 The structure of AdvGAN-based camouflage signal generation network of transmission system

基于 AdvGAN 的伪装信号生成方法的工作流程如算法 1 所示。通过对抗性训练，生成器从随机噪声空间 z 映射到伪装信号空间，产生伪装信号 $\eta(t)=G(z)$，不采用图像领域的 AdvGAN 中将原始样本作为输入的方法；判别器学习区分干净原始样本 $x(t)$ 和叠加伪装后的样本 $y(t)=x(t)+\eta(t)$。

算法 1 基于 AdvGAN 的伪装信号生成方法

组成部分：生成器 G 和判别器 D；调制识别模型 f

输入：原始信号 $x(t)$；真实类别标签 l；迭代次数 N

输出：

$$\min \| y(t)-x(t) \|_{\infty} \leqslant \varepsilon \tag{1}$$

$$f(y(t)) \neq f(x(t)) \tag{2}$$

输出为满足式(1)和(2)的伪装信号，其中 ε 为最大不可被察觉的伪装信号大小

生成器 G

输入：服从高斯分布的随机噪声样本 z

输出：伪装信号 $\eta(t)=G(z)$

判别器 D

输入：原始信号 $x(t)$，伪装后的信号 $y(t)=x(t)+\eta(t)$

输出：1 代表真，0 代表假

1. 初始化网络
2. for epoch = 0 to $N-1$ do
3. 固定生成器，根据判别器的损失函数 L_D 训练判别器
4. 固定判别器，根据生成器的损失函数 L_G 训练生成器
5. for 每个原始信号 $x(t)$ do
6. 计算伪装信号 $\eta(t)=G(z)$
7. 叠加伪装后的信号 $y(t)=x(t)+\eta(t)$
8. 输入智能调制识别模型，对 $f(y(t))$ 进行决策
9. end for
10. end for
11. return $\eta(t)$，$y(t)$

2 模型设计

2.1 调制识别模型设计

对于调制信号识别任务，一个合适的目标模型对于反侦察效果的体现至关重要。本文采用 CNN 和 VGG16 两种网络对信号序列进行特征提取和识别分类，并对网络结构进行了优化和调整，以匹配数据集。CNN 和 VGG16 的网络结构如表 1、2 所示。

表 1 CNN 的网络架构

Table 1 The network architecture of the CNN

Layers	Output Shape
Conv2D+ ReLU+ BatchNorm2D+Dropout	(batch_size,256,2,128)
Conv2D+ ReLU+ Dropout+BatchNorm2D	(batch_size,128,2,128)
Conv2D+ ReLU+ Dropout+BatchNorm2D	(batch_size,80,1,128)
Flatten	(batch_size,10240)
Linear + ReLU	(batch_size,256)
Linear + ReLU+ Dropout	(batch_size,128)
Linear +ReLU	(batch_size,64)
Linear	(batch_size,11)

表 2 VGG16 的网络架构

Table 2 The network architecture of the VGG16

Layers	Output Shape
Conv2D+BatchNorm2D+ ReLU	(batch_size,64,2,128)

Conv2D+BatchNorm2D+ReLU+Maxpooling2D	(batch_size,64,1,64)
Conv2D+BatchNorm2D+ReLU	(batch_size,128,1,64)
Conv2D+BatchNorm2D+ReLU+Maxpooling2D	(batch_size,128,1,33)
Conv2D+BatchNorm2D+ReLU	(batch_size,256,3,35)
Conv2D+BatchNorm2D+ReLU	(batch_size,256,3,35)
Conv2D+BatchNorm2D+ReLU+Maxpooling2D	(batch_size,256,1,17)
Conv2D+BatchNorm2D+ReLU	(batch_size,512,3,19)
Conv2D+BatchNorm2D+ReLU	(batch_size,512,3,19)
Conv2D+BatchNorm2D+ReLU+Maxpooling2D	(batch_size,512,1,9)
Conv2D+BatchNorm2D+ReLU	(batch_size,512,1,9)
Conv2D+BatchNorm2D+ReLU	(batch_size,512,3,11)
Conv2D+BatchNorm2D+ReLU+Maxpooling2D	(batch_size,512,1,5)
Flatten	(batch_size,2 560)
Linear+ ReLU+ Dropout	(batch_size,4 096)
Linear + ReLU+ Dropout	(batch_size,4 096)
Linear	(batch_size,11)

2.2 生成器设计

对于生成器而言，输入为随机噪声分布，输出得到伪装信号，其形状与原始信号保持一致。生成器的损失函数由 L_{G_1}、L_{G_2} 和 L_{G_3} 组成。L_{G_1} 表示最小化原始信号 $x(t)$ 和叠加伪装后的信号 $y(t)$ 的差异，即生成器希望叠加伪装后的信号能使得判别器信以为真；L_{G_2} 表示扰动损失，计算伪装信号在 L2 范数下的平均值，用于限制伪装信号的大小。

$$L_{G_1} = -D(x+G(z)) \tag{3}$$

$$L_{G_2} = \| G(z) \|_2 \tag{4}$$

式中：G 和 D 分别为生成器和判别器；x 为原始信号；z 为随机噪声；$\| \cdot \|_2$ 为 L2 范数。L_{G_3} 表示对抗性损失，使叠加伪装后的信号能够欺骗调制识别模型，借鉴 CW 攻击的损失函数对 L_{G_3} 进行了详细设计，如图 4 所示。这一过程有助于优化生成伪装信号的质量。

最后，设置 3 个权重系数 λ_{G_1}、λ_{G_2} 和 λ_{G_3} 控制不同损失项在训练中的相对重要性，综上给出生成器的损失函数

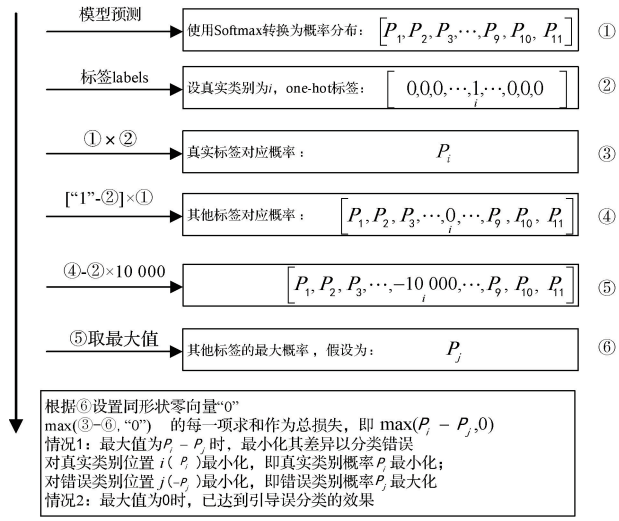


图 4 对抗性损失设计

Fig. 4 The design of adversarial loss

如下：

$$L_G = \lambda_{G_1} \cdot L_{G_1} + \lambda_{G_2} \cdot L_{G_2} + \lambda_{G_3} \cdot L_{G_3} \tag{5}$$

生成器模型如表 3 所示。

表 3 生成器的网络架构

Table 3 The network architecture of the generator

Layers	Output Shape
Flatten	(batch_size,1,2,128)
Conv2D+InstanceNorm2D+ReLU+ Dropout	(batch_size,64,2,128)
Conv2D+InstanceNorm2D+ReLU+ Dropout	(batch_size,128,2,128)
Conv2D+InstanceNorm2D+ReLU+ Dropout	(batch_size,256,2,128)
ConvTranspose2D+InstanceNorm2D+ ReLU	(batch_size,128,2,128)
ConvTranspose2D+InstanceNorm2D+ ReLU	(batch_size,64,2,128)
ConvTranspose2D+Tanh	(batch_size,1,2,128)

2.3 判别器设计

对判别器而言，它的输入为生成器产生的伪装信号加上原始信号，输出为 1 或 0，分别代表真和假。考虑到判别器的目标是区分叠加伪装后的信号和原始信号，核心任务是进行二分类判断，根据判别器对两种信号的判断结果，以此调整生成器的参数，构造更逼真的样本，同时也在不断提高自身判别真假的能力。针对判别器的损失函数设计，公式如下：

$$L_D = -0.5 \cdot D(x) + 0.5 \cdot D(x+G(z)) \tag{6}$$

式(6)确保负责判别真实数据和生成数据的两项在损

失函数中得到平等的对待，以增强判别器区分二者的能力。判别器模型如表4所示。

表4 判别器的网络架构

Table 4 The network architecture of the discriminator

Layers	Output Shape
Flatten	(batch_size,256)
Linear + BatchNorm1D+ Dropout+LeakyReLU	(batch_size,128)
Linear + BatchNorm1D+ Dropout+LeakyReLU	(batch_size,64)
Linear + BatchNorm1D+ Dropout+LeakyReLU	(batch_size,32)
Linear +Sigmoid	(batch_size,1)

3 实验设置和结果分析

本文使用Pytorch框架进行网络模型的训练和测试，操作系统为Windows10，GPU为RTX 3080 Ti。CNN、VGG16学习率设置为0.001，损失函数使用交叉熵损失函数(cross-entropy loss，CE)；伪装信号生成模型的生成器和判别器学习率分别为0.000 35和0.000 55，损失函数使用均方误差损失函数(mean squared error，MSE)。所有模型均使用Adam优化器。

3.1 数据集

本文使用无线电机器学习调制信号数据集RADIOML2016.10A进行实验，这个公开数据集专门用于机器学习在调制信号识别方面的研究和评估。该数据集包含了220 000个调制类型信号样本，覆盖了20种信噪比条件(−20～18 dB，步长2 dB)，包括11种调制方式，其中8种数字调制方式(BPSK、QPSK、8PSK、GFSK、CPFSK、PAM4、QAM16、QAM64)和3种模拟调制方式(AM-SSB、AM-DSB、WBFM)，然后在这些信号样本上添加加性高斯白噪声(AWGN)来模拟真实通信环境中的噪声干扰，信道特性包括中心频率偏移、采样率偏移、多径传播和衰落效应，在调制和信道建模之后，信号被归一化并封装成220 000个长度为128的同相和正交分量样本，每个样本维度为(128,2)。

本文对数据集样本进行预处理操作并随机划分，将80%的数据作为训练集，剩余20%作为测试集。

3.2 评估指标

对于调制信号识别任务，测试在不同信噪比下的识别准确率，作为评估模型的主要指标。其中，信噪比是衡量信号强度的重要指标，它被定义为信号功率与背景噪声功率之比：

$$SNR = 10\log\frac{P_{\text{signal}}}{P_{\text{noise}}} \tag{7}$$

在进行反侦察任务的评估时，考虑了多个关键指标。除了识别准确率下降程度和时域波形失真程度外，使用信伪比(signal-to-camouflage ratio，SCR)来衡量叠加的伪装信号与原始信号之间的量级关系，其计算方式如下：

$$SCR = 10\log\frac{P_{\text{signal}}}{P_{\text{camouflage}}} \tag{8}$$

更大的SCR值说明伪装信号对原始信号的影响越微弱，对通信性能的影响较小。

3.3 实验结果

1)调制识别效果和性能分析

考虑到数据集中包含20种不同信噪比的调制信号数据。首先，将所有信噪比的数据合并成一个数据集进行训练。然后，在测试阶段，针对不同信噪比水平进行独立验证和测试，评估模型的识别准确率。CNN和VGG16这两种智能调制识别模型随信噪比变化的识别准确率，以及在信噪比条件为0和18 dB时的混淆矩阵如图5所示。

其中CNN模型对QAM16和WBFM两种调制类型信号的识别准确率较低，VGG16模型对WBFM调制类型信号的识别准确率较低。除此之外，CNN和VGG16两种调制识别模型均取得了较高的识别准确率。

2)反侦察效果和性能分析

(1)白盒测试

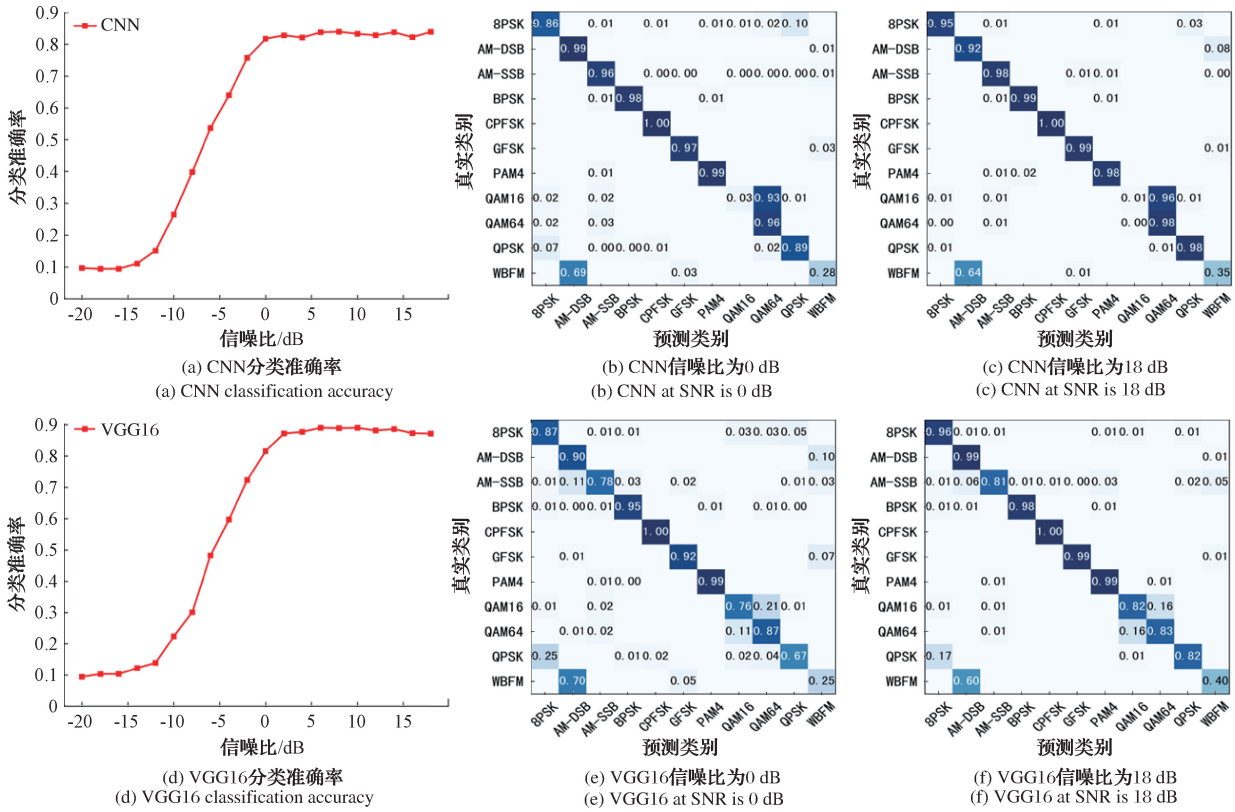
实验选取CNN作为白盒模型，设置扰动因子$\varepsilon = 0.0015$，比较了3种伪装信号生成方法，FGSM，PGD和AdvGAN。白盒情况下，CNN对3种方法构造的伪装信号的识别准确率如图6所示。

为了能够进行有效对比，首先，根据伪装前后的识别准确率来评估反侦察效果；其次，从调制信号波形和SCR指标两个角度衡量伪装信号的大小。

由图6可见，针对不同的信噪比条件，3种方法的反侦察效果表现如下：信噪比为−20～−12 dB时，FGSM反侦察效果略优于PGD和AdvGAN；信噪比为−10～−2 dB时，由于PGD的迭代性质，取得了较好的反侦察效果；当信噪比大于0 dB后，经AdvGAN进行信号伪装后，CNN智能调制识别模型的准确率仅为20%左右，反侦察效果明显优于FGSM和PGD。基于AdvGAN的方法最大的优点在于，伪装信号生成网络一旦训练完毕，就可以稳定的生成伪装信号，实现信号伪装的同时，便于实际部署。

仅从对侦收方智能调制分类器造成的影响这一角度比较不同方法存在一定的局限性，为说明伪装信号对通信信号造成的影响，仿真分析信噪比10 dB条件下，PAM4和8PSK两种调制信号分别采用FGSM、PGD以及AdvGAN 3种伪装方法后信号波形的失真情况，并与原始信号样本进行可视化对比。图7(a)～(c)为PAM4信号伪装前后对比，图7(d)～(f)为8PSK信号伪装前后对比。

由图7可见，3种方法对信号的失真均较小，计算采用FGSM、PGD以及AdvGAN 3种方法生成伪装信号的



(a) CNN分类准确率
(a) CNN classification accuracy

(b) CNN信噪比为0 dB
(b) CNN at SNR is 0 dB

(c) CNN信噪比为18 dB
(c) CNN at SNR is 18 dB

(d) VGG16分类准确率
(d) VGG16 classification accuracy

(e) VGG16信噪比为0 dB
(e) VGG16 at SNR is 0 dB

(f) VGG16信噪比为18 dB
(f) VGG16 at SNR is 18 dB

图 5　不同网络的准确率和混淆矩阵

Fig. 5　The accuracy and confusion matrix results of different networks

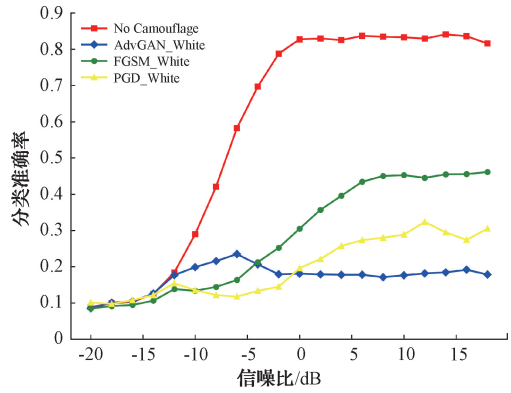


图 6　白盒反侦察效果

Fig. 6　White box anti-reconnaissance effect

信伪比并综合图 6 的实验结果，可得到 3 种方法的反侦察性能对比情况如表 5 所示。

由表 5 可见，一方面，所提基于 AdvGAN 的伪装信号生成方法造成目标模型识别准确率恶化的程度比 FGSM 和 PGD 分别高 27%和 11%左右，有效地反制了侦收方的智能识别模型。另一方面，从信号波形失真的角度来看，通过 SCR 指标评估，可知 AdvGAN 生成的伪装信号对通信信号的损伤较小，对通信性能的影响更为有限，一定程度上保证了信号的完整性，利于合作接收方开展后续工作。

综上所述，基于 AdvGAN 的伪装信号生成方法表现出低失真高干扰的特性，即使加入的伪装信号的量级不如 FGSM 和 PGD(在相对较小的伪装条件下)，AdvGAN 却能够在通信信号失真最小的同时，有效地恶化侦收方智能调制识别模型的准确率。

(2)黑盒测试

在实际情况下，对于通信方而言，侦收方的智能调制识别模型的结构往往是未知的。因此，为进一步验证算法的实用性，仿真分析了针对侦收方智能识别模型未知情况下，所提方法的性能，即黑盒模型条件下的性能。

实验选择 CNN 作为本地模型构造伪装信号，VGG16 作为黑盒模型进行测试，3 种伪装信号生成方法的黑盒反侦察效果如图 8 所示。

由图 8 可见，3 种方法基于本地模型 CNN 上构造的伪装信号，仍能够使黑盒模型 VGG16 的识别准确率大幅下降。相较于白盒模型，基于 AdvGAN 和 FGSM 的伪装性能仅下降了 5%，而基于 PGD 的伪装性能下降了约 15%，这是由于 PGD 方法的迭代特性要求连续访问目标模型的信息，然而，在黑盒情况下，模型的具体结构信息是无法获取的。

由此可见，对于侦收方调制识别模型结构未知的黑盒情况，所提方法仍具有很好的反侦察性能。

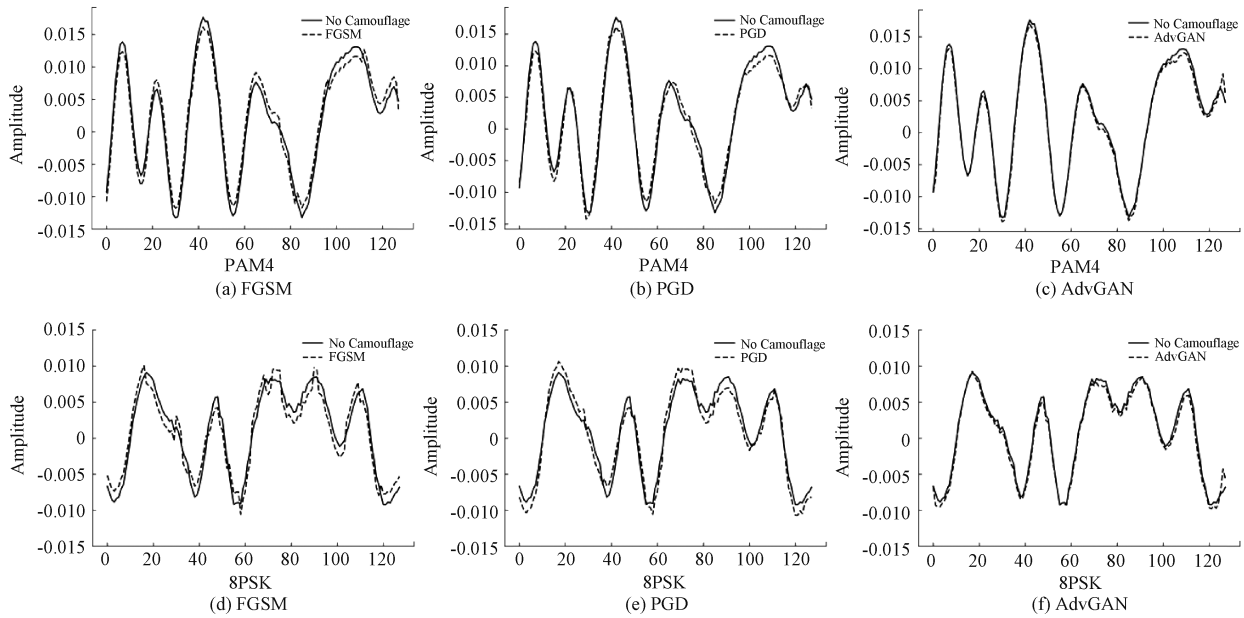


图 7　信噪比为 10 dB 时 PAM4 和 8PSK 信号波形

Fig. 7　PAM4 and 8PSK signal waveforms at SNR is 10 dB

表 5　信噪比为 10 dB 时反侦察性能对比

Table 5　Comparison of counter-surveillance performance at SNR is 10 dB

	准确率/%	调制信号	SCR/dB
FGSM	45	PAM4	15.25
		8PSK	11.92
PGD	29	PAM4	16.81
		8PSK	13.19
AdvGAN	18	PAM4	20.49
		8PSK	18.78

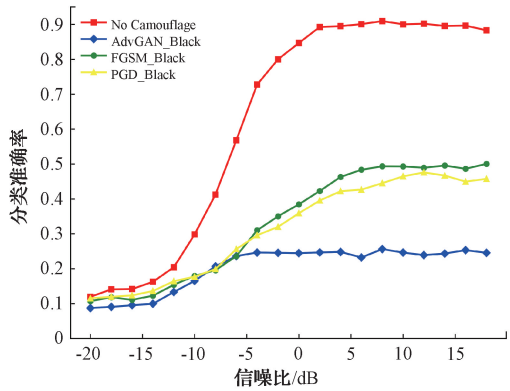


图 8　黑盒反侦察效果

Fig. 8　black box anti-reconnaissance effect

4　结　论

本文提出的基于 AdvGAN 的无线通信信号反侦察方法能够实现很好的反侦察效果。首先，本文方法仅需要使用目标模型的输出来指导生成器的训练，并不需要直接访问模型的架构和参数，便于实际部署，在面对已知模型和未知模型时均能以很高的概率使其识别错误。其次，本文方法对通信信号所造成的影响十分微小，不会带来通信性能的大幅下降，有利于合作接收方开展后续工作。但是本文方法也存在着一些有待改进之处，如计算复杂度较高、并未考虑侦收方在采取防御措施的基础上进行仿真实验。

本文重点分析了信号伪装后对侦收方的影响，取得了一定的反侦察效果，但并未对合作接收方的影响进行深入研究。未来工作将深入研究伪装信号对通信性能的恶化程度并研究合作接收方去除伪装信号的技术途径。

参 考 文 献

[1]　唐震，乔晓强，张涛，等. 基于深度残差收缩网络的辐射源个体识别方法[J]. 电子测量技术，2022，45(9)：168-174.
TANG ZH，QIAO X Q，ZHANG T，et al. Individual radiator identification method based on deep residual shrinkage network [J]. Electronic Measurement Technology，2022，45(9)：168-174.

[2]　王锦卫，杜奕航，张江，等. 基于深度神经网络和随机森林集成模型的 ADS-B 辐射源个体识别[J]. 国外电子测量技术，2023，42(3)：1-7.
WANG J W，DU Y H，ZHANG J，et al. Individual recognition of ADS-B emitter based on deep neuralnetwork and random forest ensemble model[J]. Foreign Electronic Measurement Technology，2023，42(3)：1-7.

[3] O'SHEA T J, CORGAN J, CLANCY T C. Convolutionalradiomodulation recognition networks[C]. Engineering Applications of Neural Networks: 17th International Conference. Springer International Publishing,2016: 213-226.

[4] O'SHEA T J, WEST N. Radio machine learning dataset generation with GNU radio[C]. Proceedings of the GNU Radio Conference, 2016.

[5] 王超,魏祥麟,田青,等.基于特征梯度的调制识别深度网络对抗攻击方法[J].计算机科学,2021,48(7):25-32.
WANG CH, WEI X L, TIAN Q, et al. Feature gradient-based adversarial attack on modulation recognition-oriented deep neural networks [J]. Computer Science,2021,48(7):25-32.

[6] 杨小蒙,张涛,庄建军,等.基于多模态融合和深度学习的调制信号识别[J].计算机科学,2023,50(S2):705-711.
YANG X M, ZHANG T, ZHUANG J J, et al. Modulation signal recognition based on multimodal fusion and deep learning [J]. Computer Science, 2023,50(S2):705-711.

[7] DOAN V S, HUYNH-THE T, HUA C H, et al. Learningconstellation map with deep CNN for accurate modulation recognition [C]. GLOBECOM 2020 — 2020 IEEE Global Communications Conference. IEEE, 2020:1-6.

[8] LIN Y, TU Y, DOU Z, et al. Contour stella imageand deep learning for signal recognitionin the physical layer[J]. IEEE Transactions on Cognitive Communications and Networking, 2020, 7(1): 34-46.

[9] 夏吉业,张海勇,尚教凯.海上编队通信反侦察策略研究[J].舰船电子工程,2020,40(1):9-12,53.
XIA J Y, ZHANG H Y, SHANG J K, et al. Threat analysis and strategy research on marine formation anti-reconnaissance communication [J]. Ship Electronic Engineering, 2020,40(1):9-12,53.

[10] SZEGEDY C, ZAREMBA W, SUTSKEVER I, et al. Intriguing properties of neural networks[J]. arXiv preprint arXiv:1312.6199, 2013.

[11] GOODFELLOW I J, SHLENS J, SZEGEDY C. Explaining and harnessing adversarial examples[J]. arXiv preprint arXiv:1412.6572, 2014.

[12] KURAKIN A, GOODFELLOW I J, BENGIO S. Adversarial examples in the physical world [C]. International Conference on Learning Representations, 2017.

[13] MADRY A, MAKELOV A, SCHMIDT L, et al. Towards deep learning models resistant to adversarial attacks[J]. arXiv preprint arXiv:1706.06083, 2017.

[14] SADEGHI M, LARSSON E G. Adversarial attacks on deep-learning based radio signal classification[J]. IEEE Wireless Communications Letters, 2018, 8(1): 213-216.

[15] LIN Y, ZHAO H, TU Y, et al. Threats of adversarial attacks in DNN-based modulation recognition[C]. IEEE INFOCOM 2020 — IEEE Conference on Computer Communications. IEEE, 2020: 2469-2478.

[16] HAMEED M Z, GYORGY A, GUNDUZ D. The best defenseis a good offense: Adversarial attacks to avoidmodulation detection[J]. IEEE Transactions on Information Forensics and Security, 2020, 16: 1074-1087.

[17] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets[J]. Advances in Neural Information Processing Systems, 2014, DOI: 10.3156/JSOFT.29.5_177_2.

[18] XIAO C, LI B, ZHU J Y, et al. Generating adversarial examples with adversarial networks[J]. arXiv preprintarXiv:1801.02610, 2018.

作者简介

陈奕功,硕士研究生,主要研究方向为调制信号识别、对抗攻击、深度学习。
E-mail:chenyigong2002@163.com

张江,博士,主要研究方向为智能频谱感知、频谱管理、信号处理等。
E-mail:god2525775@163.com

乔晓强(通信作者),研究员,主要研究方向为电磁频谱智能感知、电磁频谱安全与控制等。
E-mail:qxq0527@163.com

龙伟军,教授,博士生导师,研究员级高级工程师,主要研究方向为智能感知与信息融合、新体制雷达、雷达通信一体化、信号处理等。
E-mail:chinacohit@163.com