引用格式：武哲纬, 余瑞龙, 刘启和, 等. 基于特征正则对抗训练的视觉跟踪对抗鲁棒性提升方法 [J]. 电子科技大学学报, 2026, 55(2): 252-262.
WU Z W, YU R L, LIU Q H, et al. Improving adversarial robustness of visual trackers via feature regularized adversarial training[J]. Journal of University of Electronic Science and Technology of China, 2026, 55(2): 252-262.

基于特征正则对抗训练的视觉跟踪对抗鲁棒性提升方法

武哲纬[1]，余瑞龙[1]，刘启和[1*]，吴春江[2]，叶　飞[1]，周世杰[1]

（1. 电子科技大学信息与软件工程学院，成都 610054；2. 成都信息工程大学软件工程学院，成都 610225）

摘要：该文首先基于对抗特征与原始特征在不同卷积尺度下产生分离这一经验观察，提出了特征正则化损失以使二者在特征空间中实现对齐。其次，针对目标跟踪任务的双图像输入特性，创新性地设计了适配深度跟踪网络的对抗训练框架。该框架利用特征正则化损失指导对抗样本生成，有效引导网络在对抗训练中学习鲁棒特征表示，从而显著提升目标跟踪模型的对抗鲁棒性。最后，在公开数据集上的对比实验证明，提出的方法能够在自适应攻击场景下获得最优的性能，同时能够在异构跟踪器间实现有效迁移，且在干净样本上保持有限的精度损失。

关键词：视觉单目标跟踪；对抗防御技术；对抗训练；计算机视觉

中图分类号：TP37　　**文献标志码：**A　　DOI: 10.12178/1001-0548.2024351

Improving adversarial robustness of visual trackers via feature regularized adversarial training

WU Zhewei[1], YU Ruilong[1], LIU Qihe[1*], WU Chunjiang[2], YE Fei[1], and ZHOU Shijie[1]

(1. School of Information and Software Engineering, University of Electronic Science and Technology of China, Chengdu 610054, China;
2. School of Software Engineering, Chengdu University of Information Technology, Chengdu 610225, China)

Abstract: Visual object tracking (VOT), as a crucial downstream task in computer vision, has consistently garnered significant attention due to its widespread applications. In recent years, adversarial attack methods for VOT have emerged, which disrupt tracker predictions by injecting adversarial perturbations into input data. However, corresponding adversarial defense approaches remain scarce and suffer from multiple limitations: inadequate defense performance against adaptive attacks, excessive computational overhead introduced by preprocessing modules, and poor transferability across heterogeneous trackers. To address these challenges, this paper proposes a feature regularization loss based on the empirical observation that adversarial features and original features exhibit divergence across different convolutional scales, aiming to achieve feature space alignment between them. Second, considering the dual-image input characteristics of visual tracking tasks, an adversarial training framework tailored for visual tracker is designed. This framework effectively guides the network to learn robust feature representations by leveraging the feature regularization loss, thereby enhancing the adversarial robustness of the tracker. Finally, comparative experiments on public benchmarks demonstrate that our method achieves state-of-the-art performance under adaptive attack scenarios while maintaining limited accuracy degradation on clean samples. Notably, the proposed approach exhibits superior transferability across heterogeneous tracking architectures compared to existing defense methods.

Key words: visual single object tracking; adversarial defense; adversarial training; computer vison

视觉目标跟踪（visual object tracking, VOT）依据在首帧中指定的目标形态，在后续到达的连续图像序列中对目标位置进行预测。作为计算机视觉领域的一项重要下游任务，视觉目标跟踪一直以来

收稿日期：2024−12−24
基金项目：国家自然科学基金（62272089）；四川省自然科学基金（2025ZNSFSC0510）；厅市共建智能终端四川省重点实验室开放课题（SCITLAB-30003）
作者简介：武哲纬，博士生，主要从事对抗学习、计算机视觉方面的研究。
*通信作者 E-mail: qiheliu@uestc.edu.cn

都受到关注，并且在生活、工业等方面应用广泛[1-3]。同时，关于对抗样本 （adversarial sample）的研究证明了在输入图像中添加人眼不可察觉的扰动能够使深度神经网络发生重大预测错误[4]。因此，近年来许多研究人员开始关注针对目标跟踪任务的对抗攻击方法。这些方法根据精心设计的目标函数指导网络模型[5-6] 或迭代方法[7-9] 生成人眼不可察觉的对抗扰动，此类扰动能够使视觉跟踪器在正常工作过程中丢失跟踪目标。如何增强视觉跟踪器的对抗鲁棒性是当前迫切需要解决的问题。

可是，当前针对目标跟踪任务的对抗防御方法的研究却仍然相对有限。如 AADN[10] 利用分类损失与回归损失训练了部署在跟踪器前方的附加防御网络，该网络通过对输入图像开展防御性变换以消除图像中含有的潜在对抗扰动。LRR[11] 利用来自历史帧中的时序信息与描述被跟踪物体类别的自然语言信息训练了一个对输入图像进行重建的预处理网络。同时，通过文献调研我们发现，现有方法均基于输入预处理的防御范式，然而此类方法却存在诸多缺陷。首先，基于预处理网络的防御方法在面对自适应攻击（adaptive attack）时会出现较大程度的防御性能下降，这严重制约了此类方法的现实应用[12-13]。其次，由于引入预处理网络将不可避免地增加网络计算开销，故 AADN 和 LRR 将分别导致跟踪器的工作速率发生 57% 与 243% 的降低，因此基于预处理网络的防御方法难以满足延迟敏感应用场景的部署需求。最后，在现有的防御策略中，对抗训练被证明是应对对抗样本最有效的防御方法[14]，并在图像分类等任务上已经拥有成熟的研究脉络，但由于目标跟踪任务与图像分类任务在输入形式、工作原理与网络结构上存在诸多差异，现有的对抗训练方法难以直接迁移至目标跟踪任务[10,15]。

针对现有防御方法存在的上述问题，本文针对目标跟踪任务提出了一种基于特征正则化的对抗训练方法（feature regularized adversarial training，FRAT）。首先，我们观察到随着骨干网络卷积层数的加深，被对抗扰动干扰后的潜在特征（下简称对抗特征）与未被扰动的潜在特征（下简称原始特征）会在更深的卷积层中产生更大的分离。基于这一实践观察，我们为原始跟踪损失设计了特征正则化项，以使对抗特征能够分别在浅、中、深层的特征空间中与原始特征实现对齐。其次，结合目标跟踪任务将双图像作为输入的显著特性，为目标跟踪网络设计了包含模板图像扰动生成与搜索图像扰动生成的对抗训练流程。对抗训练能够根据训练时的梯度信息，充分探索训练样本邻域空间内存在的潜在对抗样本，以构筑对抗鲁棒性更强的跟踪网络。特别地，本文提出的方法不会改变跟踪器的网络结构，故在完成训练后不会引入任何额外的计算开销。

1　相关工作

1.1　视觉目标跟踪任务

孪生神经网络首先被 SiamFC[16] 应用于目标跟踪任务，其结构灵活且具有出色的精度与效率，因此在后续相当长的时间内奠定了目标跟踪任务的基本范式。SiamRPN[17] 首次将区域建议网络（region proposal network, RPN）引入跟踪任务，利用 RPN 网络生成一组包含对目标潜在位置信息的置信度提案，并通过非极大值抑制方法获取最优的预测结果。SiamRPN++[18] 调整与融合了搜索区域与模板区域的多尺度特征来形成对二者的多尺度特征表示，并且首次提出了深度互相关方法以实现更加准确的特征匹配，该方法常被作为基于孪生神经网络的基准跟踪方法。SiamMask[19] 设置了独立的分支以完成对目标的实例化分割任务。在这之后，研究人员开始研发基于 Transformer 的跟踪方法，该类方法利用自注意力机制代替卷积网络以实现特征提取与特征匹配，同样取得了良好的跟踪精度。TransT[20] 提出了基于自注意力的上下文增强模块和基于交叉注意力的特征增强模块，可以自适应地关注有用信息，并建立长距离特征之间的关联。SiamTPN[21] 利用轻量级骨干网络的多尺度特征，结合 Transformer 构造了对目标外观的鲁棒性表示，并通过池化注意力模块降低计算开销。总体而言，基于 Transformer 的跟踪器在跟踪精度上优于基于孪生网络的跟踪器，但跟踪效率较低，诸多研究正着力解决该问题。

1.2　目标跟踪任务上的对抗防御技术

研究者针对对抗样本[4] 生成与防御进行了大量研究，但绝大多数都聚焦在图像分类这一基础视觉任务上，而针对目标跟踪任务的对抗防御技术目前仍然缺乏研究。较之于图像分类任务，视觉目标跟踪的对抗防御面临以下关键差异。首先，图像分类模型以单张静态图像为输入，而视觉跟踪器需同时处理搜索区域与模板区域两类输入，并需在连续帧序列中动态预测目标状态。除了保护帧内空间信息不受扰动，视觉目标跟踪的对抗防御技术还需要考

虑对帧间时序信息开展防护。其次，主流视觉跟踪器（如基于孪生网络或 Transformer）通常包含多阶段特征匹配与融合模块，其网络复杂度远超传统分类模型。此外，为适应目标外观的动态变化，跟踪器常采用参数在线更新策略，故静态防御方法需结合在线更新过程进行动态适配，以抵御持续攻击。最后，视觉跟踪任务需满足实时处理需求，而现有基于预处理的对抗防御技术常因计算开销过高难以部署，故目标跟踪任务的对抗防御框架不应在推理时引入显著的计算开销。

AADN[10] 首次对该技术的难点进行了系统性阐述，并且提出了一个被设置在跟踪器孪生骨干网络前的预处理防御网络，它以即插即用的方式适配多种视觉跟踪器。AADN 利用强化目标粗定位的前后景分类损失与细化目标精确位置的回归损失在训练过程中的梯度信息生成对抗样本，以提高网络的对抗鲁棒性。LRR[11] 同样遵循预处理防御网络的形式，利用被扰动像素在与其时空相邻的像素间具有相似外观这一先验知识，建立了时空隐式表示机制。同时，LRR 预定义了一组描述跟踪模板潜在类别的自然语言集合，通过跨模态信息指导预处理网络对输入图像进行重建。RTAA[22] 首先提出了一个迭代式攻击方法，该方法利用带有错误信息的伪真值标签生成对抗样本，对抗样本将会最大化跟踪器的分类损失与回归损失，最终导致攻击失效。在这之后，RTAA 基于相似的先验策略提出了对应的防御方案以估计输入图像中含有的对抗扰动。CMDN[23] 利用掩码图像建模（masked image modeling）对输入图像进行随机遮蔽再进行重建以消除对抗扰动。在训练过程中，CMDN 采用特征损失与感知损失来指导重建网络的生成图像具有与干净样本相似的特征表示与视觉表示。

但现有的为目标跟踪任务开发的对抗防御方法仍然存在诸多问题。首先，它们大多是基于预处理的防御方法，此类方法在自适应攻击的场景下性能下降明显。其次，预处理网络的引入将不可避免地降低跟踪过程的处理时延，这对于时延敏感的应用场景而言是难以接受的。为了解决这些问题，本文提出了基于特征正则对抗训练方法 FRAT。FRAT 通过特征正则化项对对抗样本与干净样本在特征空间中添加对齐约束，并且在对抗训练过程中分别为搜索区域与模板图像生成对抗扰动。在训练完成后，跟踪网络无需额外的计算开销即可实现对抗鲁棒性的提升。

2 特征正则化的对抗训练方法

选择 SiamRPN++网络作为基准模型，FRAT 的训练流程如图 1 所示。

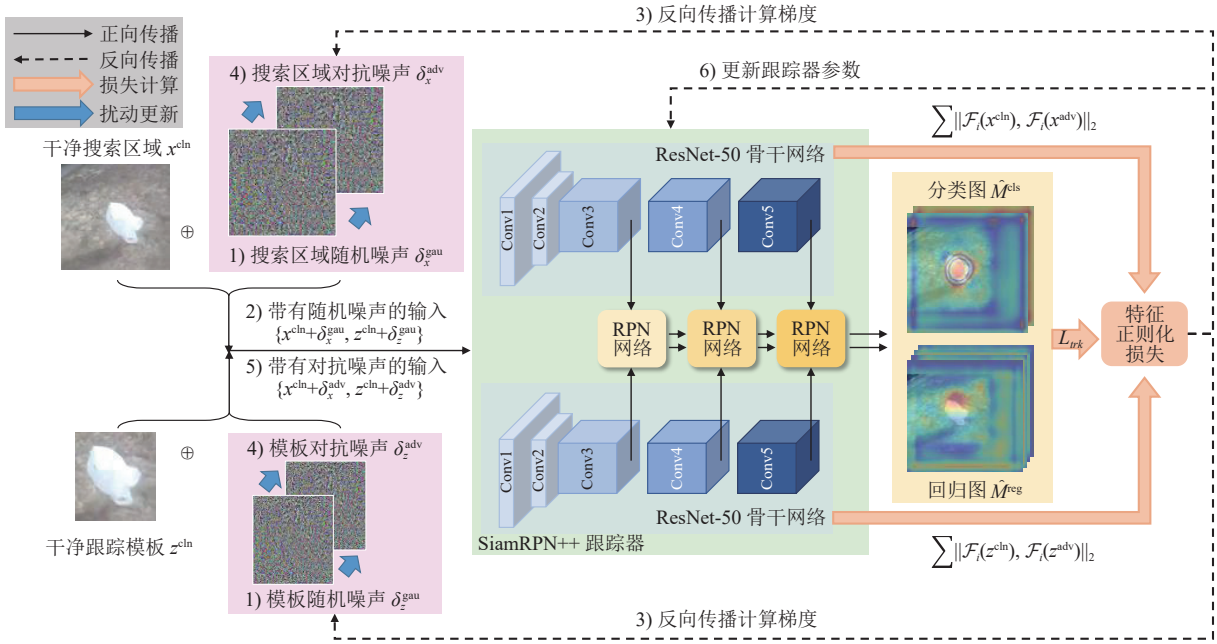


图 1 特征正则化的对抗训练方法流程图

2.1 符号约定

令 $\boldsymbol{I}_i(i\in 1,2,3\cdots)$表示输入跟踪器的包含$n$张连续图像的序列。令$\boldsymbol{b}_i^{gt}(i\in 1,2,3\cdots)$表示在第$i$帧图像中被跟踪目标的真实位置边框，其中$\boldsymbol{b}=\{b_x,b_y,$

$b_w, b_h\}$分别表示边框相对于图像左上角的坐标以及宽高。令$\{\boldsymbol{y}_i^{\text{cls}}, \boldsymbol{y}_i^{\text{reg}}\}$代表在图像$\boldsymbol{I}_i$中与$\boldsymbol{b}_i^{\text{gt}}$对应的分类标签与回归标签。

基于孪生网络的目标跟踪流程通常可以被概括如下：前处理过程首先通过真实边框$\boldsymbol{b}_1^{\text{gt}}$在初始帧$\boldsymbol{I}_1$中裁剪得到跟踪模板图像$\boldsymbol{z}$，并且根据前序帧跟踪结果$\boldsymbol{b}_{i-1}^{\text{pre}}(i \in 2,3\cdots n)$从当前帧$\boldsymbol{I}_i(i \in 2,3,\cdots,n)$中裁剪得到搜索区域图像$\boldsymbol{x}_i$；随后，将样本对$\{\boldsymbol{z}, \boldsymbol{x}_i\}$输入至参数为$\theta$的跟踪网络$f(\cdot,\theta)$。网络将输出含有表征目标位置信息的分类图$\boldsymbol{M}_i^{\text{cls}}$与回归图$\boldsymbol{M}_i^{\text{reg}}$：

$$\left\{\boldsymbol{M}_i^{\text{cls}}, \boldsymbol{M}_i^{\text{reg}}\right\} = f(\boldsymbol{z}, \boldsymbol{x}_i, \theta) \tag{1}$$

在后处理过程中，跟踪器根据$\left\{\boldsymbol{M}_i^{\text{cls}}, \boldsymbol{M}_i^{\text{reg}}\right\}$计算目标在$\boldsymbol{I}_i$中的位置预测边框$\boldsymbol{b}_i^{\text{pre}}$。需要注意的是，前处理与后处理过程并不涉及网络模型的前向传播，仅需结合超参数进行图像处理与数值计算。

基于前述定义，我们的目标是得到具有鲁棒性参数$\hat{\theta}$的跟踪器，它能够对带有对抗扰动的图像输入保持准确的预测，即具有最小化的预测损失：

$$\hat{\theta} = \arg\min_{\theta} L\left(f\left(\boldsymbol{z}^{\text{adv}}, \boldsymbol{x}^{\text{adv}}, \theta\right), \left\{\boldsymbol{y}^{\text{cls}}, \boldsymbol{y}^{\text{reg}}\right\}\right) \tag{2}$$

式中，$L(\cdot)$代表评估跟踪预测结果与标签相似程度的损失函数；$\left\{\boldsymbol{z}^{\text{adv}}, \boldsymbol{x}^{\text{adv}}\right\}$代表输入的可能带有对抗扰动的模板图像与搜索区域图像。

2.2　特征正则化损失

设计特征正则化损失的动机来源于对不同卷积尺度下的对抗特征与原始特征将会产生不同程度分离的观察。图 2 展示了对抗特征与干净特征从 ResNet-50 骨干网络的第 3、4、5 个卷积块中提取出的潜在特征，并通过 t-SNE 进行可视化降维以观测二者的分离情况。来自 3 个卷积块的结果分别被展示在图 2 中。如图 2a 所示，从浅层卷积块中提取的特征在经过对抗扰动后，其与原始特征产生较小的分离，进而出现明显的“点对”情况。随着卷积层的进一步加深，二者的分离情况发生了进一步加剧。在图 2c 中，对抗特征与原始特征在深层卷积输出中产生了显著的分离，甚至出现了明确的边界。这表明带有对抗扰动的输入图像随着卷积层的加深，其引入的干扰将会被逐渐放大，并最终导致重大的预测错误。其内在机制可归因于卷积操作矩阵在将输入对抗扰动的映射至高维空间时，其误差将在线性权重矩阵乘法操作中被放大[4]，以及 ReLU 等激活函数对正负扰动的不对称响应特性[24]导致部分通道的扰动被选择性保留甚至增强。此外，深层网络的级联非线性操作通过雅可比矩阵的谱范数扩张[25]，进一步扭曲了特征空间对拓扑结构，使得初始微小的对抗扰动演变为不可忽视的语义特征畸变。为了解决这个问题，我们设计了特征正则化损失L，它由原始跟踪损失L_{trk}与特征正则化项L_{fea}组成。首先，L_{trk}的各个子项对跟踪网络的子结构进行了针对性的设计，能够在训练过程中指导跟踪器进行参数更新并最终实现准确的预测，L_{fea}分别对来自跟踪器骨干网络浅、中、深层的潜在特征添加约束，使各个对抗特征分别在各自的尺度贴近原始特征，为后续的特征匹配与预测流程提供准确的特征表示。其次，L也用于在对抗训练过程中为每一个训练批次（training batch）中针对跟踪器的不同子结构生成对抗样本，以充分提升跟踪网络整体的防御鲁棒性，此部分将在 2.3 节详细说明。

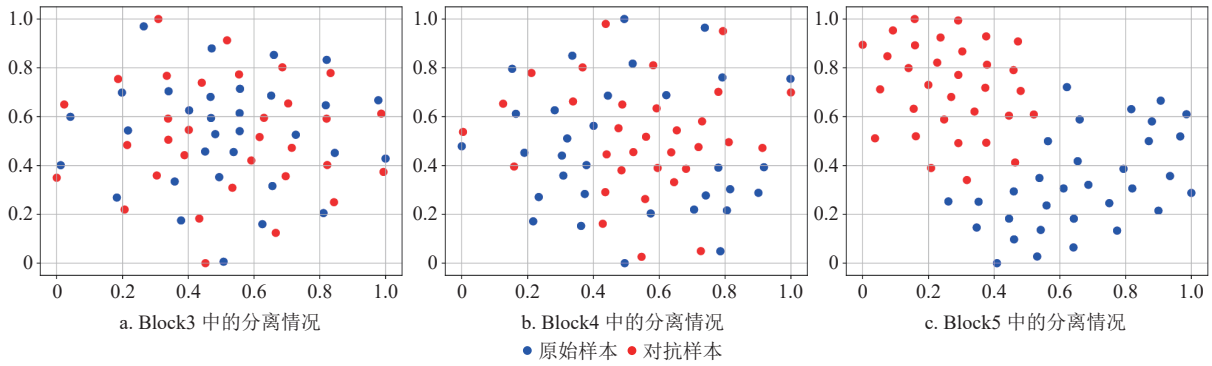


图 2　对抗特征与原始特征在 ResNet50 网络不同卷积尺度下的分离情况

综上所述，L可以写：

$$L = L_{\text{fea}} + L_{\text{trk}} \tag{3}$$

式中，L_{fea}为：

$$L_{\text{fea}} = \sum_{i}^{\{3,4,5\}} \left[\left\| \mathcal{F}_i\left(\boldsymbol{x}^{\text{cln}}\right), \mathcal{F}_i\left(\boldsymbol{x}^{\text{adv}}\right) \right\|_2 + \left\| \mathcal{F}_i\left(\boldsymbol{z}^{\text{cln}}\right), \mathcal{F}_i\left(\boldsymbol{z}^{\text{adv}}\right) \right\|_2 \right] \tag{4}$$

式中，$\mathcal{F}_i(\cdot)$代表来自第i个卷积块的潜在特征；

$\boldsymbol{x}^{\text{cln}}$代表干净搜索区域；$\boldsymbol{x}^{\text{adv}}$代表含有对抗扰动的搜索区域；$\boldsymbol{z}^{\text{cln}}$代表干净跟踪模板；$\boldsymbol{z}^{\text{adv}}$代表含有对抗扰动的跟踪模板。以 SiamRPN++为例，分别对来自第 3、4、5 个卷积块的对抗特征与干净特征进行对齐，因为它们在后续的特征匹配阶段分别用于获取不同尺度下的目标位置响应。同理，当L_{fea}应用于其他跟踪器时，亦可通过对特征匹配阶段所采用的多尺度潜在特征进行对齐。

式 (3) 中的L_{trk}为跟踪器所采用的原始训练损失。以 SiamRPN++为例，L_{trk}由强化目标粗定位的前后景分类损失L_{cls}与细化目标精确位置的回归损失L_{reg}组成。前者通过交叉熵损失计算，用于评估跟踪器分类分支的准确性；后者通过平滑L_1损失计算，用于评估跟踪器回归分支的准确性，可以写作：

$$L_{\text{reg}}=\sum_{t}\text{Smooth}L_1(d_t,\sigma)\quad t\in\{x,y,w,h\} \tag{5}$$

式中，σ为超参数阈值；d_t表示预测边框$\boldsymbol{b}^{\text{pre}}$与真实边框$\boldsymbol{b}^{\text{gt}}$各个分量之间的归一化距离，可以分别通过式 (6)～式 (9) 计算：

$$d_x=\frac{b_x^{\text{gt}}-b_x^{\text{pre}}}{b_x^{\text{pre}}} \tag{6}$$

$$d_y=\frac{b_y^{\text{gt}}-b_y^{\text{pre}}}{b_y^{\text{pre}}} \tag{7}$$

$$d_w=\ln\frac{b_w^{\text{gt}}}{b_w^{\text{pre}}} \tag{8}$$

$$d_h=\ln\frac{b_h^{\text{gt}}}{b_h^{\text{pre}}} \tag{9}$$

2.3 特征正则化的对抗训练

如前所述，特征正则化损失L用于在训练过程中生成同时具有攻击跟踪器分类分支、回归分支与骨干网络的对抗扰动。通过将生成的对抗样本拓展至训练集，可以让跟踪网络在训练过程中具备对对抗样本的鲁棒性。由于跟踪器的输入为跟踪模板与搜索区域组成的图像对，因此需要为两张图像分别生成对抗扰动。故可以定义如式 (10) 的最小−最大化问题来描述对抗训练的目标：

$$\min_{\theta}\max_{\substack{\|\boldsymbol{\delta}_{\boldsymbol{x}}\|_2<\varepsilon\\ \|\boldsymbol{\delta}_{\boldsymbol{z}}\|_2<\varepsilon}}L\left(f\left(\boldsymbol{z}^{\text{cln}}+\boldsymbol{\delta}_{\boldsymbol{z}},\boldsymbol{x}^{\text{cln}}+\boldsymbol{\delta}_{\boldsymbol{x}},\theta\right),\left\{\boldsymbol{y}^{\text{cls}},\boldsymbol{y}^{\text{reg}}\right\}\right) \tag{10}$$

式中，$\boldsymbol{\delta}_{\boldsymbol{z}}$与$\boldsymbol{\delta}_{\boldsymbol{x}}$代表在对抗训练过程中被添加至干净跟踪模板$\boldsymbol{z}^{\text{cln}}$与干净搜索区域$\boldsymbol{x}^{\text{cln}}$中的训练扰动，而$\|\cdot\|_2<\varepsilon$表示将它们分别约束在以干净样本为圆心，扰动预算$\varepsilon$为半径的$l_2$球内，以保证扰动的不可察觉性。

式 (10) 内部的最大化问题是为被训练模型找到导致最坏输出的对抗扰动，外部的最小化问题是训练一个对对抗扰动鲁棒的模型参数θ。每个批次的训练数据将会在网络中前向传播两次，两次前向传播计算得到的特征正则化损失L将分别用于更新训练扰动和网络参数。

具体而言，在第一次前向传播中，首先将带有随机噪声$\boldsymbol{\delta}_{\boldsymbol{z}}^{\text{gau}}$与$\boldsymbol{\delta}_{\boldsymbol{x}}^{\text{gau}}$的跟踪模板与搜索区域中输入至参数为$\theta$的跟踪网络$f(\cdot,\theta)$。网络将输出分类图与回归图$\left\{\hat{\boldsymbol{M}}^{\text{cls}},\hat{\boldsymbol{M}}^{\text{reg}}\right\}$，二者将分别与标签$\left\{\boldsymbol{y}^{\text{cls}},\boldsymbol{y}^{\text{reg}}\right\}$参照式 (3) 计算特征正则化损失$L$。此时可以根据$L$使用单步 FGSM 策略将$\boldsymbol{\delta}^{\text{gau}}$更新对抗扰动$\boldsymbol{\delta}^{\text{adv}}$。以搜索区域扰动更新举例，更新方式为：

$$\boldsymbol{\delta}_{\boldsymbol{x}}^{\text{adv}}=\text{clip}\left(\boldsymbol{\delta}_{\boldsymbol{x}}^{\text{gau}}+\varepsilon\text{sign}\left(\nabla_{\boldsymbol{x}^{\text{cln}}}L\right),\varepsilon,-\varepsilon\right) \tag{11}$$

跟踪模板扰动更新同理。

在第二次前向传播中，将带有对抗扰动的跟踪模板$\boldsymbol{z}^{\text{cln}}+\boldsymbol{\delta}_{\boldsymbol{z}}^{\text{adv}}$与搜索区域$\boldsymbol{x}^{\text{cln}}+\boldsymbol{\delta}_{\boldsymbol{x}}^{\text{adv}}$输入$f(\cdot,\theta)$，以同样的方式再次计算$L$。最后根据此时的梯度信息更新跟踪网络的参数。

综上所述，完整的 FRAT 流程总结如下。

算法 1　特征正则化的对抗训练流程

输入：初始化的跟踪网络参数θ, 训练集D, 训练轮次T, 批大小S, 学习率γ, 扰动预算ε

输出：优化后的跟踪网络参数$\hat{\theta}$

For i in range $[1,T]$ do:

For random batch $\left\{\boldsymbol{x},\boldsymbol{z},\boldsymbol{y}^{\text{cls}},\boldsymbol{y}^{\text{reg}}\right\}\in D$ do:

初始化训练扰动为高斯扰动

$$\boldsymbol{\delta}_{\boldsymbol{z}}^{\text{gau}}\sim U[-\varepsilon,\varepsilon],\boldsymbol{\delta}_{\boldsymbol{x}}^{\text{gau}}\sim U[-\varepsilon,\varepsilon]$$

第一次前向传播

$$\left\{\hat{\boldsymbol{M}}^{\text{cls}},\hat{\boldsymbol{M}}^{\text{reg}}\right\}=f\left(\boldsymbol{z}^{\text{cln}}+\boldsymbol{\delta}_{\boldsymbol{z}}^{\text{gau}},\boldsymbol{x}^{\text{cln}}+\boldsymbol{\delta}_{\boldsymbol{x}}^{\text{gau}},\theta\right)$$

参照式 (3) 计算特征正则化损失L

参照式 (11) 更新高斯扰动$\boldsymbol{\delta}_{\boldsymbol{z}}^{\text{gau}}$与$\boldsymbol{\delta}_{\boldsymbol{x}}^{\text{gau}}$为对抗扰动$\boldsymbol{\delta}_{\boldsymbol{z}}^{\text{adv}}$与$\boldsymbol{\delta}_{\boldsymbol{x}}^{\text{adv}}$

第二次前向传播$\left\{\hat{\boldsymbol{M}}^{\text{cls}},\hat{\boldsymbol{M}}^{\text{reg}}\right\}=f\left(\boldsymbol{z}^{\text{cln}}+\boldsymbol{\delta}_{\boldsymbol{z}}^{\text{adv}},\boldsymbol{x}^{\text{cln}}+\boldsymbol{\delta}_{\boldsymbol{x}}^{\text{adv}},\theta\right)$

参照式 (3) 依据第二次前向传播的结果重新计算特征正则化损失L

利用 SGD 优化器更新网络参数

End For

End For

Return $\hat{\theta}$

同时，图 3 展示了 FRAT 在面对对抗扰动时的对抗鲁棒性，证明了经过 FRAT 训练的 SiamRPN++跟踪器能够有效缓解因对抗扰动导致的跟踪失败现象。其中，绿色、红色、蓝色的边框分别代表 SiamRPN++跟踪器在干净样本上、被 CSA 方法攻击后与被 FRAT 防御后的预测边框。

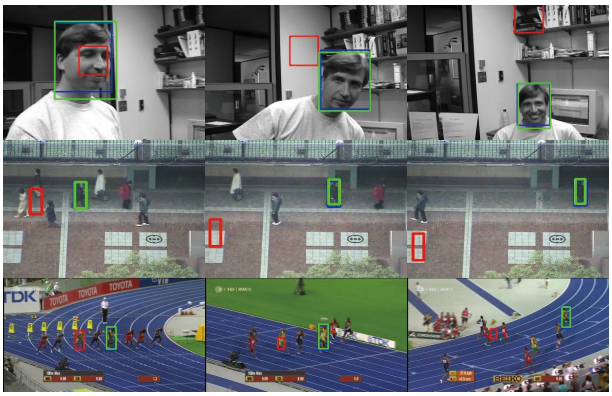


图 3　FRAT 可视化防御结果

3　实验结果与分析

3.1　实验设置

本文实验的训练环境为 Dell R750xa 平台，搭载 2× Intel Xeon 4310 2.1 GHz、3× Nvidia Tesla A100 80 GB，运行内存 256 GB。实验采用的基准跟踪器为 SiamRPN++跟踪器，同时通过端到端的训练方式训练跟踪网络，采用 SGD 优化器，整个训练过程包含 20 个轮次（Iteration），批大小（Batchsize）设置为 16，训练时学习率动态变化，范围为 [0.005,0.001] 以避开陷入局部最优状态并最大化训练收益。扰动预算ϵ 设置为 8/255，模板图像与搜索图像的尺寸分别为 127×127 与 255×255 像素。

3.2　训练与测试数据集

本文在训练阶段采用了 ImageNet VID 和 COCO 数据集。在测试阶段，本文在 3 个广泛使用的目标跟踪测试数据集上开展量化评估，包括 OTB100[26]、VOT2018[27] 与 LaSOT[28] 数据集。OTB100 数据集包含 100 个来自日常生活场景的高难度序列，并标注了 9 个属性来表示特定的高难度场景，包括光照变化、尺度变化、遮挡、变形、运动模糊、快速运动、平面内旋转等。VOT2018 是另一个具有挑战性的跟踪基准，它同时测量跟踪器的准确性和鲁棒性。该基准包括 60 个视频，采用期望平均覆盖面积对跟踪器的性能进行排名。LaSOT 是最近发布的大规模单目标跟踪基准，共包含 280 个跟踪序列，其中包含数字挑战场景和 70 多类目标，每帧都有高质量的人工标注。

3.3　评价指标

在本文的测试数据集中，主要通过计算成功得分（success score, Succ.）、精度得分（precision score, Prec.）、归一化精度得分（normalized precision score, N.Prec.）、期望平均覆盖面积（expected average overlap, EAO）与鲁棒性（robustness, R）以评估跟踪器的性能表现。

成功得分[26, 28] 表示在给定阈值T的情况下，计算模型预测位置边框$\boldsymbol{b}^{\mathrm{pre}}$ 与目标真实位置边框$\boldsymbol{b}^{\mathrm{gt}}$之间的交并比 (intersection over union, IoU)。若得分高于T，则该帧计为跟踪成功；否则为失败。以此计算在测试集中所有图像帧的跟踪成功率。通过计算自变量T关于成功率的曲线与坐标轴的曲线下面积（area under curve, AUC），则可得到成功率得分。对于包含N帧的测试集，成功得分的计算方法为：

$$\mathrm{Succ.}=\frac{1}{N}\int_0^1\sum_{i=1}^{N}\left|\frac{\mathrm{IoU}\left(\boldsymbol{b}_i^{\mathrm{pre}},\boldsymbol{b}_i^{\mathrm{gt}}\right)}{T}\right|\mathrm{d}T \tag{12}$$

精度得分[26] 的计算方式与成功得分一致，仅将交并比更改为计算$\boldsymbol{b}^{\mathrm{pre}}$与$\boldsymbol{b}^{\mathrm{gt}}$的中心位置误差（center location error, CLE）。类似地，精度得分为：

$$\mathrm{Prec.}=\frac{1}{N}\int_0^1\sum_{i=1}^{N}\left|\frac{\mathrm{CLE}\left(\boldsymbol{b}_i^{\mathrm{pre}},\boldsymbol{b}_i^{\mathrm{gt}}\right)}{T}\right|\mathrm{d}T \tag{13}$$

但式 (13) 中采用的距离度量方法为绝对欧氏距离，此种方式无法克服由于输入图像尺寸不同所带来的误差。为了缓解该问题，LaSOT 数据集将其绝对欧氏距离更改为归一化欧氏距离[28]，并将此指标更改为归一化精度得分，计算方法为：

$$\mathrm{N.Pre.}=\frac{1}{N}\int_0^1\sum_{i=1}^{N}\left|\frac{\mathrm{N.CLE}\left(\boldsymbol{b}_i^{\mathrm{pre}},\boldsymbol{b}_i^{\mathrm{gt}}\right)}{T}\right|\mathrm{d}T \tag{14}$$

期望平均覆盖面积[27] 通过统计整个数据集的视频长度得到长度极值$[N_{\mathrm{lo}},N_{\mathrm{hi}}]$，计算每个视频在$[N_{\mathrm{lo}},N_{\mathrm{hi}}]$区间内所有图像帧中的平均交并比：

$$\mathrm{EAO}=\frac{1}{N_{\mathrm{hi}}-N_{\mathrm{lo}}}\sum_{i=N_{\mathrm{lo}}}^{N_{\mathrm{hi}}}\mathrm{IoU}\left(\boldsymbol{b}_i^{\mathrm{pre}},\boldsymbol{b}_i^{\mathrm{gt}}\right) \tag{15}$$

鲁棒性指标[27]直接统计整个数据集上平均每100帧出现目标丢失的次数，计算公式为：

$$R=\frac{100}{N}\sum_{i=1}^{N}\left|\frac{\mathrm{IoU}\left(\boldsymbol{b}_i^{\mathrm{pre}},\boldsymbol{b}_i^{\mathrm{gt}}\right)}{T}\right| \tag{16}$$

3.4 在对抗攻击下的防御鲁棒性分析

有研究表明针对原始模型的静态攻击不足以评估防御方法的真实效果，而自适应攻击能够根据防御机制的响应和反馈动态调整对抗样本的生成，故此类方法被认为是评估防御方法的重要途径[12−13]。因此，本文进行了在自适应攻击场景下的实验。在自适应攻击的场景下部署了迭代式黑盒攻击方法IoU Attack[7]、生成式白盒攻击方法DFA[6]和生成式黑盒攻击方法CSA[5]，并且和最新的基线目标跟踪防御方法AADN[10]、RTAA[20]、CMDN[21]以及典型的对抗训练方法TRADES[29]、PGD-AT[30]开展了防御性能比较。

具体而言，在自适应攻击场景下，将迭代式攻击方法的查询对象设置为经过FRAT训练的SiamRPN++跟踪器，并基于该网络重新训练了生成式攻击方法的扰动生成器。类似地，也用此方式测试了TRADES与PGD-AT在自适应攻击场景下的防御效果。

如表1所示，提出的特征正则化对抗训练FRAT在所有数据集的主要指标上均获得了最佳的防御效果。在OTB100数据集上，FRAT能够为原始跟踪器在成功得分与精度得分上分别带来32.86%与35.56%的提升；在LaSOT数据集上，FRAT能够提升30.84%与28.94%的成功得分与归一化精度得分。这显示了FRAT具有良好的防御鲁棒性，能够有效抵御对抗攻击对跟踪器带来的干扰。特别地，在针对对深度特征施加干扰的DFA方法的防御结果中，FRAT能够取得最优的结果，其核心原因在于特征正则损失项能够对对抗特征实施对齐，从而排除对抗扰动在特征空间中的影响。但在VOT2018数据集上，对IoU Attack方法开展防御时，FRAT的准确性指标（0.575）较之于最优的AADN（0.647）存在差距。其原因在于FRAT尽管能够保证预测边框与真实边框具有最大的期望重叠，但二者的中心位置误差却不能够实现最佳。由此导致了准确性指标的差距。最后，在TRADES与PGD-AT上的实验结果也证实了为分类器设计的对抗训练方法在直接迁移至目标跟踪任务上时效果不佳，这与现有研究观点保持一致[10, 28]。

表1 在自适应攻击场景下的防御鲁棒性比较

攻击方法	防御方法	OTB100数据集				VOT2018数据集						LaSOT数据集			
		Succ.	ΔSucc./%	Pre.	ΔPre./%	EAO	ΔEAO/%	A	ΔA/%	$R\downarrow$	$\Delta R\downarrow$/%	Succ.	ΔSucc/%	N.Prec.	ΔN.Prec/%
IoU Attack	无防御	0.499	–	0.644	–	0.129	–	0.568	–	1.171	–	0.334	–	0.387	–
	AADN	0.627	25.65	0.826	28.26	0.334	158.91	**0.647**	**13.91**	0.431	63.19	0.432	29.34	0.495	27.91
	RTAA	0.538	7.82	0.724	12.42	0.178	37.98	0.472	−16.90	0.829	29.21	0.350	4.79	0.426	10.08
	CMDN	0.613	22.85	0.817	26.86	0.195	51.16	0.561	−1.23	0.740	36.81	0.428	28.14	0.486	25.58
	TRADES	0.481	−3.61	0.653	1.40	0.131	1.55	0.537	−5.46	1.433	22.37	0.321	−3.89	0.360	−6.98
	PGD-AT	0.511	2.40	0.665	3.26	0.139	7.75	0.544	−4.23	0.967	−17.42	0.346	3.59	0.379	−2.07
	FRAT	**0.663**	**32.87**	**0.873**	**35.56**	**0.342**	**165.12**	0.575	1.23	**0.262**	**77.63**	**0.437**	**30.84**	**0.499**	**28.94**
DFA	无防御	0.336	–	0.517	–	0.071	–	0.382	–	2.056	–	0.161	–	0.124	–
	AADN	0.486	44.64	0.663	28.24	0.122	71.83	0.395	3.40	1.161	43.53	0.256	59.01	0.197	58.87
	RTAA	0.470	39.88	0.622	20.31	0.113	59.15	0.386	1.05	1.339	34.87	0.249	54.66	0.299	141.13
	CMDN	0.491	46.13	0.630	21.86	0.204	187.32	0.582	52.36	0.698	66.05	0.277	72.05	0.355	186.29
	TRADES	0.365	8.63	0.552	6.77	0.099	39.44	0.408	6.81	1.667	−18.92	0.188	16.77	0.148	19.35
	PGD-AT	0.357	6.25	0.538	4.06	0.087	22.54	0.400	4.71	1.762	−14.30	0.173	7.45	0.134	8.06
	FRAT	**0.515**	**53.27**	**0.668**	**29.21**	**0.218**	**207.04**	**0.593**	**55.24**	**0.676**	**67.12**	**0.287**	**78.26**	**0.361**	**191.13**
CSA	无防御	0.349	–	0.491	–	0.069	–	0.541	–	1.147	–	0.180	–	0.219	–
	AADN	0.461	32.09	0.636	29.53	0.104	50.72	0.549	1.48	1.573	−37.14	0.261	45.00	0.303	38.36
	RTAA	0.413	18.34	0.588	19.76	0.126	82.61	0.562	3.88	0.879	23.37	0.235	30.56	0.279	27.40
	CMDN	0.614	75.93	0.824	67.82	0.250	262.32	0.583	7.76	0.567	50.57	0.397	120.56	0.436	99.09
	TRADES	0.367	5.16	0.524	6.72	0.086	24.64	0.478	−11.65	1.122	−2.18	0.212	17.78	0.256	16.89
	PGD-AT	0.358	2.58	0.511	4.07	0.076	10.14	0.465	−14.05	1.128	−1.66	0.202	12.22	0.253	15.53
	FRAT	**0.623**	**78.51**	**0.831**	**69.25**	**0.252**	**265.22**	**0.599**	**10.72**	**0.560**	**51.18**	**0.422**	**134.44**	**0.453**	**106.85**

3.5 在异构跟踪器上的防御迁移性分析

为了探究防御方法在异构跟踪器上的迁移性，将基线防御方法分别迁移至基于Transformer的跟踪器（SiamTPN[27]和TransT[19]）与基于孪生网络的跟踪器（SiamRPN[16]）上，并在OTB100数据集

与 VOT2018 数据集上部署 IoU Attack 来测试防御性能表现。

值得注意的是，由于不同跟踪器所采用的骨干网络不尽相同，故在实验时依据前面所提的方式对参与特征匹配的潜在特征进行对齐。具体而言，对使用 ShuffleNetv2 骨干网络的 SiamTPN，对齐了来自第 2、4、6 层的潜在特征；对使用 ResNet50 骨干网络的 TransT 与 SiamRPN，分别对齐了其第 4、5 层潜在特征。

实验结果如表 2 所示，本文方法能够在 OTB100 上超越现有防御方法，获得最优的结果。FRAT 能够使 SiamTPN 跟踪器较无防御时获得 19.38% 的成功得分提升；使 TransT 跟踪器获得 5.67% 的成功得分提升；使 SiamRPN 跟踪器获得 20.20% 的成功得分提升。在 VOT2018 数据集上，FRAT 能够在所有跟踪器上的期望平均覆盖与鲁棒性指标中获得最优。这证明了 FRAT 具有更加强大的防御迁移性，能够对基于 Transformer 与孪生网络的跟踪器带来对抗鲁棒性提升。但与前面结果类似，在准确性指标上，FRAT 在 SiamRPN 跟踪器（0.575）上与最佳结果（0.581）存在差距。其原因可能是 SiamRPN 仅使用骨干网络的最终特征输出进行预测，故本文的特征正则化损失无法完全发挥约束作用，因此导致了性能上的差距。

表 2　在异构跟踪器上的防御迁移性比较

跟踪器	防御方法	OTB100数据集				VOT2018数据集					
		Succ.	ΔSucc./%	Pre.	ΔPre./%	EAO	ΔEAO/%	A	ΔA/%	$R\downarrow$	Δ$R\downarrow$/%
SiamTPN	无攻击无防御	0.702	–	0.902	–	0.462	–	0.61	–	0.131	–
	无防御	0.552	–	0.735	–	0.236	–	0.534	–	0.895	–
	AADN	0.465	−15.76	0.629	−14.42	0.134	−43.22	0.521	−2.43	1.091	−21.90
	RTAA	0.621	12.50	0.842	14.56	0.307	30.08	0.595	11.42	0.522	41.68
	CMDN	0.611	10.69	0.826	12.38	0.313	32.63	0.607	13.67	0.419	53.18
	FRAT	**0.659**	**19.38**	**0.855**	**16.33**	**0.371**	**57.20**	**0.620**	**16.10**	**0.362**	**59.55**
TransT	无攻击无防御	0.695	–	0.899	–	0.264	–	0.589	–	0.403	–
	无防御	0.617	–	0.834	–	0.170	–	0.57	–	0.848	–
	AADN	0.405	−34.36	0.564	−32.37	0.150	−11.76	0.541	−5.09	0.974	−14.86
	RTAA	0.634	2.76	0.830	−0.48	0.203	19.41	0.588	3.16	0.589	30.54
	CMDN	0.626	1.46	0.828	−0.72	0.199	17.06	0.567	−0.53	0.587	30.78
	FRAT	**0.652**	**5.67**	**0.835**	**0.12**	**0.236**	**38.82**	**0.592**	**3.86**	**0.455**	**46.34**
SiamRPN	无攻击无防御	0.629	–	0.847	–	0.383	–	0.586	–	0.276	–
	无防御	0.490	–	0.699	–	0.185	–	0.562	–	1.927	–
	AADN	0.517	5.51	0.734	5.01	0.198	7.03	**0.581**	**3.38**	0.714	62.95
	RTAA	0.535	9.18	0.742	6.15	0.207	11.89	0.578	2.85	0.679	64.76
	CMDN	0.544	11.02	0.756	8.15	0.211	14.05	0.580	3.20	0.778	59.63
	FRAT	**0.589**	**20.20**	**0.819**	**17.17**	**0.242**	**30.81**	0.575	2.31	**0.632**	**67.20**

3.6　在干净样本上的泛化性分析

由于神经网络在对抗样本上的鲁棒性提升的代价是牺牲其在干净样本上的精度，并且在现实场景中防御部署者往往无法得知输入图像中是否含有对抗扰动，故防御网络不仅应具有对对抗扰动的鲁棒性，还应该保持在干净样本上的预测性能。为了探究提出的方法在干净样本上的性能表现，在干净样本上开展了实验。

如表 3 所示，FRAT 在大多数评价指标上取得了最佳的结果。但在 VOT2018 数据集的期望平均覆盖指标（0.281）与 CMDN（0.283）存在小幅差距，其原因可能在于在对抗训练过程中跟踪网络对干净样本出现了轻度的欠拟合。然而 FRAT 获得了最优的鲁棒性指标，这表示本文方法能够有效缓解因对抗扰动导致的目标丢失现象，使跟踪过程更加平稳。在 LaSOT 数据集上的归一化精度（0.405）得分指标中，本文方法与 CMDN（0.502）存在差距。这可能的原因有二：首先，有研究指出在训练权重相同的情况下，对抗训练对检测模型的定位精度损害显著高于分类精度[28]；同时，对抗训练更加倾向于缓解目标丢失问题，而在 LaSOT 数据集所聚焦的长时跟踪场景下，由于训练方式或跟踪过程中引入的定位偏差可能在连续帧中累积，进而放大了成功得分良好但精度得分出现较大程度下

降的现象，这也能够被来自 TRADES与 PGA-AT 的实验结果所验证。此类现象可以通过提高回归任务在损失函数中的权重，或者通过特征解耦增强回归任务特征通道的方式得到缓解。

表 3　在干净样本上的泛化性比较

防御方法	OTB100数据集		VOT2018数据集			LaSOT数据集	
	Succ.	Pre.	EAO	A	$R\downarrow$	Succ.	N.Prec.
无防御	0.695	0.905	0.414	0.600	0.234	0.496	0.575
AADN	0.640	0.847	0.245	0.581	0.510	0.432	0.493
CMDN	0.644	0.856	**0.283**	0.578	0.393	0.444	**0.502**
TRADES	0.621	0.818	0.236	0.555	0.569	0.428	0.471
PGD-AT	0.632	0.835	0.249	0.562	0.547	0.434	0.484
FRAT	**0.652**	**0.860**	0.281	**0.588**	**0.382**	**0.446**	0.405

3.7　对训练阶段计算开销的分析

相比于基于预处理的防御方法，本文提出的 FRAT 无需在推理阶段引入额外的计算，从而保证跟踪器在推理时的处理效率。但对抗训练方法在训练阶段的计算开销依旧不容忽视。因此，与基线对抗训练方法 TRADES 与 PGD-AT 对比了在 SiamRPN++跟踪器上完成 20 轮训练的耗时，实验结果如表 4 所示。

表 4　对训练阶段计算开销的比较

训练方法	训练耗时/min	时间开销倍率
原始训练	1 521	–
TRADES	11 964	7.86
PGD-AT	9 922	6.52
FRAT	**4 617**	**3.03**

如表 4 所示，本文提出的 FRAT 的时间开销约为原始训练的 3.03 倍。结合表 1 中的实验结果，与基线对抗训练方法 TRADES 与 PGD-AT 相比，FRAT 能够使用约 46% 的时间开销获得更加强大对抗鲁棒性提升，这进一步证明了 FRAT 具有更加优秀的应用潜力。

3.8　对特征正则项的消融实验

如前面所述，特征正则项通过对基线跟踪器骨干网络浅、中、深层的潜在特征添加约束，以消除对抗扰动在特征空间中对潜在特征产生的影响。为明确特征正则化层数对对抗鲁棒性的贡献，基于 SiamRPN++跟踪器，通过组合不同特征正则化层数并在 OTB100 数据集上部署 IoU Attack 进行消融实验。

如表 5 所示，在仅选择一层潜在特征进行正则化时，可以观察到越深层次的潜在特征对鲁棒性提升的贡献越大。而在同时正则化两层潜在特征时，选择正则化第 4、5 层的潜在特征会具有最佳的对抗鲁棒性，而正则化第 3、5 层与第 3、4 层所带来的对抗鲁棒性提升会依次降低。

这表明对抗扰动引入的干扰在骨干网络的卷积层中被持续放大，在深层次潜在特征中添加正则化切断了对抗扰动的传播链路。其次，由于浅层特征包含了较多纹理和局部结构信息，但其对全局任务的判别性的贡献有限。而深层特征融合了更多抽象的高级语义信息，它们更加适用于视觉目标跟踪的语义匹配。因此正则化的潜在特征层数越深，其带来的对抗鲁棒性提升越显著。

表 5　不同层潜在特征对鲁棒性的贡献

特征正则化层数			OTB100数据集			
Conv3	Conv4	Conv5	Succ.	ΔSucc./%	Pre.	ΔPre./%
			0.499	—	0.644	—
√			0.553	+10.82	0.735	+14.13
	√		0.598	+19.84	0.788	+22.36
		√	0.634	+27.05	0.842	+30.75
√	√		0.642	+28.66	0.853	+32.45
	√	√	0.655	+31.26	0.866	+34.47
√		√	0.649	+30.06	0.851	+32.14
√	√	√	**0.663**	**+32.87**	**0.873**	**+35.56**

3.9　对扰动预算的消融实验

前面我们通过扰动预算ϵ限制生成的对抗扰动大小。为探究超参数ϵ对防御鲁棒性的影响，在 OTB100 数据集上部署 IoU Attack 进行消融实验，结果如图 4 所示。

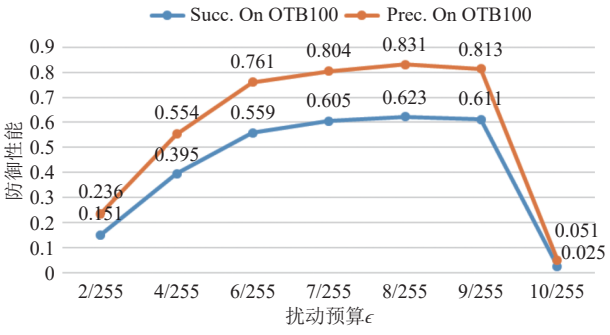


图 4　扰动预算对训练性能的影响

如图所示，ϵ过小或过大均将加剧网络模型对对抗样本的脆弱表现。在扰动预算逐渐从 2/255 扩大至 10/255 的过程中，防御性能将会逐渐提高，并在ϵ=8/255 处达到最大值。当ϵ=9/255 时将会导致轻微的性能下降，进一步扩大至 10/255 时，防御性能将会急剧下降，此时跟踪器的预测结果出现了灾难性的错误。出现这一现象的原因首先是过小的扰动预算设置将会使网络模型难以学习到有用的

对抗扰动，从而无法获得有效的防御鲁棒性。其次，过大的扰动预算将会使模型在训练时过拟合到巨大的对抗扰动上，这和对抗攻击所采用的人眼不可察觉的扰动是不相符的。换言之，过大的扰动预算使网络模型学习到的扰动超出了对抗扰动的存在范围，从而无法对对抗攻击保持有效的防御性能。再次，在对抗攻击的通用设置中，研究者们通常将小于 8/255 的对抗扰动认为是人眼不可感知的扰动，而实验结果也印证了该先验知识的有效性。综上所述，本文将ϵ设置为 8/255 是合理的。

4　结束语

本文首先指出了现有目标跟踪防御方法均为基于预处理的防御方法这一研究现状，进而阐述了此类方法在面对自适应攻击时将导致显著的性能下降这一缺点。其次，基于在不同卷积尺度下被对抗特征与干净原始特征具有不同程度的分离这一经验观察，提出了特征正则化项以使跟踪器的对抗特征在不同尺度下与干净特征实现对齐。最后，结合目标跟踪任务具有双图像输入的特性，设计了适用于目标跟踪网络的对抗训练策略，通过以特征正则化损失引导的对抗训练，能够使目标跟踪网络获得更强大的对抗鲁棒性提升。在 OTB100、VOT2018 与 LaSOT 数据集上开展的实验证明了提出的方法相比同类方法具有最优的性能。

参 考 文 献

[1] LI P X, JIN J Y. Time3D: End-to-end joint monocular 3D object detection and tracking for autonomous driving[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2022: 3875-3884.

[2] LUO C X, YANG X D, YUILLE A. Exploring simple 3D multi-object tracking for autonomous driving[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. New York: IEEE, 2021: 10468-10477.

[3] SANDOVAL L A C. Low cost object tracking by computer vision using 8 bits communication with a viper robot[C]//Proceedings of the 8th International Conference on Control and Robotics Engineering. New York: IEEE, 2023: 232-237.

[4] GOODFELLOW I J, SHLENS J, SZEGEDY C. Explaining and harnessing adversarial examples[EB/OL]. [2024-04-14]. https://www.researchgate.net/publication/269935591_Explaining_and_Harnessing_Adversarial_Examples.

[5] YAN B, WANG D, LU H C, et al. Cooling-shrinking attack: Blinding the tracker with imperceptible noises[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. [S. l.]: IEEE, 2020: 990-999.

[6] SUTTAPAK W, ZHANG J F, ZHANG L Q. Diminishing-feature attack: The adversarial infiltration on visual tracking[J]. Neurocomputing, 2022, 509: 21-33.

[7] JIA S, SONG Y B, MA C, et al. IoU attack: Towards temporally coherent black-box adversarial attack for visual object tracking[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. [S. l.]: IEEE, 2021: 6709-6718.

[8] LIU S A, CHEN Z Y, LI W, et al. Efficient universal shuffle attack for visual object tracking[C]//Proceedings of the ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing. New York: IEEE, 2022: 2739-2743.

[9] GUO Q, XIE X F, JUEFEI-XU F, et al. SPARK: spatial-aware online incremental attack against visual tracking[M]//Computer Vision – ECCV 2020. Cham: Springer International Publishing, 2020: 202-219.

[10] WU Z W, YU R L, LIU Q H, et al. Enhancing tracking robustness with auxiliary adversarial defense networks [M]//Computer Vision – ECCV 2024. Cham: Springer Nature Switzerland, 2024: 198-214.

[11] CHEN J L, REN X H, GUO Q, et al. LRR: Language-driven resamplable continuous representation against adversarial tracking attacks[EB/OL]. [2024-04-15]. https://arxiv.org/abs/2404.06247.

[12] TRAMER F, CARLINI N, BRENDEL W, et al. On adaptive attacks to adversarial example defenses[J]. Advances in Neural Information Processing Systems, 2020, 33: 1633-1645.

[13] CARLINI N, WAGNER D. Adversarial examples are not easily detected: Bypassing ten detection methods[C]// Proceedings of the 10th ACM Workshop on Artificial Intelligence and Security. New York: ACM, 2017: 3-14.

[14] BAI T, LUO J Q, ZHAO J, et al. Recent advances in adversarial training for adversarial robustness[EB/OL]. [2024-04-20]. https://arxiv.org/abs/2102.01356.

[15] ZHANG H C, WANG J Y. Towards adversarially robust object detection[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. New York: IEEE, 2019: 421-430.

[16] BERTINETTO L, VALMADRE J, HENRIQUES J F, et al. Fully-convolutional Siamese networks for object tracking[M]//HUA G, JÉGOU H, eds. Computer Vision – ECCV 2016 Workshops. Cham: Springer International Publishing, 2016: 850-865.

[17] LI B, YAN J J, WU W, et al. High performance visual tracking with Siamese region proposal network[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2018: 8971-8980.

[18] LI B, WU W, WANG Q, et al. SiamRPN++: Evolution of Siamese visual tracking with very deep networks[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE, 2019: 4282-4291.

[19] WANG Q, ZHANG L, BERTINETTO L, et al. Fast online object tracking and segmentation: A unifying approach[C]//

Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2019: 1328-1338.

[20] CHEN X, YAN B, ZHU J, et al. Transformer tracking[C]// Proceedings of the IEEE/CVF Conference On Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2021: 8126-8135.

[21] XING D T, EVANGELIOU N, TSOUKALAS A, et al. Siamese transformer pyramid networks for real-time UAV tracking[C]//Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. New York: IEEE, 2022: 1898-1907.

[22] JIA S, MA C, SONG Y B, et al. Robust tracking against adversarial attacks[M]//VEDALDI A, BISCHOF H, BROX T, et al, eds. Computer Vision – ECCV 2020. Cham: Springer International Publishing, 2020: 69-84.

[23] YU R L, WU Z W, LIU Q H, et al. CMDN: Pre-trained visual representations boost adversarial robustness for UAV tracking[J]. Drones, 2024, 8(11): 607.

[24] SZEGEDY C, ZAREMBA W, SUTSKEVER I, et al. Intriguing properties of neural networks[EB/OL]. [2024-05-2]. https://www.researchgate.net/publication/259440613_Intriguing_properties_of_neural_networks.

[25] MIYATO T, KATAOKA T, KOYAMA M, et al. Spectral normalization for generative adversarial networks [EB/OL]. [22024-05-10]. https://arxiv.org/abs/1802.05957.

[26] WU Y, LIM J, YANG M H. Object tracking benchmark[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1834-1848.

[27] KRISTAN M, LEONARDIS A, MATAS J, et al. The sixth visual object tracking VOT2018 challenge results[C]//Computer Vision – ECCV 2018 Workshops. Cham: Springer International Publishing, 2019: 3-53.

[28] FAN H, LIN L T, YANG F, et al. LaSOT: A high-quality benchmark for large-scale single object tracking[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2019: 5369-5378.

[29] ZHANG H Y, YU Y D, JIAO J T, et al. Theoretically principled trade-off between robustness and accuracy [EB/OL]. [2024-05-20]. https://arxiv.org/abs/1901.08573.

[30] RICE L, WONG E, KOLTER Z. Overfitting in adversarially robust deep learning[C]//International Conference on Machine Learning. [S.l.]: PMLR, 2020: 8093-8104.

责任编辑 叶 芳