

# 基于 GAN 的红外小目标检测合成 数据增强技术研究

杜 锋

(江苏电子信息职业学院, 江苏 淮安 223003)

**摘要:** 近年来, 卷积神经网络在红外小目标检测方面取得了良好的效果; 但由于该技术中公共训练数据数量有限, 在一定程度上导致分析结果不够精确; 为了解决训练数据的稀缺问题, 提出了一种生成红外小目标检测合成训练数据的方法; 采用生成对抗网络框架, 其中合成背景图像和红外小目标在两个独立的过程中生成; 在第一阶段, 通过将可见光图像转化为红外图像来合成红外图像; 在第二阶段, 将转换后的图像上植入目标掩码, 将所提出的强度调制网络合成了真实的目标对象, 可以从进一步的图像处理产生; 实验结果表明, 当使用由真实图像和合成图像组成的数据集训练各种检测网络时, 检测网络比只使用真实数据产生更好的性能, 优化了传统数据处理方法, 使得训练结果更为精确。

**关键词:** 卷积神经网络; 生成对抗网络; 图像转换; 红外小目标; 合成数据增强

## Research on GAN-Based Synthetic Data Augmentation Technology for Infrared Small Target Detection

DU Feng

(Jiangsu Vocational College of Electronics and Information, Huai'an 223003, China)

**Abstract:** Recently, convolutional neural networks have made a great progress in infrared small target detection. However, due to limited public training data in this technology, the analysis results are inaccurate a certain extent. To address the scarcity of training data, this paper proposes a method for generating synthetic training data for infrared small target detection, and adopts a generative adversarial network framework, where the synthesized background image and infrared small targets are generated in two independent processes. In the first stage, the infrared images are synthesized by converting visible light images into infrared images. In the second stage, the target mask is implanted on the converted image, and the proposed intensity modulation network is synthesized into the real target object, which can be generated from further image processing. Experimental results show that, when the dataset on real and synthetic images trains various detection networks, the detection network of the real and synthetic images has better performance than that of only real images, optimizing the traditional data processing methods and making the training results more accurate.

**Keywords:** convolutional neural network; generative adversarial network; image translation; infrared small target; synthetic data augmentation

### 0 引言

检测红外图像中的小目标是红外搜索和跟踪 (IRST) 系统中的一项重要任务。它在军事领域有许多应用, 如红外瞄准制导、预警、防空和反导弹<sup>[1]</sup>。红外小目标检测的主要挑战在于缺乏足够的结构特征, 因为目标在图像中通常由少量的像素表示。此外, 背景杂波可能会导致严重的误报, 使目标识别更加困难。

传统的滤波方法通过人工设计的滤波器来解决这个问题, 在目标区域产生高响应, 因为目标的外观通常类似于红外图像中的小亮点。通常采用高斯滤波器的拉普拉斯式<sup>[2]</sup>来利用目标的这一空间特性。另一方面, 二维最小二乘滤波器<sup>[3]</sup>通过使用周围的像素来估计当前区域的背景信

号。然后, 利用输入背景和预测背景之间的差值来检测目标。然而, 这种针对目标和背景<sup>[4]</sup>的简单模型不仅难以适应各种场景, 而且还可能产生许多误报。受人类视觉系统的启发, 提出了一种局部对比测量 (LCM)<sup>[5]</sup>来增强目标信号和抑制背景信号。文献 [6] 对 LCM 进行了扩展, 以提高了计算效率和性能。为了有效地利用目标和背景的结构信息, 文献 [7] 还提出了一种基于多方向滤波器的检测方法。最近, 一种新的形态学表示法<sup>[8]</sup>, 如不同形状的树, 已经被引入, 通过补充空间信息来弥补红外图像中详细信息的不足。

红外贴片图像 (IPI) 模型及其变体<sup>[9]</sup>提供了传统滤波方法的另一种替代方案。IPI 模型将红外小目标检测作为一

收稿日期: 2024-04-30; 修回日期: 2024-05-23。

基金项目: 2023 年江苏省产学研合作项目 (BY20231025)。

作者简介: 杜 锋 (1977-), 男, 硕士, 副教授。

引用格式: 杜 锋. 基于 GAN 的红外小目标检测合成数据增强技术研究[J]. 计算机测量与控制, 2024, 32(11): 118-124.

个低秩(背景)和稀疏信号(目标)分解问题。虽然这些方法在一些图像中显示出了良好的结果,但由于无法捕获高级语义,它们仍然难以区分复杂和非重复的背景区域中小的和模糊的目标。深度卷积神经网络(CNNs)的快速进展促使了提出的学习端到端模型的问题<sup>[10]</sup>,该模型将输入的红外图像转换为检测地图。这些基于学习的方法比传统的滤波或低秩近似方法产生更好的检测结果。

一般来说,基于 CNN 的方法的性能随着训练样本数量的增加而提高。然而,与一般的目标检测和分类相比,用于红外小目标检测的现有公共数据集<sup>[11-12]</sup>的规模非常有限。收集和手动注释红外小目标检测的大数据集需要相当长的时间、精力和成本,原因主要有以下几个方面:

1) IRST 系统安装在平台上,如飞机或船舶等,然后在操作过程中获取红外图像。为了从存储的图像中构建数据集,首先,应该从大量的图像集中选择包含感兴趣的目标的图像。其次,应该对原始数据进行预处理为作为 CNN 模型的输入(例如,去除不好的像素,因为它可能会与真正的目标相混淆)。最后,能够区分目标和背景杂波的专家应该生成地面真实标签。这些过程需要花费大量的时间和精力。

2) 应准备一个不同的训练数据,以提高训练模型的通用性。为此目的,配备了 IRST 系统的平台应该会在不同的环境中收集大量的图像,这就导致了成本的增加。例如,即使对于同一场景,所获得的图像的外观也可能会严重不同,因为孔径辐射不仅取决于背景表面,而且还取决于大气条件。

使用合成数据进行训练是一种有效的方法,因为可以显著减轻构建真实数据集的负担。一个合成图像生成器模拟图像形成的物理过程,并使用建模的成像管道创建人工图像。通过调整控制生成过程的参数,可以获得大量多样化的训练图像。此外,合成图像生成器可以在此过程中自动产生注释。合成数据集已广泛应用于 CNN 模型的各种应用,如合成孔径雷达数据<sup>[13]</sup>中的自动目标识别、高光谱图像<sup>[14]</sup>中的目标检测、航空图像<sup>[15]</sup>中的飞机检测。

另一种综合训练数据的方法是利用生成式对抗网络(GAN)<sup>[16]</sup>。根据生成的样本来源,基于 GAN 的数据合成可分为两类:样本合成和样本迁移<sup>[17]</sup>。在样本合成中,生成器试图输出虚假但可信的数据,而鉴别器被训练来区分真实和虚假的数据。训练结束后,生成器产生的数据尽可能地符合真实数据分布的假数据。第二类方法,也称为图像数据的图像到图像转换<sup>[18]</sup>,通过不仅保留内在源内容,而且转移外部目标样式,学习将数据从源域转换到目标。与基于模拟的数据生成相比,基于 GAN 的方法的优点是,它们可以不建模成像管道或物理世界而获得合成数据<sup>[19]</sup>。

## 1 红外小目标检测

### 1.1 基于深度学习的红外小目标检测

与传统的过滤方法相反,数据驱动的方法,即基于深度学习的方法,不假设任何关于目标和背景的先验知识,

并从数据中学习它。文献[17]采用条件对抗性学习策略对目标分割网络进行训练。在他们的工作中,有两个生成器可以产生分割结果。一个负责最小化错误警报,另一个负责减少误检测。这两个生成器都使用了扩展的卷积来收集上下文线索。一个鉴别器可以区分分割结果与地面真相。最终结果是通过平均两个分割输出得到。由于条件对抗网络对图像转换问题的成功,文献[20]提出了一种训练将红外图像转换为目标地图的检测网络的方法。基于 U-Net<sup>[21]</sup>的生成器和鉴别器通过成对训练数据的对抗性学习进行训练。由于在实践中很难获得真实的成对训练数据,他们提出了一个合成成对训练数据的过程。

最近,文献[22]基于特征金字塔网络(FPN),提出了注意局部对比度(ALC)网络来从红外图像中分割目标区域。在 ALC 网络中,利用扩张卷积和特征图的循环位移,在特征水平上进行多尺度 LCM。它们还合并了一个新的注意模块,以合并不同层的特征。在另一项工作文献[23]中,引入了一个非对称上下文调制(ACM)模块,并集成到 U-Net 和 FPN 中。文献[24]提出了一种基于 UNet+ 的密集嵌套检测网络(DNA-Net)

### 1.2 图像到图像的转换

图像到图像的转换是在给定足够的训练数据时,将场景的一种可能表示转换为另一种可能表示的任务<sup>[25]</sup>。自从作为一种通用解决方案引入以来,它取得了巨大的成功,并应用于许多视觉任务。在 Pix2pix<sup>[26]</sup>中,转换网络是在条件 GAN 的框架内进行训练的,其中输入图像是一个条件变量。该网络由 L1 损失引导,迫使平移后的图像靠近目标域中相应的图像。尽管 Pix2pix 取得了巨大的成功,但其训练方法还是带来了一些问题。第一个是模式崩溃,其中转换网络不能产生不同的输出。

另一个问题是需要成对的源-目标图像,因为人们很少能获得足够数量的成对训练样本。为了使转换网络能够生成多模态输出,BicycleGAN<sup>[27]</sup>引入了潜在代码(即转换网络的输入噪声向量)和输出之间的双射一致性。在 CycleGAN<sup>[28]</sup>中,配对训练数据集的约束通过一个循环一致性来放松,将图像从源域转换到目标域,然后再回到源域,这应该是标识函数。

### 1.3 模态转移

现有的一种模态数据样本可以转移到另一种模态中,以补充不足的训练数据集<sup>[29]</sup>。由于在可见光域存在一些大型数据集可用于视觉任务,可以利用这些数据集来弥补红外域数据的稀缺性。在文献[30]中,可见光光谱中标记的视频序列被转换成红外视频序列。与原始领域的标签一起,转换后的视频被用来训练一个热红外跟踪器。在该工作中, Pix2pix<sup>[31]</sup>和 CycleGAN<sup>[32]</sup>被用于转换。与上述方法一样,采用 Pix2pixHD 进行转换用语义标签标注的可见图像标注到热图像中,用于热图像语义分割。

根据数据采集的时间和地点,遥感数据的特征可能有很大差异。因此,使用在特定时期或特定区域收集的数据

集训练的模型在测试时可能表现出较差的性能。模态转移也可以用来减少训练数据集和测试数据集之间的域差距。在文献 [33] 中, 训练一个由一个尺度和一个移位矩阵组成的简单生成器, 将训练图像的光谱分布转换为测试图像的光谱分布。然后, 利用变换后的训练图像对语义分割网络进行微调。在文献 [34] 中, 采用了一种两阶段的方法来提高分割网络的泛化能力。首先, 数据增强网络使用样式传输技术来学习多个域之间的映射。然后, 从增强网络生成的多样化样本中训练分割网络。

当转换包含许多不同对象实例的复杂图像时, 源图像中的对象可能会被错误地映射到目标域中 (即, 转换后的对象的数据分布与目标域中的对象的数据分布不一致)。为了解决图像级转换的问题, 提出了几种实例级转换的方法。这些方法训练它们的转换网络, 使网络对背景和每个对象实例应用自适应映射。在这项工作中, 提出了一种新的合成图像生成方法, 将图像和感兴趣的物体的转换顺序统一起来, 从而允许利用真实的合成图像进行红外小目标检测。所提出的数据增强方案可以缓解该领域数据集不足的限制。

## 2 合成训练数据生成方法

在本工作中, 介绍了一种新的方法来创建合成训练数据的红外小目标检测。在这里, 可以在图像到图像转换的背景下处理这个问题。与以前在单个转换过程中同时处理背景实例和对象实例的方法不同, 采用了两阶段的方法。这一原因如下:

1) 在红外小目标检测中, 感兴趣的标记位于较远的距离。在这种情况下, 由于可见光谱中的相对较低的大气传输, 在可见图像中很可能看不到可见红外图像中的可观测目标。因此, 很难获得两个光谱域内的配对目标数据。

2) 与可见物体相反, 红外区域小目标的外观没有明显的特征。此外, 背景也不能很好地区分目标的空间分布。因此, 当分别处理将背景和标记从一个域转换到另一个域的问题时, 转换过程可以更容易处理。

在该方法的第一阶段, 应用图像-图像转换方法将可见光图像转移到红外域, 以生成合成的背景图像。然后, 为这些图像生成目标掩模。为了合成真实的目标, 提出了一个在 GAN 框架内训练的强度调制网络。强度调制网络调整掩模的强度以适应周围环境, 从而欺骗鉴别器。另一方面, 训练鉴别器来区分真实目标和强度调制目标掩模。值得注意的是, 文献 [2] 中的研究方法与本站的工作类似, 提出了一个将可见图像转换为红外图像的框架。然而, 他们的转换网络是使用在同一场景中获得的图像数据集进行训练和测试的。此外, 在他们的实验中使用的数据集处理扩展的目标。本工作的主要贡献如下:

1) 将生成合成训练数据的问题分解为两个子问题。a) 利用图像到图像的转换技术, 从真实的可见光谱图像中生成合成的背景图像。b) 基于所提出的强度调制网络, 采用 GAN 框架在合成的背景图像上渲染真实的目标。

2) 比较了 5 个深度网络, 包括基线模型和 4 个最先进的模型, 以显示所提出的数据增强方法的有效性。

3) 在消融研究中, 实验结果表明, 所提出的方法的每个阶段都可以逐步有助于提高检测性能。

从网络上收集到的可见图像被转换为红外域。本研究采用能够从一个源图像中生成多个目标图像的 BicycleGAN 进行背景平移。为了在合成的背景图像中渲染目标, 根据真实数据的地面真实值建立了各种目标掩模。首先, 随机移动目标掩模的位置, 然后, 通过应用膨胀、侵蚀和恒等式之间的其中一种操作, 任意修改目标掩模的形状。将处理后的掩模植入合成的背景图像中。

所提出的强度调制网络将合成图像和掩模作为输入, 调制目标区域的强度值, 将植入的掩模转化为真实的目标。该鉴别器旨在通过只关注目标区域来区分真实目标和合成目标。为了简单起见, 省略了用作鉴别器附加输入的目标掩模。对抗性损失迫使强度调制网络学习从以固定像素值表示的植入目标掩模映射到具有空间的合成目标, 且分布类似于真正的目标。

## 3 合成图像生成过程

### 3.1 合成背景图像

如第一节引言所述, 用于红外小目标检测的公开数据集的大小是有限的。为了创建各种背景图像, 使用图像到图像转换的方法将可见图像转换为红外图像。在文献 [35] 中, 考虑了两种图像到图像的转换方法, 即 Pix2pix 和 CycleGAN, 并比较了从标记的真实彩色视频中生成标记的合成红外视频, 用于热红外跟踪问题。通过主观地和客观地检查转换图像的质量, 他们得出结论, 使用配对训练数据集的 Pix2pix 优于使用未配对训练数据集的 CycleGAN。基于这一结果, 利用 BicycleGAN, 它是 Pix2pix 的扩展, 具有产生多模态输出的能力。由于以下原因, 要求转换网络具有一个额外的特性, 即多模态性:

1) 多模态输出模拟了红外图像的多样性。对于同一场景, 捕获的红外图像的强度分布可以根据一天中的时间和环境条件而变化。在多模态合成背景图像上进行训练, 红外小目标检测网络对这种变化具有更强的鲁棒性。

2) 对于相同数量的可见图像, 可以增加由图像到图像转换模块生成的合成背景图像的数量。

表 1 几种图像转换方法的比较

方法	是否成对	是否多模态
Pix2pix	是	否
CycleGAN	否	否
BicycleGAN	是	是

总之, 选择 BicycleGAN 进行背景图像转换, 因为它是多模态的, 比使用未配对训练数据集的方法产生更好的结果。表 1 比较了所介绍的 3 种图像转换方法。

从网络上, 收集了天空、山、云、海岸和地平线的各

种图像。随机生成的潜在向量和收集到的图像被用作训练后的 BicycleGAN 的输入。转换后的红外图像的例子如图 1 和 2 所示。但并非所有的转换图像都与输入的可见图像一致，如图 2 所示。通常观察到可见光和平移红外图像之间的结构不匹配和失真，特别是在平滑区域。如图 3 可以看出，即使模式不同，两幅图像的轮廓应该是相似的。这个问题可以通过最近提出的多模态转换方法得到缓解。在这项工作中，当生成的红外图像的剪影与可见光图像有明显不同时，手动选择合适的转换图像用于第二阶段。



图 1 BicycleGAN 多模态输出成功案例

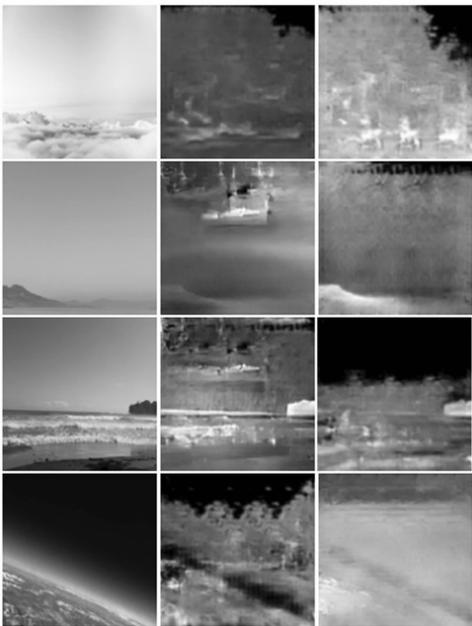


图 2 BicycleGAN 故障案例

### 3.2 强度调制网络

设  $M$  为二进制目标掩码，为该过程的输出，背景区域

的值为 0，否则为 1。给定平移后的背景图像  $I$ ，一个植入目标掩码的图像如下：

$$I_t = 255 \times M + I \otimes (1 - M) \quad (1)$$

其中： $\otimes$  表示元素级的乘法。在这里，假设处理一个 8 位的图像。对于所提出的强度调制网络  $G$ ，使用  $I_t$  作为输入。 $G$  预测了  $I_t$  的缩放图  $S$ ：

$$S = G(I_t) \quad (2)$$

通过仅将  $S$  应用于目标区域，可以得到最终的合成图像  $I_f$  如下：

$$I_f = S \otimes I_t \otimes M + I_t \otimes (1 - M) \quad (3)$$

为了鼓励  $\min_D L_{adv}(D) = \frac{1}{2} E_{I_r} [(D(I_r) - 1)^2] + \frac{1}{2} E_{I_f} [(D(I_f))^2]$  生成一个缩放图，将植入的掩模改变为真实的目标，使用了一个鉴别器  $\min_G L_{adv}(G) = \frac{1}{2} E_{I_r} [(D(I_f) - 1)^2]$  进行对抗性学习。即使对  $I_r$  没有可用的监督， $L_{cont}(G) = - \sum_{x \in \mathbb{R}_s} \ln(\frac{I_f(x)}{M_{sr}})$  也应该适应围绕目标掩模的背景区域的像素值。对抗性训练可以解决这些问题。为了训练  $M_{sr}$  和  $R_{tg}$ ，使用目标函数最小二乘 GAN (LS-GAN)<sup>[36]</sup>。目标函数如下：

$$\min_D L_{adv}(D) = \frac{1}{2} E_{I_r} [(D(I_r) - 1)^2] + \frac{1}{2} E_{I_f} [(D(I_f))^2] \quad (4)$$

$$\min_G L_{adv}(G) = \frac{1}{2} E_{I_r} [(D(I_f) - 1)^2] \quad (5)$$

其中： $I_r$  是真实的图像。在原始的 GAN 框架中，鉴别器基于从整个图像中提取的特征产生输出。然而，由于所提出的强度调制网络的设计仅用于调整掩蔽区域的像素值，因此鉴别器的输出应仅来自于目标区域的特征。与传统的全局 GAN 相反，PatchGAN<sup>[37]</sup> 中的鉴别器在输入图像上滑动一个窗口，并生成一个分数地图，指示每个补丁是真实的还是假的。因此，利用中心位于目标区域内的局部斑块的对抗性损失来训练强度调制网络。

包括一个额外的损失来训练强度调制网络。一般情况下，在图像中出现一个红外小目标作为一个光点，说明目标的强度值应远高于周围区域。为了惩罚目标和周围区域之间的对比度逆转，引入对比度损失，如下：

$$L_{cont}(G) = - \sum_{x \in \mathbb{R}_s} \ln(\frac{I_f(x)}{M_{sr}}) \quad (6)$$

其中， $M_{sr}$  为周围区域的平均值，计算目标区域  $R_{tg}$  的总和。那么，可以定义强度调制网络的总损失为：

$$L_{total}(G) = \lambda_{adv} L_{adv} + \lambda_{cont} L_{cont} \quad (7)$$

其中： $\lambda_{adv}$  和  $\lambda_{cont}$  是调整不同损失贡献的超参数。

### 3.3 目标检测的基线网络

为了展示所提出的合成数据增强方法的好处，在基线检测网络和最先进的基于深度学习的检测算法上进行了实验。基线检测网络是基于原始的 U-Net。首先，设计了 U-Net 用于生物医学图像的自动分割。由于 U-Net 在许多视觉相关问题上的



图 3 训练图像对的样本

优异性能，它已成为设计新的网络架构的首选起点。U-Net 由两个子部分组成：卷积编码器和带有跳过连接的解码器。基本的  $3 \times 3$  卷积操作之后是一个 BN 层和一个 ReLU 激活。该编码器通过  $2 \times 2$  个最大池化操作，以 2 个池化大小为  $2 \times 2$ ，步幅为 2，逐渐降低特征映射的空间维度。

这个过程在网络中重复了三次。在每个降采样步骤后，特征通道的数量会增加一倍。在编码器执行的操作相反发生在解码器。解码器使用  $2 \times 2$  转置卷积操作对特征映射进行上采样，然后将上采样的特征映射与来自编码器的相应特征映射连接起来。同时，连接的特征通道数量被减半。在最后一层，使用  $1 \times 1$  卷积和 Sigmoid 激活函数来产生最终的结果。

在编码器和解码器之间，插入空间注意 (SA) 模块，自动学习在特征图中的焦点位置。最初，SA 模块是卷积块注意模块的组成部分之一，并在通用图像分类和目标检测中显示了其有效性。利用特征之间的空间关系，SA 模块输出一个二维权重图。为了计算权重映射，沿通道轴对输入特征映射应用最大池化和平均池操作。然后，将两个池结果连接起来，由  $7 \times 7$  卷积算子进一步处理。最后，利用 Sigmoid 型激活函数将处理后的二维映射转换为权重映射。输出特征映射是通过将输入特征映射与计算出的权重映射进行元素级乘法得到的。

## 4 实验结果与分析

### 4.1 数据集和评估指标

如前所述，最终目标是提高在有限样本下训练的基于 CNN 的检测网络的性能。首先描述了红外小目标检测数据集，在其上评估提出的合成训练数据生成过程。如单帧红外小目标 (SIRST) 数据集从数百个红外序列中，选择了 427 个具有代表性的帧并手动标注。大约 90% 的图像帧包含单个目标。SIRST 数据集被分为 50% 用于训练，20% 用于验证，30% 用于测试。

为了训练 BicycleGAN，利用了两对可见红外数据集，KAIST 多光谱行人数据集和 CVC-14 数据集。这两个数据集的图像都是由安装在移动车辆上的一个红外线摄像机和一个可见光摄像机捕获的。KAIST 多光谱行人数据集包含约 95.3 k 张图像，CVC-14 数据集包含约 8.4 k 张图像。其中，只排除

了夜间序列，并且只从白天序列中采样训练图像。共 6 212 对可见红外图像用于 BicycleGAN 模型。

传统的滤波方法是通过诸如背景抑制因子和信号到杂波增益等指标来评估的。这些指标旨在评估滤波方法在减轻背景信号和改善 (或至少保留) 目标信号方面的成功程度。然而，深度网络输出一个二进制掩码，其中这些度量值的值在大多数情况下将是无穷大的。为了评估基于深度学习的方法的检测性能，使用了 4 种不同的评估指标。

第一个评价度量是联合上的归一化交集 (nIoU)，此外，在文献 [12] 和文献 [15] 中采用。nIoU 被定义为：

$$\text{nIoU} = \frac{1}{N} \sum_i \frac{TP[i]}{T[i] + P[i] - TP[i]} \quad (8)$$

其中： $TP$ 、 $T$  和  $P$  分别为真阳性、真阳性和正阳性。 $N$  为总样本数。为了获得更高的 nIoU 分数，检测结果应与地面真实值有最大的重叠，误报也最小。因此，nIoU 量化了目标定位的精度。

还采用了检测概率 ( $Pd$ ) 和误报 ( $Fa$ ) 来评估所提出的数据增强过程的有效性。在实际的红外小目标检测系统中，一组像素被报告为单一报警。通过计算检测地图中像素的簇数来计算  $Pd$  和  $Fa$ 。如果真实数据的中心和检测到的位置距离在阈值 (3 个像素) 内，则检测被认为是正确的预测，否则是假警报。为了获得检测图，将阈值固定为 0.5。

### 4.2 实验细节

为了从真实可见图像中合成背景红外图像，从头开始训练 BicycleGAN 网络。使用了与文献 [26] 相同的网络架构。使用 Adam 优化器训练转换网络，批量大小为 2，学习速率为  $2 \times 10^{-4}$ 。从网络上下下载的真实可见光图像通过经过训练的转换网络转换为合成的红外背景图像。在将真实图像输入转换网络之前，收集到的真实图像被调整到固定的分辨率并进行裁剪，就像在训练中所做的那样。为了构建目标掩码库，利用了 SIRST 数据集中的注释数据。具体来说，SIRST 数据集中的每个二进制目标映射都被移动，使目标的中心位于地图的中间。然后，将每个移位的注释数据注册到目标掩码库中。强度调制网络由 Adam 优化器进行训练，批处理大小为 2，学习速率为  $1 \times 10^{-4}$ 。设置了超参数  $\lambda_{adv} = 0.03$  和  $\lambda_{cont} = 0.03$ 。仅使用 SIRST 数据集中的训练集进行对抗性训练。实验中使用的所有网络都是用 PyTorch 实现的，并在 NVIDIA Geforce RTX 3090 上运行。

### 4.3 真实数据和合成数据的结果

首先，对真实数据和合成数据进行可视化比较。合成目标的外观类似于一个真实的目标。目标的边界自然被抹去，目标的像素值不再是图像中最亮的。

再次注意，这项工作的最终目标是通过合成训练数据来提高检测网络的性能，而不是渲染真实的红外图像。因此，通过将真实的训练数据与不同数量的合成训练数据相结合来分析检测结果。选择了 5 个检测网络，包括 4 个最先进的检测模型和基线，以显示所提出的合成数据增强的有效性。所选的模型为 MDvsFA<sup>[13]</sup>、IRSTD-GAN<sup>[1]</sup>、DNA-Net<sup>[30]</sup> 和 ACM U-Net<sup>[28]</sup>。表 2 显示了不同合成量的 nIoU、 $Pd$  和  $Fa$  值训练数据在总的训练集中。可以看到以下几点。首先，仅通过合成训练数据训练

的检测网络产生的定位 ( $nIoU$ ) 和检测 ( $P_d$  和  $F_a$ ) 结果的精度要低得多。这意味着在真实数据和合成数据之间仍然存在一个领域的差距。其次，通过将真实训练数据与综合训练数据相结合，可以提高目标定位的精度和检测性能。具体来说，所提出的数据增强方法总是能提高目标定位的精度。在检测性能方面，使用组合训练数据训练的模型仅在两个指标中的至少一个方面比使用真实训练数据训练的模型具有更好的性能。这两种观察结果在所有 5 种模型中都是一致的。

表 2 总体性能比较

训练图数量	真实	160	160	160	160	160	0
	生成	0	68	160	262	376	376
	总和	160	228	320	422	536	376
基线方法	$nIoU$	0.743	0.748	0.765	0.769	0.747	0.459
	$P_d$	0.971	0.990	0.980	0.980	0.961	0.775
	$F_a(10^{-5})$	5.062	2.801	4.486	1.176	3.286	66.706
MDvsFA	$nIoU$	0.340	0.483	0.501	0.534	0.606	0.372
	$P_d$	0.676	0.833	0.804	0.892	0.951	0.804
	$F_a(10^{-5})$	5.225	14.68	6.024	3.564	7.669	20.238
IRSTD-GAN	$nIoU$	0.660	0.688	0.688	0.694	0.704	0.331
	$P_d$	0.961	0.961	0.980	0.951	0.961	0.569
	$F_a(10^{-5})$	5.285	4.911	7.506	4.582	3.521	10.343
DNA-Net	$nIoU$	0.722	0.719	0.730	0.743	0.736	0.441
	$P_d$	0.951	0.980	0.990	0.981	1.000	0.686
	$F_a(10^{-5})$	2.889	1.081	0.946	1.864	1.232	3.699
ACM U-Net	$nIoU$	0.709	0.724	0.733	0.738	0.748	0.300
	$P_d$	0.961	0.961	0.980	0.971	0.951	0.716
	$F_a(10^{-5})$	3.203	3.509	4.967	2.861	2.265	255.366

对于检测结果的定性比较描述了这两个模型的结果：一个由 160 张真实图像训练，另一个由 160 张真实图像和 262 张合成图像的组合训练。与仅用真实图像训练的模型相比，用组合图像训练的模型产生的误警较少，定位精度较高。

#### 4.4 强度调制网络中的消融研究

本节中的消融研究使用基线检测网络进行。首先，研究超参数  $\lambda_{cont}$  对检测性能的影响。使用给定的  $\lambda_{cont}$  值来生成合成数据集。然后，使用 160 张真实图像和 262 张合成图像来训练检测网络。对检测性能的比较结果见表 3。引入对比度损失可以提高目标定位的精度，如表 3 所示。虽然  $\lambda_{cont}=0.03$  提供了最高的  $nIoU$  值，但大于 0.03 的  $\lambda_{cont}$  也实现了类似的目标定位性能。在检测性能方面，所有比较模型给出相似的  $P_d$ ，只有  $F_a$  不同。如表 3 所示，当  $\lambda_{cont}$  设置为 0.1 时， $P_d$  达到最高，当  $\lambda_{cont}$  设置为 0.03 时， $F_a$  达到最低。

表 3  $\lambda_{cont}$  对检测结果的影响

$\lambda_{cont}$	$nIoU$	$P_d$	$F_a(10^{-5})$
0	0.745	0.980	3.461
0.03	0.769	0.980	1.176
0.1	0.761	0.990	4.437
0.3	0.760	0.980	2.774

接下来，评估了强度调制网络的有效性。将所提出的强度调制网络与两种为目标区域分配像素值的变体进行了比较：一种是直接使用植入的目标，另一种是为目标设置一个随机值。当随机选择目标的像素值时，只考虑大于目标周围区域的平均值的值。然后，对目标区域应用高斯模糊处理。用  $nIoU$ 、 $P_d$ 、 $F_a$  评价的检测结果见表 4。即使所提出的强度调制网络没有应用于数据生成过程中，与仅训练真实图像的模型相比，采用人工目标处理的转换图像可以提高  $nIoU$  评分，降低  $F_a$ 。通过采用所提出的强度调制网络，可以进一步提高  $nIoU$  评分和  $P_d$ 。

表 4 强度调制网络上的消融研究

处理方法	$nIoU$	$P_d$	$F_a(10^{-5})$
不处理	0.751	0.971	1.113
随机值	0.746	0.971	4.900
强度调制	0.769	0.980	1.176

## 5 结束语

本文提出了一种新的基于 GAN 的合成数据生成过程来补充训练数据集。在没有复杂的场景和成像过程模拟模型的情况下，可以使用从网络中获得的可见光图像和最近发布的用于红外小目标检测的公共数据集来生成合成的训练数据。虽然生成的图像与真实图像之间存在域上的差距，但实验结果表明，经过真实数据和合成数据混合训练的各种检测网络比单独使用真实图像表现更好。消融研究表明，强度调制网络有助于有效地提高检测性能。在未来的工作中，计划探索领域自适应技术，以填补合成和真实之间的领域空白。

#### 参考文献：

- [1] ZHAO B, WANG C, FU Q, et al. A novel pattern for infrared small target detection with generative adversarial network [J]. IEEE Transactions on Geoscience and Remote Sensing, 2020, 59 (5): 4481 - 4492.
- [2] KIM S, LEE J. Scale invariant small target detection by optimizing signal-to-clutter ratio in heterogeneous background for infrared search and track [J]. Pattern Recognition, 2012, 45 (1): 393 - 406.
- [3] BAE T W, ZHANG F, KWEON I S. Edge directional 2D LMS filter for infrared small target detection [J]. Infrared Physics & Technology, 2012, 55 (1): 137 - 145.
- [4] CHEN C P, LI H, WEI Y, et al. A local contrast method for small infrared target detection [J]. IEEE Transactions on Geoscience and Remote Sensing, 2013, 52 (1): 574 - 581.
- [5] 胡睿杰, 车 逗. 红外小目标检测方法综述 [J]. 计算机与现代化, 2023, 13 (8): 79.
- [6] HAN J, MA Y, ZHOU B, et al. A robust infrared small target detection algorithm based on human visual system [J]. IEEE Geoscience and Remote Sensing Letters, 2014, 11 (12): 2168 - 2172.
- [7] YANG P, DONG L, XU W. Detecting small infrared maritime

- targets overwhelmed in heavy waves by weighted multidirectional gradient measure [J]. *IEEE Geoscience and Remote Sensing Letters*, 2021, 19 (1): 1–5.
- [8] ZHAO M, LI W, LI L, et al. Three-order tensor creation and Tucker decomposition for infrared small-target detection [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2021, 60 (8): 1–16.
- [9] ZHAO M, LI W, LI L, et al. Single-frame infrared small-target detection: A survey [J]. *IEEE Geo-science and Remote Sensing Magazine*, 2022, 10 (2): 87–119.
- [10] GAO C, MENG D, YANG Y, et al. Infrared patch-image model for small target detection in a single image [J]. *IEEE Transactions on Image Processing*, 2013, 22 (12): 4996–5009.
- [11] 焦建彬, 杨舒, 刘峰. 基于人工神经网络的红外小目标检测 [J]. *控制工程*, 2010, 17 (5): 611–613.
- [12] DAI Y, WU Y, SONG Y, et al. Non-negative infrared patch-image model: Robust target background separation via partial sum minimization of singular values [J]. *Infrared Physics & Technology*, 2017, 81 (10): 182–194.
- [13] WANG H, ZHOU L, WANG L. Miss detection vs. false alarm: Adversarial learning for small object segmentation in infrared images [C] // *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019: 8509–8518.
- [14] 史鹏飞, 韩松, 倪建军, 等. 结合数据增强和改进 YOLOv4 的水下目标检测算法 [J]. *电子测量与仪器学报*, 2022, 36 (3): 113–121.
- [15] DAI Y, WU Y, ZHOU F, et al. Asymmetric contextual modulation for infrared small target detection [C] // *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2021: 950–959.
- [16] DRIGGERS R, POLLAK E, GRIMMING R, et al. Detection of small targets in the infrared: an infrared search and track tutorial [J]. *Applied Optics*, 2021, 60 (16): 4762–4777.
- [17] KIM S, LEE J. Small infrared target detection by region-adaptive clutter rejection for sea-based infrared search and track [J]. *Sensors*, 2014, 14 (7): 13210–13242.
- [18] INKAWHICH N, INKAWHICH M J, DAVIS E K, et al. Bridging a gap in SAR-ATR: Training on fully synthetic and testing on measured data [J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2021, 14 (5): 2942–2955.
- [19] KIM J H, KIM J, JOUNG J. Siamese hyperspectral target detection using synthetic training data [J]. *Electronics Letters*, 2020, 56 (21): 1116–1118.
- [20] LIU W, LUO B, LIU J. Synthetic data augmentation using multiscale attention CycleGAN for aircraft detection in remote sensing images [J]. *IEEE Geoscience and Remote Sensing Letters*, 2021, 19 (3): 1–5.
- [21] 陈佛计, 朱枫, 吴清潇, 等. 生成对抗网络及其在图像生成中的应用研究综述 [J]. *计算机学报*, 2021, 44 (2): 347–369.
- [22] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets [J]. *Advances in Neural Information Processing Systems*, 2014, 27 (2): 18–20.
- [23] 肖进胜, 申梦瑶, 雷俊锋, 等. 基于生成对抗网络的雾霾场景图像转换算法 [J]. *计算机学报*, 2020, 43 (1): 165–176.
- [24] SUN X, WANG B, WANG Z, et al. Research progress on few-shot learning for remote sensing image interpretation [J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2021, 14 (21): 2387–2402.
- [25] ISOLA P, ZHU J Y, ZHOU T, et al. Image-to-image translation with conditional adversarial networks [C] // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017: 1125–1134.
- [26] YU F, KOLTUN V. Multi-scale context aggregation by dilated convolutions [J]. *ArXiv Preprint ArXiv: 1511.07122*, 2015.
- [27] 李宗翰, 刘佳, 柯彦, 等. 基于图像翻译的载体选择式图像隐写方案 [J]. *应用科学学报*, 2019, 37 (5): 733–743.
- [28] RONNEBERGER O, FISCHER P, BROX T. U-net: Convolutional networks for biomedical image segmentation [C] // *Medical Image Computing and Computer-Assisted Intervention-MICCAI 2015: 18th International Conference, Munich, Germany, 2015: 234–241*.
- [29] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection [C] // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017: 2117–2125.
- [30] LI B, XIAO C, WANG L, et al. Dense nested attention network for infrared small target detection [J]. *IEEE Transactions on Image Processing*, 2022, 32 (25): 1745–1758.
- [31] 徐光生, 王士同. 基于潜在的低秩约束的不完整模态迁移学习. [J]. *Journal of Frontiers of Computer Science & Technology*, 2022, 16 (12): 519–520.
- [32] ZHAO Y, WU R, DONG H. Unpaired image-to-image translation using adversarial consistency loss [C] // *Computer Vision-ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part IX 16. 2020: 800–815*.
- [33] HWANG S, PARK J, KIM N, et al. Multispectral pedestrian detection: Benchmark dataset and baseline [C] // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015: 1037–1045.
- [34] 赵佳佳, 唐峥远, 杨杰, 等. 基于图像稀疏表示的红外小目标检测算法 [J]. *红外与毫米波学报*, 2011, 30 (2): 156–162.
- [35] GONZÁLEZ A, FANG Z, SOCARRAS Y, et al. Pedestrian detection at day/night time with visible and FIR cameras: A comparison [J]. *Sensors*, 2016, 16 (6): 820.
- [36] WANG T C, LIU M Y, ZHU J Y, et al. High-resolution image synthesis and semantic manipulation with conditional GANs [C] // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018: 8798–8807.
- [37] MAO X, LI Q, XIE H, et al. Least squares generative adversarial networks [C] // *Proceedings of the IEEE International Conference on Computer Vision*, 2017: 2794–2802.