



文献 CSTR:
32001.14.11-6035.csd.2024.0196.zh



文献 DOI:
10.11922/11-6035.csd.2024.0196.zh
数据 DOI:
10.57760/sciencedb.j00001.01190

文献分类: 计算机科学技术

收稿日期: 2024-11-21
录用日期: 2025-07-24
发表日期: 2025-09-22

* 论文通信作者
张文龙: wenlong@nudt.edu.cn

红外弱小运动目标检测与分割半仿真数据集

冯子成^{1,2}, 张文龙^{1,2*}, 孙晓亮^{1,2}, 郭良超^{1,2}, 刘冬辉^{1,2}

1. 国防科技大学空天科学学院, 长沙 410073
2. 图像测量与视觉导航湖南省重点实验室, 长沙 410073

摘要: 红外弱小运动目标检测是目标感知领域的研究热点, 在安防监控、侦察预警等领域具有广泛的应用前景。近年来, 基于深度学习的红外弱小目标检测算法快速发展, 对数据的规模和质量提出了更高要求。然而, 真实数据的大量获取与精细标注难度较大, 现有数据集通常采用高斯合成或抽帧简化方法, 导致目标真实性不足和运动轨迹不连续。针对上述问题, 本数据集提出了融合时空特征的生成对抗网络, 学习真实目标外观变化特性以生成高仿真红外目标, 然后结合已有数据集的丰富场景和多样目标运动轨迹, 最终构建了一套大规模复杂背景下红外弱小运动目标检测与分割半仿真数据集。生成的目标尺寸为 $4\times 4\sim 9\times 9$ 像素, 局部信噪比为 $-5\sim 5$ dB, 符合弱小目标的定义规范。本数据集涵盖建筑、植被、云层和水域四类场景, 包含三种不同目标数量情况, 并提供中心点、目标框和掩膜三种标注形式, 以及实例级和像素级两类评价指标。数据集包含200个序列, 共60000帧图像, 平均分为训练集和测试集两部分, 可为复杂背景下红外弱小运动目标检测、分割和跟踪等研究提供基础数据。

关键词: 红外弱小运动目标; 复杂背景; 目标检测; 目标分割; 序列图像

数据库(集)基本信息简介

数据库(集)名称	红外弱小运动目标检测与分割半仿真数据集
数据作者	冯子成, 张文龙, 孙晓亮, 郭良超, 刘冬辉
数据通信作者	张文龙 (wenlong@nudt.edu.cn)
数据时间范围	2020–2021 年
地理区域	中国
数据量	~10 GB
数据格式	*.png, *.json
数据服务系统网址	https://doi.org/10.57760/sciencedb.j00001.01190
基金项目	国家自然科学基金 (12202485)
数据库(集)组成	本数据集包含 200 个图像序列, 共 60000 帧图像, 均分为训练集和测试集各 100 个图像序列。每个序列对应一个文件夹, 文件夹中包含图像序列文件 (*.png)。标注分为两个文件夹, 一个包含像素级掩膜标注文件 (*.png), 命名方式与图像序列文件相同, 另一个包含实例级标注文件 (*.json), 以图像序列名命名, 具体包含目标尺寸、相对速度和局部信噪比等属性信息, 以及每一帧图像对应的目标数、目标位置坐标和尺寸。

引言

红外目标检测是一项处理分析红外图片和视频以搜寻目标的技术。其中，红外图片和视频通过红外成像技术获得，该技术利用红外传感器感知物体表面的热辐射并将其转换成可视的图像，具有感知距离远、全天时、全天候工作等优势。受益于此，红外目标检测成为目标感知研究的关键核心技术之一，并广泛应用于军事、安保、交通等领域。本数据集旨在仿真空中飞行平台对空中运动目标的红外探测，模拟空中平台预警、监测、制导等实际应用场景。在上述场景中，红外目标检测任务被细分为红外弱小运动目标检测，其具体特点和定义为：

(1) 远距离目标的成像尺寸较小，空间分辨率低，根据国际光学工程学会的定义，将像素面积不超过 9×9 的红外目标定义为红外小目标。

(2) 目标与背景的红外辐射强度差异较小，亮度特征对比度较弱，局部信噪比不大于 $5\text{dB}^{[1]}$ 。

(3) 背景环境复杂，包括云层、植被、建筑、水域等多样场景，存在诸多干扰物体，易于产生虚警。

(4) 目标和背景动态变化，不仅目标和背景的位置发生改变，目标的形态、亮度等外观特征也会随时间变化。

红外弱小目标的检测是一项具有挑战性的课题，也是当前研究的热点，其目的是实现潜在隐藏或微小热源目标的高灵敏度探测。在红外弱小目标检测研究领域，仅使用单幅图像进行目标检测的方法称为单帧红外弱小目标检测，相关算法仅利用空域信息检测弱小目标；利用连续多帧图像进行运动目标检测的方法称为序列红外弱小目标检测或红外弱小运动目标检测，相关算法可以联合利用目标的空域信息和时域运动信息。

当前，单帧红外弱小目标检测技术日趋成熟，相关算法实现了较高检测率和较低的虚警率；而序列红外弱小目标检测技术则发展缓慢，相关算法的性能未能充分体现增加时域运动信息的优势。造成这种情况的原因是近年来发布了大量丰富的单帧目标检测数据集，并使用了更为精确的像素级目标掩膜标注，这为基于深度学习的目标检测模型提供了更强的训练监督模式，使得基于图像分割的目标检测算法成为单帧红外弱小目标检测的主流方向。然而，相比较标注单帧图片，标注序列图像的工作量急剧增加，因此现有的红外弱小运动目标检测数据集数量较少，且大部分采用了实例级的目标框标注方式，或者更为简单的目标中心点标注方式。回丙伟等^[2]发布了红外相机拍摄的小型固定翼无人机目标数据集，其使用了中心点坐标进行标注，难以在基于分割的目标检测模型上应用。孙晓亮等^[3]发布了使用中心点标注的红外弱小运动目标检测仿真数据集，其包含了各类复杂场景和多样目标运动轨迹，但其使用高斯分布模拟的红外点目标具有固定、易辨识的外观特征，与真实目标的外观特征存在较大差异。孙衡等^[4]发布了唯一的使用目标掩膜标注的红外弱小运动目标检测真实数据集。但是，一方面该数据集对原视频进行了抽帧处理，损坏了目标运动轨迹的连续性，不利于提取目标时域运动特征；另一方面，探测视角多为由地面观察天空中目标，缺少平视或者下视视角，背景场景较为单一。综上所述，在红外弱小运动目标检测研究领域，存在数据集匮乏和现有数据不能满足需求的问题，限制了相关检测算法的研究。

本数据集综合上述数据集[3]和[4]的优势，即孙晓亮的数据集中复杂的场景和多样的目标运动轨迹，以及孙衡的数据集中真实的目标外观和精确的掩膜标注，采用半仿真的方式生成红外弱小运动目标视频序列。具体实现过程中，为生成更逼真的视频数据，本数据集设计了一个生成对抗网络，

通过学习孙衡的数据集中真实目标的时序外观特征变化模式，生成多样化的目标序列并植入到孙晓亮的数据集的复杂背景中。通过上述方法，本数据集生成了场景丰富、目标真实、标注精准的高质量数据，旨在促进红外弱小运动目标检测、分割与跟踪领域研究的发展。

1 数据采集和处理方法

1.1 数据集生成模型总体框架

本数据集采用生成对抗网络来生成空中飞行平台对运动目标的红外探测数据。生成对抗网络（generative adversarial network, GAN）是一种广泛使用的深度学习模型，应用于图像生成、图像修复等任务，其核心思想是通过设计生成器和判别器进行对抗性训练，使得两者相互竞争，最终输出高质量的数据样本。其中，生成器（generator, G）负责生成尽可能真实的样本，而判别器（discriminator, D）负责判断输入的样本是真实样本还是生成样本，两者通过交替训练进行优化，形成一个相互博弈过程。目前，研究人员使用生成对抗网络实现了可见光图像到红外图像的转化、红外弱小目标图片数据的生成^[5]等相关任务，得到了逼真的生成数据。然而，这些网络模型仅能利用图像的空域特征，并不适合在时、空域中存在特征变化的红外弱小运动目标数据生成任务。

鉴于上述分析，本数据集设计了一种融合时空特征的生成对抗网络，总体框架如图 1 所示。该网络模型以孙晓亮数据集的背景样本和孙衡数据集的目标样本为输入，经过生成路径得到红外弱小目标视频帧和相应的标注，该路径共包含四个步骤：预处理、生成目标、目标嵌入和生成标注。判别路径仅用在模型训练过程中，该路径可分为两个步骤：生成目标判别和真实目标判别，最终得到的对抗损失用以优化网络模型参数。模型中的生成器和判别器使用了卷积长短期记忆网络^[6]（convolutional long short term memory network, ConvLSTM）来提取和融合目标的时空特征。

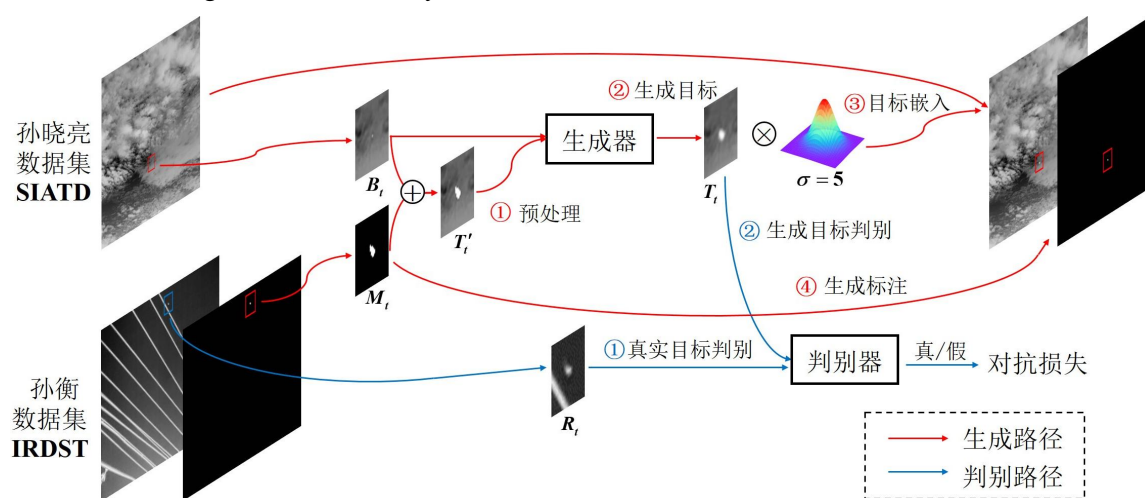


图 1 数据生成模型总体框架

Figure 1 Overall framework of the data generation model

1.2 数据生成模型及训练方法

1.2.1 生成路径

生成路径包含四个步骤：预处理、生成目标、目标嵌入和生成标注。其中，目标嵌入和生成标注仅用于数据生成的后处理，不参与网络模型训练过程。

(1) 预处理

预处理步骤旨在得到生成器的规范化输入，具体处理步骤如下：

- 1) 分别随机选取孙晓亮数据集和孙衡数据集中的一个视频序列作为背景和目标序列。
- 2) 对于时刻 t ，根据目标位置裁剪目标、背景和掩膜的局部图像块，得到真实目标 R_t 、局部背景 B_t 和局部掩膜 M_t ，其尺寸均为 32×32 ，通道数均为 1。
- 3) 对 R_t 、 B_t 和 M_t 进行标准化处理：将 R_t 和 B_t 映射到 $[-1, 1]$ ；将 M_t 映射到 $[0, 1]$ ，其中 0 表示背景，1 表示目标。
- 4) 将局部掩膜 M_t 嵌入到局部背景 B_t 中，得到初始的生成目标 T'_t ，计算过程如下：

$$T'_t = M_t + B_t \odot (1 - M_t) \quad (1)$$

其中， $\odot$ 表示按像素乘法，部分结果如图 2 所示。

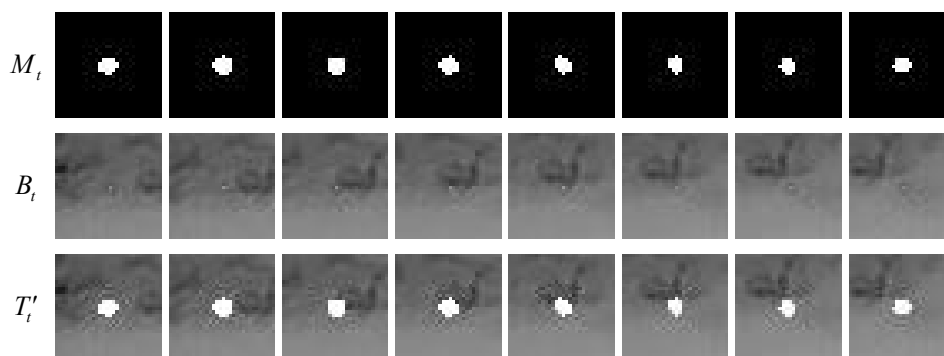


图 2 预处理的初始生成目标结果

Figure 2 Initial generated targets after preprocessing

(2) 生成目标

生成器的结构如图 3 所示，其处理流程表达为公式 (2)。生成器的输入为按通道维度拼接的局部背景 B_t 和初步生成目标 T'_t ，其中，前者负责提供完整的背景信息，后者负责提供基本的目标形状信息；经过多尺度空间特征提取模块、时空特征融合模块和生成输出模块后，得到最终的生成目标 T_t 。

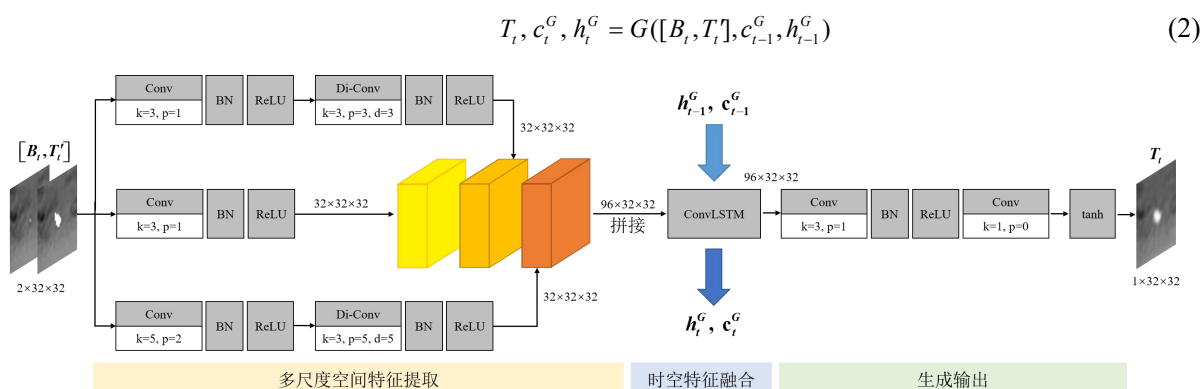


图 3 生成器模型

Figure 3 The generator model

1) 多尺度空间特征提取模块

该模块利用不同卷积核的卷积层和不同扩张率的空洞卷积层 (dilated convolution, Di-Conv) 实现对输入图像的多尺度空间特征提取。空洞卷积通过在卷积核元素之间添加空格来扩张卷积核的

尺寸, 如图 4 所示, 它能够在不增加参数的前提下增大感受野。

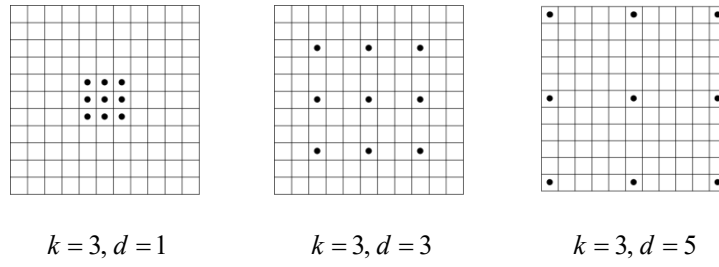


图 4 不同扩张率的空洞卷积

Figure 4 Dilated convolutions with different dilations

多尺度空间特征提取模块如图 3 左侧部分所示, 其中 k, d, p 分别表示卷积核尺寸、扩张率和边界填充宽度, 该模块分别通过三个分支提取了三个尺度的空间特征: 中间分支使用了卷积核为 3×3 的卷积层, 可以提取感受野为 3×3 的小尺度空间特征; 上分支使用了卷积核为 3×3 的卷积层以及卷积核为 3×3 、扩张率为 3 的空洞卷积层, 可以提取感受野为 9×9 的中等尺度空间特征; 下分支使用了卷积核为 5×5 的卷积层以及卷积核为 3×3 、扩张率为 5 的空洞卷积层, 可以提取感受野为 15×15 的大尺度空间特征。

2) 时空特征融合模块

时空特征融合模块使用 ConvLSTM 来融合来自不同时刻的目标空间特征, 实现目标的时空特征提取。ConvLSTM 是传统一维长短期记忆网络 (long short term memory network, LSTM) 拓展到二维的变体, 通过用卷积层代替原来的全连接层, 可以适用于二维图像数据的输入, 其结构如图 5 所示。其中, c_t 表示 t 时刻的单元状态, 起到记忆存储的作用; h_t 表示 t 时刻的隐变量, 即模块输出; f_t 表示遗忘门, 用来控制保留上一时刻的记忆; i_t 表示输入门, 用来控制更新当前时刻的记忆; o_t 表示输出门, 用来控制当前时刻记忆到隐变量的输出。通过门控机制, LSTM 可以比循环神经网络 (recurrent neural network, RNN) 更好地捕捉序列数据中的长期依赖关系。

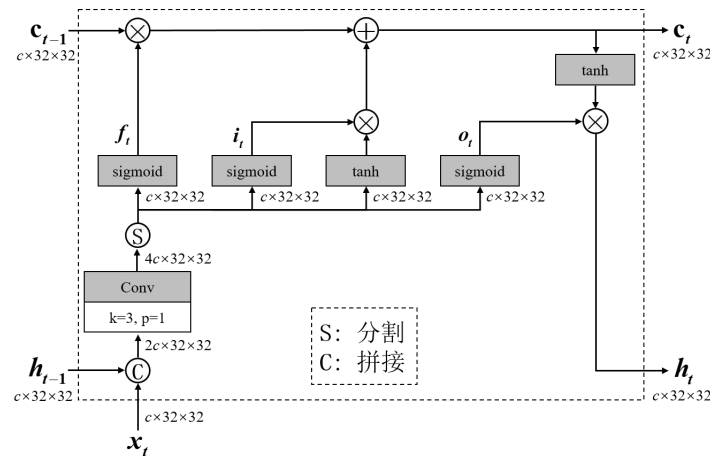


图 5 卷积长短时记忆网络

Figure 5 Convolutional long short-term memory network

3) 生成输出模块

如图 3 右侧部分所示, 该模块使用了一个 CBR 结构, 即卷积层、批标准化层 (batch normalization, BN) 和 ReLU 激活函数, 然后使用一个卷积核为 1×1 的卷积层将通道数降为 1, 最后经过 $\tanh$ 激活函数得到最终的生成目标 T_i 。部分生成结果如图 6 所示。

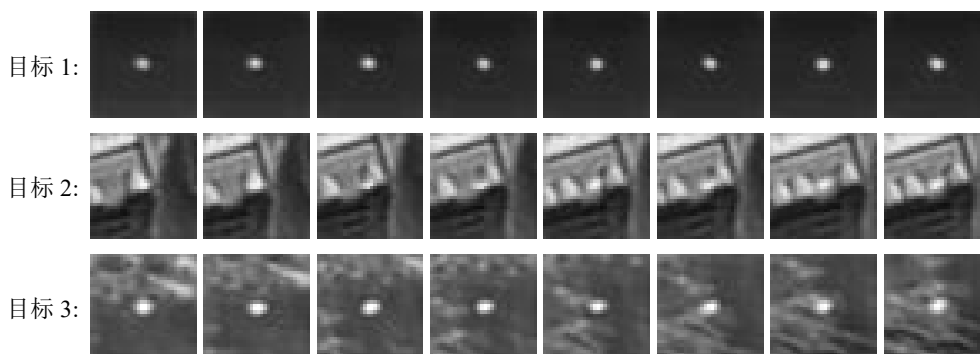


图 6 最终生成目标结果

Figure 6 Final generated targets

(3) 目标嵌入

该步骤用于将生成目标 T_i 植入背景中的目标位置, 得到期望的红外弱小目标视频帧。因为 T_i 中的局部背景经过生成器会产生轻微的亮度变化, 若直接将其嵌入背景会形成明显的正方形接缝, 如图 7 第一行所示。为解决这个问题, 目标嵌入过程中引入了高斯权重, 二维高斯加权公式如下:



图 7 是否使用高斯加权进行目标嵌入的结果对比

Figure 7 Comparison of results with or without Gaussian weighting for target implantation

$$G = \exp\left(-\frac{1}{2\sigma^2}(x^2 + y^2)\right), G \in \mathbb{R}^{32 \times 32}$$

$$T_i^* = G \odot T_i + (1 - G) \odot B_i \quad (3)$$

其中, $\sigma = 5$ 为标准差, 坐标 (x, y) 以区域中心为原点, T_i^* 为加权后的目标区域, 它将直接替换背景中相应的局部区域以完成目标嵌入, 最终嵌入的结果如图 7 第二行所示。图 8 展示了使用不同 σ 的结果: 当 σ 较小时, T_i 中心区域的高斯权重过低, 导致目标特征被损坏; 当 σ 较大时, T_i 边界区域的高斯权重过高, 导致接缝无法被彻底消除; 整体而言, $\sigma = 5$ 为合适的设置。

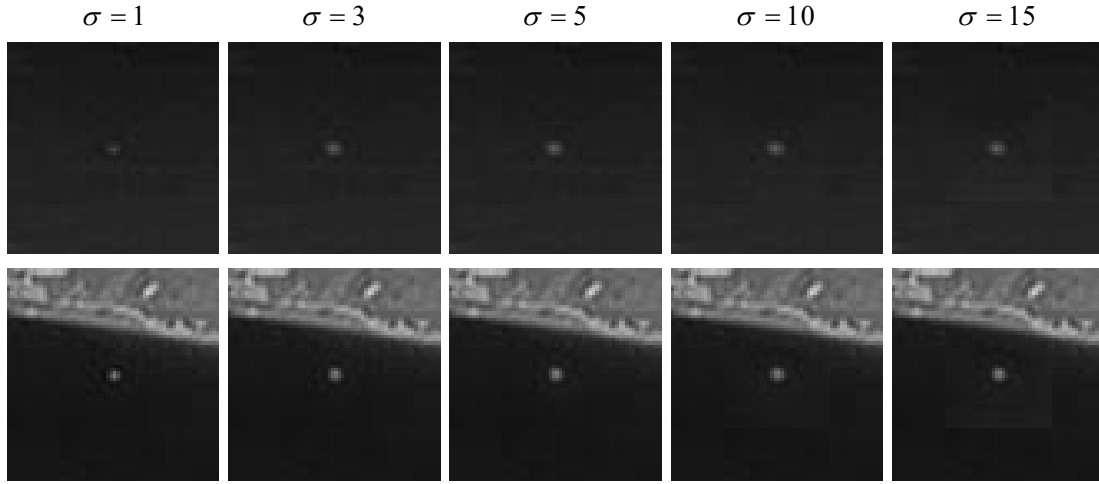


图 8 使用不同标准差 σ 的结果对比

Figure 8 Comparison of results using different standard deviations

(4) 生成标注

根据相应的目标位置, 局部掩膜 M_i 被嵌入到与视频帧同尺寸的空白图像中, 得到像素级的目标掩膜标注, 其中目标像素值为 255, 背景像素值为 0。然后, 通过计算掩膜的质心和外接矩形, 可以得到中心点和目标框标注。

1.2.2 判别路径

判别路径仅用于模型训练, 主要任务是使用判别器区分输入是否为真实目标, 共包含两个步骤: 真实目标判别和生成目标判别。这两个步骤仅有输入和真值的差异, 并需要在训练过程中交替进行。

判别器的结构如图 9 所示, 其输入为真实目标 R_i 或者生成目标 T_i , 经过空间特征提取模块、时空特征融合模块和判别输出模块后, 借鉴 PatchGAN 算法^[7]的思路, 判别器需要给出图像中每个像素的判别结果, 最后与真值对比得到对抗损失。在真实目标判别步骤中, 输入真实目标 R_i , 判别器的输出为 O_i^{real} , 如公式 (4) 所示, 此时真值为 1; 在生成目标判别步骤中, 输入生成目标 T_i , 判别器的输出为 O_i^{gen} , 如公式 (5) 所示, 此时真值为 0。

$$O_t^{real}, c_t^D, h_t^D = D(R_t, c_{t-1}^D, h_{t-1}^D) \quad (4)$$

$$O_t^{gen}, c_t^D, h_t^D = D(T_t, c_{t-1}^D, h_{t-1}^D) \quad (5)$$

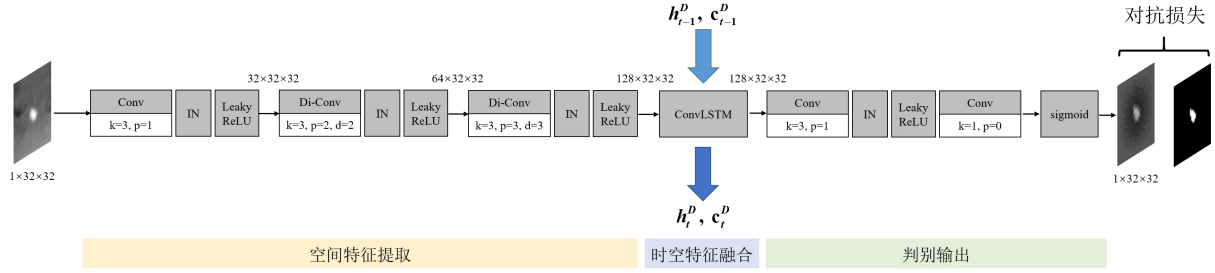


图9 判别器模型

Figure 9 The discriminator model

(1) 空间特征提取模块

如图9左侧部分所示，该模块使用一个卷积层和两个空洞卷积层提取输入图像的空间特征，其感受野为 10×10 ，并使用了实例标准化（instance normalization, IN）和 Leaky ReLU 激活函数。

(2) 时空特征融合模块

该模块与生成器中的时空特征融合模块相同，用以融合来自不同时刻的目标空间特征，实现目标的时空特征提取。

(3) 判别输出模块

如图9右侧部分所示，该模块使用了一个 CBR 结构，然后使用一个卷积核为 1×1 的卷积层将通道数降为 1，最后经过 sigmoid 激活函数得到最终的判别结果。

1.2.3 模型训练

模型训练含两个阶段，分别为判别器训练阶段和生成器训练阶段，这两个阶段交替进行。图10展示了一个样本的模型训练过程。

首先训练判别器 D ，其损失函数如下所示：

$$Loss_{D,t}^{real} = BCE(M_t \odot O_t^{real}, M_t \times 1) \quad (6)$$

$$Loss_{D,t}^{gen} = BCE(M_t \odot O_t^{gen}, M_t \times 0) \quad (7)$$

$$Loss_D = \frac{1}{N} \sum_{t=1}^N (Loss_{D,t}^{real} + Loss_{D,t}^{gen}) \quad (8)$$

其中，训练过程使用了二元交叉熵损失（binary cross-entropy loss, BCE），并使用局部掩膜 M_t 将损失计算限制在目标区域， N 表示一个批次样本中包含的连续帧数。判别器损失 $Loss_D$ 仅用于更新判别器的参数，而生成器 G 的参数在此时冻结。

然后训练生成器 G ，其损失函数如下所示：

$$Loss_{G,t} = BCE(M_t \odot O_t^{gen}, M_t \times 1) + \lambda \times L_1((1 - M_t) \odot T_t, (1 - M_t) \odot B_t) \quad (9)$$

$$Loss_G = \frac{1}{N} \sum_{t=1}^N Loss_{G,t} \quad (10)$$

其中，使用 L_1 损失促使生成的 T_t 的背景部分与真实背景 B_t 相同，并设置损失系数 $\lambda = 0.1$ 。生

成器损失 $Loss_G$ 仅用于更新生成器的参数，而判别器 D 的参数在此时被冻结。

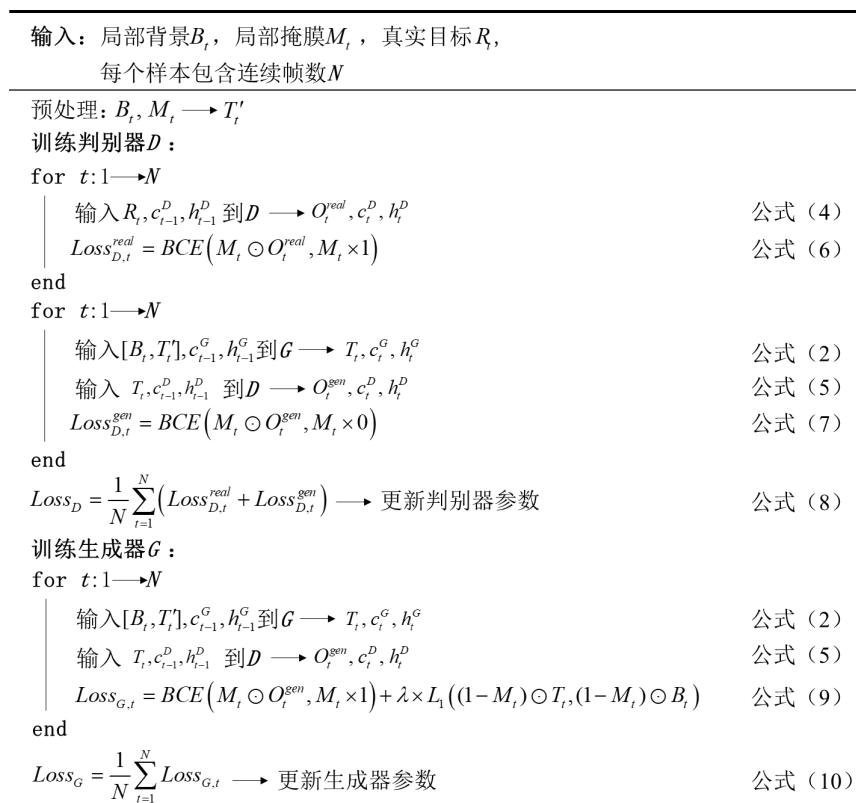


图 10 一个样本的模型训练过程

Figure 10 Model training process for a sample

模型使用 PyTorch 实现，并使用 1 个 NVIDIA RTX 3090 GPU 进行训练和推理。训练过程使用 Adam 优化器，共进行 200 次迭代，学习率为 0.001，批次大小设置为 32，每个样本中连续帧数 N 设置为 20。 c^G, h^G 和 c^D, h^D 初始化为零张量。

训练过程如图 11 所示。生成器的损失值呈现先下降后小幅波动的趋势，表明生成器成功学习到了真实红外小目标的外观特性。判别器的损失值呈现先上升、再下降、后小幅波动的趋势，其原因是：在训练开始阶段，生成器性能快速提升，生成样本逐渐逼真，导致判别器损失上升；然后判别器通过训练逐渐提高判别性能从而降低损失，并在对抗学习中性能逐渐趋于稳定。当第 110 次迭代后，判别器性能处于稳定阶段，且生成器损失 $Loss_G$ 最小，因而保存为最优模型参数。

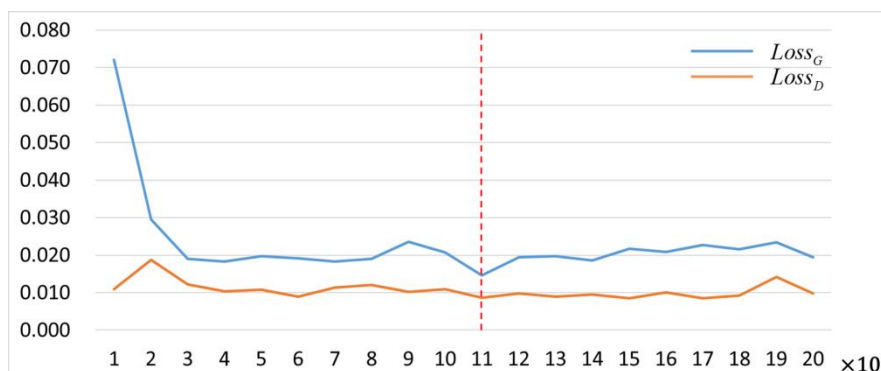


图 11 模型训练过程的损失曲线

Figure 11 Training loss curve of the model

2 数据样本描述

本数据集分为训练集和测试集，各包含 100 段图像序列，每段序列包含 300 帧，共计 60000 帧，统计全部图像均值为 98.28，标准差为 45.16。如表 1 所示，根据目标数量可分为单目标共 100 段、双目标共 49 段及三目标共 51 段。如表 2 所示，根据场景类别又可以分为建筑共 94 段、植被共 38 段、云层共 37 段及水域共 31 段。每段图像序列对应一个文件夹，文件夹以背景来源的孙晓亮数据集的图像序列号命名。

表 1 目标数量分布

Table 1 Distribution of target counts

目标数量	1	2	3
训练集	49	25	26
测试集	51	24	25

表 2 场景类别分布

Table 2 Distribution of scene category

场景类别	建筑	植被	云层	水域
训练集	42	19	21	18
测试集	52	19	16	13

图 12 展示了目标尺寸、速度和局部信噪比等三个属性分布情况。其中，目标尺寸为目标面积的平方根，计算如公式 (11) 所示， h 和 w 分别为目标的高度和宽度，单位为像素，所有目标符合尺寸不超过 9 的小目标规范。目标的速度定义为目标相对背景的运动速度，单位为像素，计算如公式 (12) 所示， Δx 和 Δy 为目标在图像上的横向、纵向偏移量， dx 和 dy 为使用相邻帧图像配准得到的横向、纵向背景偏移量。目标的局部信噪比 (local signal-to-noise ratio, LSNR) 计算如公式 (13) 所示，单位为分贝， E_T 和 E_B 分别为目标和局部背景的均值， δ_B 为局部背景的标准差，局部背景为目标四周扩展 20 个像素的区域，所有目标符合 LSNR 不大于 5 dB 的弱目标规范。如表 1、表 2 和图 12 所示，训练集和测试集在场景类别、目标数量、尺度、速度和 LSNR 等方面均具有相似分布。

$$size = \sqrt{h \times w} \quad (11)$$

$$speed = \sqrt{(\Delta x - dx)^2 + (\Delta y - dy)^2} \quad (12)$$

$$LSNR = 10 \times \log_{10} \frac{E_T - E_B}{\delta_B} \quad (13)$$

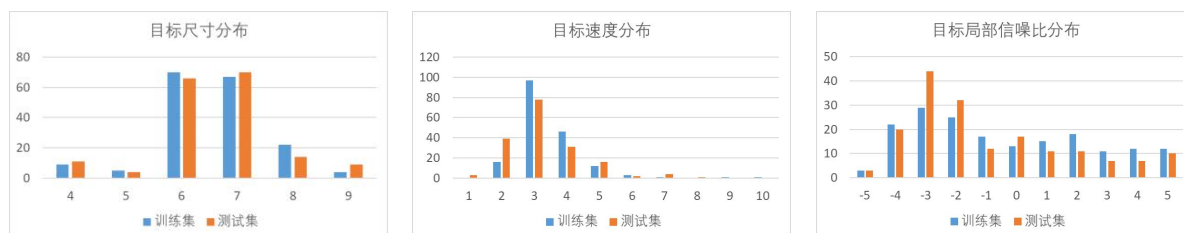


图 12 目标属性分布

Figure 12 Distribution of target attributes

2.1 序列图像数据

本数据集中红外图像分辨率为 640×512 像素，位深度为 8 bit，图像以 png 格式存储，图像序列中图像帧按顺序依次编号，即图像帧命名从“0.png”依次递增至“299.png”。图 13 给出了三种目标数目类别在四种场景下的样本图像示例。

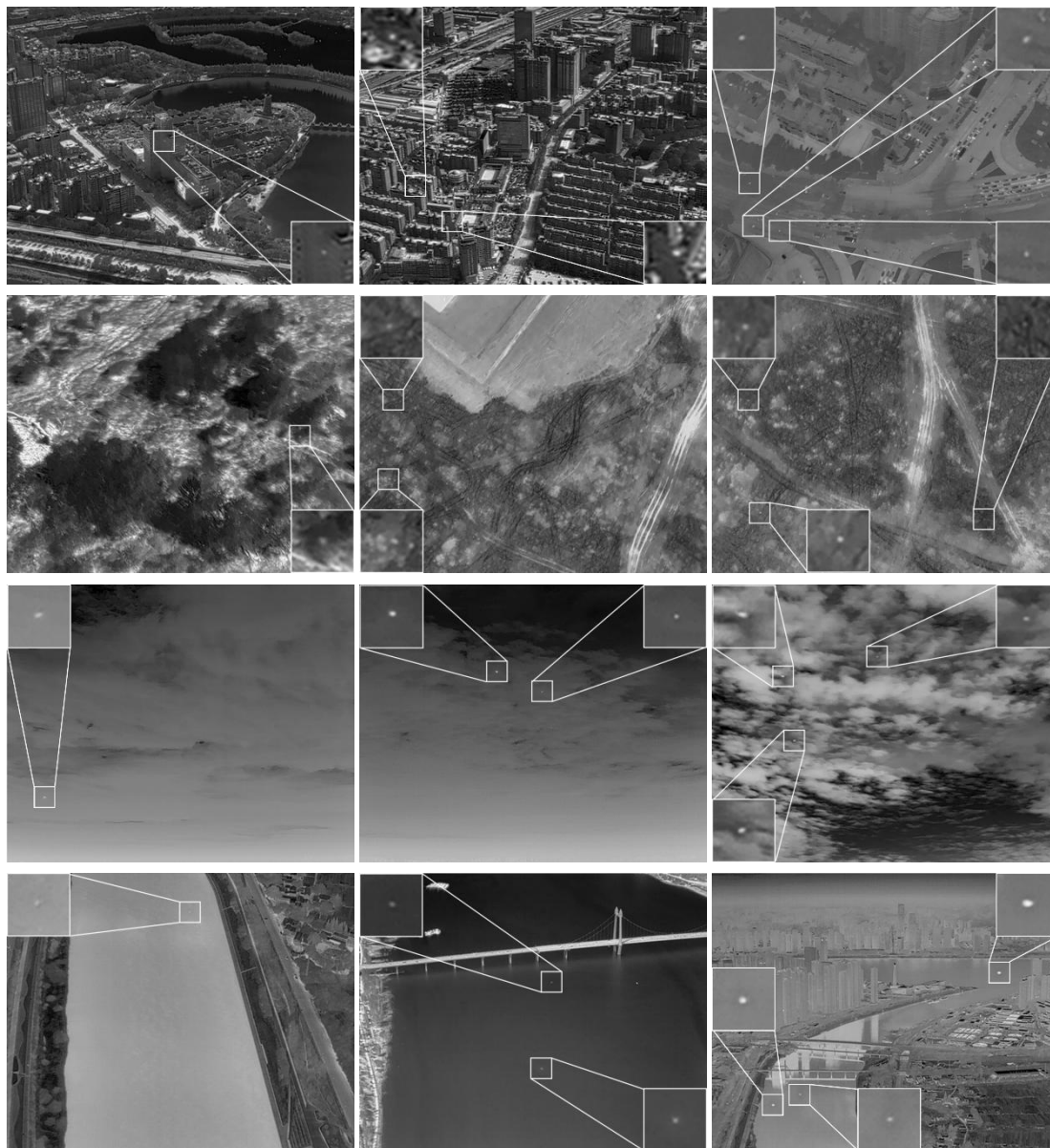


图 13 样本图像示例

Figure 13 Example image samples

2.2 标注数据文件

标注数据方式有两种，一种是像素级掩膜标注，以“*.png”格式存储于“masks”文件夹，其分辨率为 640×512 像素，位深度为 8 bit，并与红外图像命名方式相同；另一种是实例级目标框标注，以“*.json”格式存储于“labels”文件夹，每个序列对应一个 json 文件，文件以序列名命名，文件包含“meta”和“label”两个“名称/值”对。其中，“meta”又包含 3 个“名称/值”对，名称如下：

- avg_size

- avg_speed_rel
- avg_lsnr

分别对应目标的尺寸、速度和 LSNR。

“label”包含 5 个“名称/值”对，名称如下：

- sequence_id
- frame
- num_objects
- center_coords
- size

分别对应序列编号、帧编号、当前帧中目标数目、目标中心坐标值与目标尺寸。图 14 给出掩膜标注和目标框标注，示例展示了 14 号图像序列中 5 号帧图像中目标信息，该帧图像中共包含 1 个目标，并给出了目标中心位置和尺寸真值。

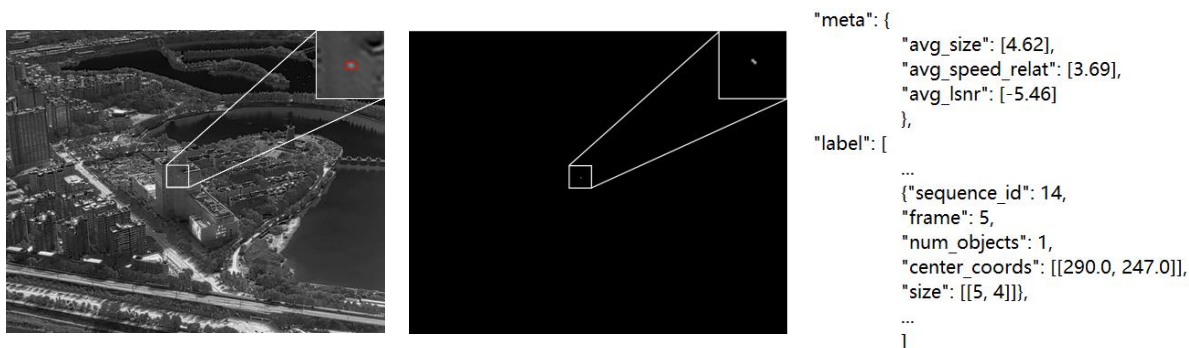


图 14 标注文件结构示例

Figure 14 Example of annotate file structure

格式相关说明如下：

- (1) 标注文件中所有数字均为阿拉伯数字。
- (2) 目标数类别对应该图像序列中包含目标数目的多少，本数据涵盖单目标、双目标和三目标三类。
- (3) 标注文件中帧编号从 0 开始，与图像帧命名一致。
- (4) 目标位置真值对应 x、y 坐标值，均为浮点型，对应图像坐标系，即原点位于图像帧左上角点，起始点坐标为 (0,0)，水平向右为 x 轴正向、竖直向下为 y 轴正向，x、y 坐标对应范围为 [0, 639]、[0, 511]。
- (5) 目标尺寸真值对应目标框的宽度和高度。

3 数据质量控制和评估

3.1 数据质量控制

为构建覆盖场景类型范围广、数据集规模大、贴近真实应用场景的复杂背景下红外弱小运动目标检测与分割数据集，本数据集采用半仿真方式生成图像序列，其中背景图像数据来自孙晓亮数据集中真实拍摄的红外场景视频，目标图像数据来自生成对抗网络的生成图像。为保证数据真实，生成对抗网络使用孙衡数据集中真实目标图像进行训练。对于生成的红外弱小运动目标的嵌入，通过

合理设计第 1.2.1（3）节阐述的目标嵌入步骤，有效解决接缝问题。数据的所有标注均通过程序自动生成，如第 1.2.1（4）节所述。

针对生成的序列图像数据和标注，两名长期从事红外弱小运动目标检测的研究人员负责数据集的质量控制。其中，一人负责计算生成数据的目标尺寸和 LSNR 指标，并剔除尺寸大于 9×9 像素或 LSNR 大于 5 dB 的目标；另一人负责视觉检测生成数据是否有丢帧、明显接缝、标注错误等问题。两名研究人员交替执行上述质量检查，共检查两个轮次，以确保所有数据严格符合弱小目标的定义规范。

3.2 与已有数据集比较

本数据集与两个数据集紧密相关，一个是孙晓亮等^[3]于 2021 年发布的复杂背景下红外运动小目标检测半仿真数据集，另一个是孙衡等^[4]于 2023 年发布的红外弱小运动目标检测数据集。表 3 给出了本数据集与这两个数据集的详细对比情况。相比使用高斯分布模拟目标的孙晓亮数据集，本数据集使用生成对抗网络生成多样且真实的目标外观，且提供了更详细的标注方式。相比具有掩膜标注的孙衡数据集，本数据集具有更大的规模，共计 200 个序列和 60000 帧图像，包含更多样的目标运动和更平滑的背景运动，且严格遵守目标尺寸小于 9×9 像素和 LSNR 小于 5dB 的弱小目标规范。另外，就场景类型分布而言，本数据集更偏重复杂的城市背景，而孙晓亮数据集和孙衡数据集则偏重较为简单的天空云层背景。

表 3 本数据集与两个相关数据集对比

Table 3 Comparison between the dataset and two existing datasets

比较项	本数据集（IDSMT）	复杂背景下红外运动小目标检测半仿真数据集（SIATD）	红外弱小运动目标检测数据集（IRDST）
数据集规模	200 个序列、60000 帧	350 个序列、160125 帧	85 个序列、40650 帧
图像分辨率	640×512	640×512	720×480、934×696
涵盖场景类型 （序列数目）	建筑（94）、植被（38）、云层（37）、水域（31）	建筑（70）、植被（70）、云层（140）、水域（70）	建筑（14）、植被（25）、云层（34）、水域（12）
视角	仰视、平视、俯视	仰视、平视、俯视	仰视、平视
标注方式	中心点标注、目标框标注、掩膜标注	中心点标注	中心点标注、目标框标注、掩膜标注
目标外观	点目标、扩展目标；生成对抗网络仿真目标，特征多样	点目标；高斯分布仿真目标，特征简单、易辨识	点目标、扩展目标；真实目标，特征多样
目标尺寸范围	4×4~9×9 像素	3×3~7×7 像素	4×4~13×13 像素
目标局部信噪比	-5~5dB	-8~3dB	-6~9db
目标运动特性	运动轨迹多样且连续性好	运动轨迹多样且连续性好	运动轨迹连续性差，存在位置跳变
背景运动特性	无人机飞行平台搭载红外相机拍摄，无抽帧处理，运动平缓	无人机飞行平台搭载红外相机拍摄，无抽帧处理，运动平缓	手持云台安装红外相机拍摄，每 3 帧抽 1 帧，运动较剧烈
评价指标	实例级评价指标和像素级评价指标，公布测试代码	实例级评价指标，公布测试代码	实例级评价指标和像素级评价指标，未公布测试代码

图 15 展示了本数据集与两个数据集中目标外观的对比，孙晓亮数据集的目标外观呈近似中心对称特点，简单且易辨识，而本数据集的目标外观与孙衡数据集的真实目标外观在亮度、形态等方面高度相似，证明本数据集使用生成对抗网络高度仿真了真实目标特征。图 16 展示了本数据集与两个数据集中目标运动轨迹的对比，相比孙衡数据集中不规则、跳变的目标轨迹，本数据集继承了孙晓亮数据集中目标运动轨迹的平滑、连续、多样等特点，能够为目标检测模型提供更有效的目标时域运动信息。

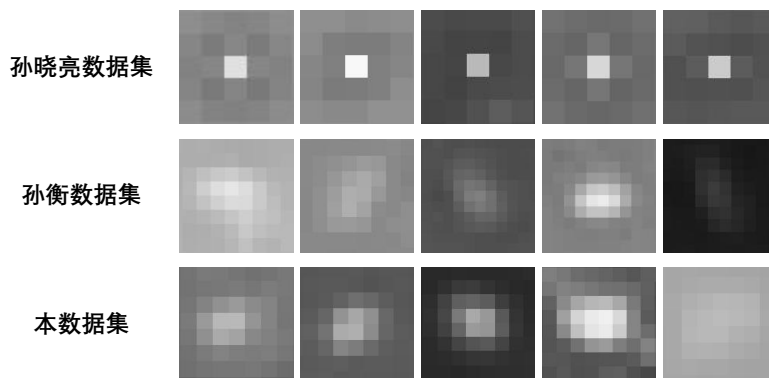


图 15 本数据集与两个数据集中目标外观对比

Figure 15 Comparison of target appearances between the dataset and two existing datasets

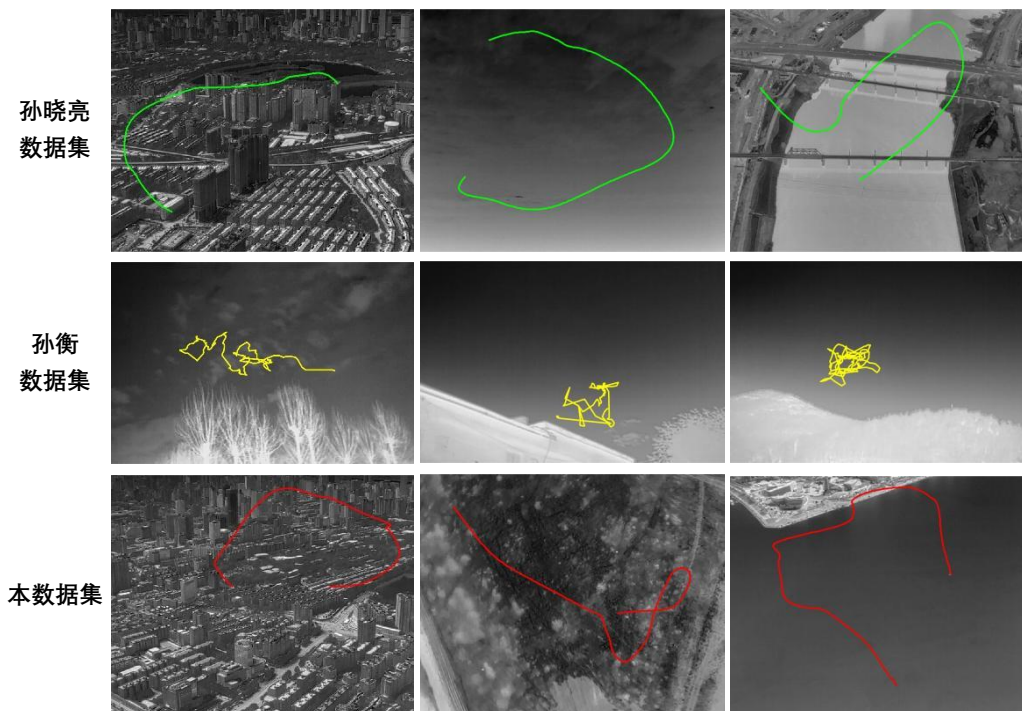


图 16 本数据集与两个数据集中目标运动轨迹对比

Figure 16 Comparison of target motion trajectories between the dataset and two existing datasets

4 数据使用方法和建议

4.1 数据集使用方法

本数据集使用方法简单，研究人员从数据网站上下载本数据集，遵循如下基本步骤使用本数据

集:

- (1) 读取数据段中数据, 运行红外弱小运动目标检测算法, 完成目标检测任务。
- (2) 按照 2.2 节中约定的标注文件格式, 保存目标检测结果。
- (3) 使用公布的测试代码计算评价指标。

4.2 评价指标

参考红外弱小运动目标检测领域研究工作, 本数据集设置了实例级和像素级两类评价指标, 前者包括精确率 P 、回归率 R 和 F_1 指标^[8], 后者包括 IoU 和虚警率 F_a 。

实例级评价指标如下所示:

$$\begin{aligned} F_1 &= \frac{2 \times P \times R}{P + R} \\ P &= \frac{TP}{TP + FP} \\ R &= \frac{TP}{TP + FN} \end{aligned} \quad (14)$$

其中, TP 、 FP 和 FN 分别对应正确检测目标、误检目标和漏检目标, 检测目标与真实目标的距离小于阈值 τ 即认为该目标为正确检测目标, 本数据集中设置 $\tau=3$ 像素。

像素级评价指标如下所示:

$$\begin{aligned} IoU &= \frac{\sum_{i=1}^n A_i^p \cap A_i^g}{\sum_{i=1}^n A_i^p \cup A_i^g} \\ F_a &= \frac{P_{false}}{P_{All}} \end{aligned} \quad (15)$$

其中, IoU 为预测目标区域 A_i^p 和真值 A_i^g 在所有 n 帧上的交并比, 虚警率 F_a 是错误预测像素 P_{false} 与所有像素 P_{All} 的比值。

4.3 参考算法结果

表 4 展示了一些高性能目标检测算法在本数据集上评价指标结果, 其中, 检测率 P_d 与回归率 R 的定义相同, 图 17 展示了相关算法的 P - R 图和 P_d - F_a 图。其中, RDIAN^[4]、MSHNet^[9]、DNANet^[10]、ISTDU-Net^[11]和 SCTransNet^[12]为基于分割的单帧红外弱小目标检测算法; SSTNet^[13]、TAD^[14]和 Tridos^[15]为基于实例的红外弱小运动目标检测算法。整体而言, 基于分割的检测算法表现出更好的性能, 而序列检测算法未能充分体现增加时域运动信息的优势。

表 4 参考算法在本数据集上评价结果

Table 4 Evaluation results of the reference algorithms on the dataset

算法	RDIAN	MSHNet	DNANet	ISTDU-Net	SCTransNet	SSTNet	TAD	Tridos
P	0.8182	0.9422	0.9391	0.9659	0.9516	0.9326	0.8985	0.9213
$R(P_d)$	0.8588	0.8101	0.8375	0.8636	0.9050	0.5876	0.6428	0.7323
F_1	0.8380	0.8711	0.8854	0.9119	0.9277	0.7210	0.7494	0.8160

算法	RDIAN	MSHNet	DNANet	ISTDU-Net	SCTransNet	SSTNet	TAD	Tridos
$F_a(\times 10^{-6})$	3.8900	1.3860	1.3546	0.5984	0.8608	—	—	—
IoU	0.6308	0.6418	0.6903	0.6774	0.7359	—	—	—

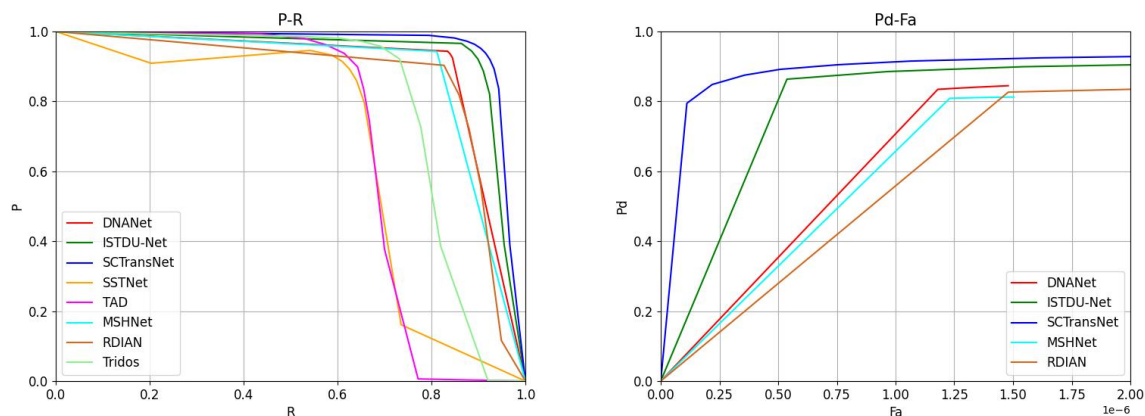


图 17 参考算法的 P-R 图和 Pd-Fa 图

Figure 17 P-R curves and Pd-Fa curves of the reference algorithms

5 数据价值

本数据集针对空中飞行平台对空中运动目标的红外探测应用，通过合理设计方案，构建了共计 200 个图像序列、60000 帧图像及详细标注的大规模数据集。从目标规范角度，所有目标符合尺寸不大于 9×9 像素且 LSNR 不大于 5dB 的弱小目标定义规范；从目标数量角度，包含了单目标、双目标和三目标的情况；从背景类型的角度，涵盖了建筑、植被、云层、水域等 4 种典型场景。该数据集是目前使用像素级掩膜标注的同类型公开发布的数据集中规模最大、涵盖情形最为广泛的数据集，能够促进基于图像分割的红外弱小运动目标检测研究领域的发展。

为满足红外弱小运动目标检测算法测试评估需求，本数据集设置了实例级和像素级两类评价指标，前者包括精确率、回归率和 F_1 指标，后者包括 IoU 和虚警率，并公布了测试代码和展示了一些高性能算法的评价结果。本数据集的发布，为复杂背景下红外弱小运动目标检测研究提供了丰富的数据测试对象，对于加速相关研究进程、提升算法研究水平、促进理论研究成果工程化应用等，具有重要意义。

致 谢

在数据集采集、整理、评审和发布等过程中，得到了课题组多位老师和学生的指导与帮助，在此表示衷心的感谢。

数据作者分工职责

冯子成（1992—），男，山东省潍坊市人，博士研究生，研究方向为目标检测与跟踪。主要承担工作：数据集方案制定，数据集制作，数据集使用评价。

张文龙（1991—），男，陕西省铜川市人，博士，讲师，研究方向为目标检测与跟踪。主要承担工作：数据集方案制定，数据集使用评价。

孙晓亮（1989—），男，安徽省亳州市人，博士，副研究员，研究方向为图像测量与视觉导航。主要承担工作：数据集方案制定，数据集发布。

郭良超（1995—），男，河南省三门峡市人，博士研究生，研究方向为红外弱小目标检测。主要承担工作：数据集使用评价，数据集发布。

刘冬辉（1999—），男，山东省烟台市人，硕士研究生，研究方向为红外弱小目标检测。主要承担工作：数据集使用评价，数据集发布。

参考文献

- [1] 樊香所. 序列图像弱小目标检测与跟踪算法研究[D]. 成都: 电子科技大学, 2019. [FAN X S. Research on dim and small targets detection and tracking algorithms in sequence image[D]. Chengdu: University of Electronic Science and Technology, 2019.]
- [2] 回丙伟, 宋志勇, 范红旗, 等. 地/空背景下红外图像弱小飞机目标检测跟踪数据集[J/OL]. 中国科学数据, 2020, 5(3): 286 – 297. DOI:10.11922/csdata.2019.0074.zh. [HUI B W, SONG Z Y, FAN H Q, et al. A dataset for infrared detection and tracking of dim-small aircraft targets under ground/air background[J/OL]. China Scientific Data, 2020, 5(3): 286 – 297. DOI:10.11922/csdata.2019.0074.zh.]
- [3] 孙晓亮, 郭良超, 张文龙, 等. 复杂背景下红外弱小运动目标检测半仿真数据集[J/OL]. 中国科学数据, 2024, 9(3): 309 – 325. DOI:10.11922/csdata.2021.0015.zh. [SUN X L, GUO L C, ZHANG W L, et al. A dataset of semi-synthetic detection for small infrared moving targets under complex backgrounds[J/OL]. China Scientific Data, 2024, 9(3): 309 – 325. DOI:10.11922/csdata.2021.0015.zh.]
- [4] SUN H, BAI J X, YANG F, et al. Receptive-field and direction induced attention network for infrared dim small target detection with a large-scale dataset IRDST[J]. IEEE Transactions on Geoscience and Remote Sensing, 2023, 61: 5000513. DOI:10.1109/TGRS.2023.3235150.
- [5] KIM J H, HWANG Y. GAN-based synthetic data augmentation for infrared small target detection[J]. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60: 5002512. DOI:10.1109/TGRS.2022.3179891.
- [6] SHI X J, CHEN Z R, WANG H, et al. Convolutional LSTM network[C]//Proceedings of the 29th International Conference on Neural Information Processing Systems - Volume 1. Montreal, Canada. ACM, 2015: 802 – 810. DOI: 10.5555/2969239.2969329.
- [7] ISOLA P, ZHU J Y, ZHOU T H, et al. Image-to-image translation with conditional adversarial networks[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, HI, USA. IEEE, 2017: 5967 – 5976. DOI:10.1109/CVPR.2017.632.
- [8] 张文龙. 动观测平台下运动目标检测方法研究[D]. 长沙: 国防科技大学, 2020. [ZHANG W L. Research on detection method of moving target under moving observation platform[D]. Changsha: National University of Defense Technology, 2020.]
- [9] LIU Q K, LIU R, ZHENG B L, et al. Infrared small target detection with scale and location sensitivity[C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle,

WA, USA. IEEE, 2024: 17490 – 17499. DOI:10.1109/CVPR52733.2024.01656.

[10] LI B Y, XIAO C, WANG L G, et al. Dense nested attention network for infrared small target detection[J]. IEEE Transactions on Image Processing, 2023, 32: 1745 – 1758. DOI:10.1109/TIP.2022.3199107.

[11] HOU Q Y, ZHANG L W, TAN F J, et al. ISTDU-Net: infrared small-target detection U-net[J]. IEEE Geoscience and Remote Sensing Letters, 2022, 19: 7506205. DOI:10.1109/LGRS.2022.3141584.

[12] YUAN S, QIN H L, YAN X, et al. SCTransNet: spatial-channel cross transformer network for infrared small target detection[J]. IEEE Transactions on Geoscience and Remote Sensing, 2024, 62: 5002615. DOI:10.1109/TGRS.2024.3383649.

[13] CHEN S J, JI L P, ZHU J W, et al. SSTNet: sliced spatio-temporal network with cross-slice ConvLSTM for moving infrared dim-small target detection[J]. IEEE Transactions on Geoscience and Remote Sensing, 2024, 62: 5000912. DOI:10.1109/TGRS.2024.3350024.

[14] LYU Y Y, LIU Z G, LI H D, et al. A real-time and lightweight method for tiny airborne object detection[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Vancouver, BC, Canada. IEEE, 2023: 3016 – 3025. DOI:10.1109/CVPRW59228.2023.00303.

[15] DUAN W W, JI L P, CHEN S J, et al. Triple-domain feature learning with frequency-aware memory enhancement for moving infrared small target detection[J]. IEEE Transactions on Geoscience and Remote Sensing, 2024, 62: 5006014. DOI:10.1109/TGRS.2024.3452175.

论文引用格式

冯子成, 张文龙, 孙晓亮, 等. 红外弱小运动目标检测与分割半仿真数据集[J/OL]. 中国科学数据, 2025, 10(4). (2025-09-22). DOI: 10.11922/11-6035.csd.2024.0196.zh.

数据引用格式

冯子成, 张文龙, 孙晓亮, 等. 红外弱小运动目标检测与分割半仿真数据集[DS/OL]. V1. Science Data Bank, 2025. (2025-02-11). DOI: 10.57760/sciencedb.j00001.01190.

A dataset of semi-synthetic infrared dim small moving targets for detection and segmentation

FENG Zicheng^{1,2}, ZHANG Wenlong^{1,2*}, SUN Xiaoliang^{1,2},
GUO Liangchao^{1,2}, LIU Donghui^{1,2}

1. College of Aerospace Science and Engineering, National University of Defense Technology, Changsha, 410073, China

2. Hunan Provincial Key Laboratory of Image Measurement and Vision Navigation, Changsha, 410073, China

Abstract: Infrared dim small moving target detection is a research hotspot in the field of target perception,

with wide applications in security monitoring, reconnaissance and early warning. Recently, infrared dim small target detection algorithms based on deep learning have been developed rapidly, placing higher demands on the scale and quality of datasets. However, it is difficult to obtain a large amount of real data with fine-grained annotations. Existing datasets either synthesize targets using Gaussian distribution or sample frames to reduce the workload, resulting in unrealistic targets and discontinuous motion trajectories. To address these issues, we proposed a generative adversarial network with spatio-temporal feature fusion to learn the appearance variation of real targets and generate highly simulated infrared targets. Then, by combining diverse scenes and various target motion trajectories from existing datasets, we finally constructed a large-scale semi-synthetic dataset for infrared dim small moving target detection and segmentation under clutter backgrounds. The generated targets range in size from 4×4 to 9×9 pixels with a local signal-to-noise ratio of -5 to 5 dB, conforming to the definition of dim and small targets. The dataset covers four scenes (building, vegetation, cloud and water), three cases (one, two and three targets), and provides three annotations (center point, bounding box and mask), as well as two evaluation metrics (instance-level and pixel-level). The dataset contains 200 image sequences, 60,000 images, equally divided into a training set and a test set. It can be used in infrared dim small moving target detection, segmentation and tracking research under complex backgrounds.

Keywords: infrared dim small moving target; clutter background; target detection; target segmentation; image sequence.

Dataset profile

Title	A dataset of semi-synthetic infrared dim small moving targets for detection and segmentation
Data authors	FENG Zicheng, ZHANG Wenlong, SUN Xiaoliang, GUO Liangchao, LIU Donghui
Data corresponding author	ZHANG Wenlong (wenlong@nudt.edu.cn)
Time range	2020 – 2021
Geographical scope	China
Data Volume	~10 GB
Data formate	*.png, *.json
Data service system	https://doi.org/10.57760/sciencedb.j00001.01190
Source of funding	National Natural Science Foundation of China (12202485)
Dataset composition	The dataset contains 200 image sequences, 60,000 images, equally divided into a training set and a test set of 100 sequences each. Each sequence is stored in a dedicated folder containing image sequence files in PNG format. Annotations are categorized into two folders: one contains pixel-level mask annotation files in PNG format, named identically to the corresponding image sequence files. The other contains instance-level files in JSON format, named after the image sequence names. These JSON files include attribute information such as target size, relative speed, and local signal-to-noise ratio, as well as the number, coordinates, and sizes of targets in each frame.

