刘琦, 韩玉林, 郭龙. 基于在线知识蒸馏及对比学习的入侵检测技术[J]. 智能计算机与应用,2025,15(9):139-145. DOI: 10.20169/j.issn.2095-2163.25031101

基于在线知识蒸馏及对比学习的入侵检测技术

刘 琦[1], 韩玉林[1], 郭 龙[2]
(1 中国海油(中国)有限海南分公司, 海口 570100; 2 中海油信息科技有限公司湛江分公司, 广东 湛江 524000)

摘 要: 随着物联网在各个行业的广泛应用,发展速度也在不断加快,其中包含的设备类型增长非常迅速,设备中多种新型漏洞构成了紧迫的网络安全威胁。大部分基于行为的入侵检测系统(Network Intrusion Detection System,NIDS)依赖人工智能模型。在多次攻防演练中,发现攻击集中于脆弱的边缘端,这些边缘端设备通常算力不足且存储空间有限,无法部署参数较多的检测模型。本文基于线性知识蒸馏梯度对齐的注意力模块,动态筛选教师模型的高价值特征,并通过对比学习,增强学生模型对对抗性样本的鲁棒性,使轻量化学生模型在有限的训练时间内蒸馏出了更多的知识。与传统的入侵检测模型比较,所提的基于卷积神经网络(Convolutional Neural Network,CNN)的在线知识蒸馏(Knowledge Distillation,KD)的学生模型在保持92%准确性和98%召回率的性能下,将推理计算的参数数量降低至教师模型的6%,为低功耗条件下的物联网设备提供了有力安全保障。

关键词: 物联网; 入侵检测; 知识蒸馏; 卷积神经网络; 对比学习
中图分类号: TP182 **文献标志码**: A **文章编号**: 2095-2163(2025)09-0139-07

Intrusion detection based on online knowledge distillation and contrastive learning

LIU Qi[1], HAN Yulin[1], GUO Long[2]
(1 Hainan Branch of China National Offshore Oil Corporation (CNOOC) (China) Co., Ltd., Haikou 570100, China;
2 Zhanjiang Branch of CNOOC Information Technology Co., Ltd., Zhanjiang 524000, Guangdong, China)

Abstract: With the widespread application of the Internet of Things (IoT) across various industries and the continuous acceleration of its development, the types of devices involved are growing rapidly, leading to an urgent network security threat due to the emergence of multiple new vulnerabilities in these devices. Most behavior-based Network Intrusion Detection Systems (NIDS) rely on artificial intelligence models. During multiple attack-defense drills, it is found that attacks concentrate on vulnerable edge devices, which often have insufficient computing power and limited storage space, making it impossible to deploy detection models with numerous parameters. This study proposes an attention module based on linear knowledge distillation gradient alignment, which dynamically selects high-value features from the teacher model. After that, the research enhances the robustness of the student model against adversarial samples by contrastive learning, allowing the lightweight student model to distill more knowledge within a limited training time. Compared to traditional intrusion detection models, the proposed student model based on Convolutional Neural Networks (CNN) employing online Knowledge Distillation (KD) reduces the number of hidden layers in the inference model to 6% of the original model while maintaining 92% accuracy and 98% recall. This provides robust security assurance for IoT devices under low-power conditions.

Key words: Internet of Things; intrusion detection; knowledge distillation; Convolutional Neural Network; contrastive learning

0 引 言

近年来,物联网边缘端设备呈现规模化发展的趋势,设备的种类和网络复杂度不断增加,如何在边缘端有效地进行入侵检测是目前亟需解决的问题[1-3]。特别是在极端环境下部署的依靠电池工作的边缘端,抗攻击能力较弱,对入侵检测算法的实时性和精准度要求更高[4]。某企业创新开展“平战结

基金项目: 国家重点研发计划深海和极地关键技术与装备专题(2022YFC2806205)。
作者简介: 刘 琦(1986—),男,系统应用工程师,主要研究方向:信息化系统应用。Email:liuqi5@cnooc.com.cn。
收稿日期: 2025-03-11

合,精准防护”网络安全专项工作,以“动态防御,主动防御,纵深防御,精准防护,整体防护,联防联控”的思想实现了云边协同的立体网络防护模式。

结合动态防御策略,边缘侧与云中心将负责不同的计算任务。在数据流量较大时,由云中心进行流量监测、设备检查等工作,收集威胁信息;边缘侧实时进行入侵检测,在边缘侧对网络通信内容进行抓包并压缩,提取类似经典的入侵检测数据集(Knowledge Discovery and Data Mining,KDD)的格式特征[5]并保存。在数据流量较少时,边缘侧将所有收集的数据上传至云中心作为训练集,构建新的入侵检测识别模型,通过知识蒸馏获得轻量级的学生模型,将在线训练好的学生模型发送给边缘端进行更新[6]。

传统的 NIDS 基于已知攻击的签名来判断攻击的类型,此外,还需要训练已知威胁演变出的相似的复杂攻击的签名[7]。典型的机器学习分类技术基于现有标记的系统活动的数据集[8-16]。在大多数情况下,物联网设备只要判断正常或恶意,即输出为二元分类,但这些已有算法的复杂性大大超过了边缘端的计算能力。

不均衡的样本分布是入侵检测中面临的主要挑战。通常,正常流量样本容易收集,而恶意流量样本、尤其是高级攻击样本,却难以获取。许多研究者提出了结合对比学习与深度学习入侵模型的方法来解决这一问题[17-18]。基于以上的背景,提出了一种由多个残差块构成的 CNN 模型的在线知识蒸馏和的对比学习方法,在算法层面针对不均衡的数据类别分布进行了训练加强,实现了轻量级入侵检测模型的判断准确性提升的目标。

1 知识蒸馏

已有大量的研究表明,深度学习模型在具备充足可训练数据的情况下,能够实现相较于机器学习模型更优秀的入侵检测效果。在入侵检测领域,深度学习模型面临着一个显著的矛盾:既需要保持高准确性,又需要具有轻量级特征。知识蒸馏的目标是通过深度学习的教师网络训练一个更紧凑、更准确的学生模型。

本文仿照 ResNet 构建了多个残差块。每个残差块包括 2 个卷积层,每个卷积层后面是池化层和 ReLU 激活函数。在不改变残差块的数量的情况下,教师网络和学生网络的结构非常相似,在对比学习时不影响分类结果。在这种设计下,可以灵活削减每个残差块中的卷积层参数数量。学生网络的每个残差块的隐层参数数量可以比教师网络的少。

教师-学生网络模型结构如图 1 所示。以下段落将介绍某企业部署于其物联网的一种在线知识蒸馏的教师-学生网络模型,比对在线的蒸馏模型和离线的蒸馏模型的学生网络准确性,同时通过 NSL-KDD 数据集证实对比学习的方法可有效改善训练数据的平衡性。

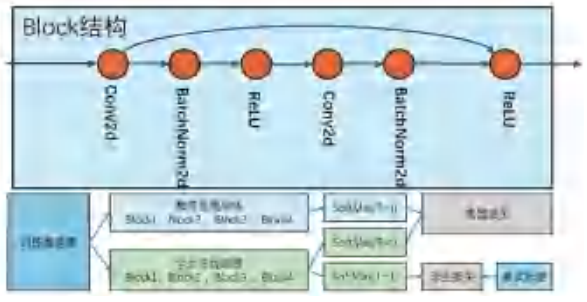

图 1 教师-学生网络模型结构图

Fig. 1 Architectural diagram of the teacher - student network model

2 数据集及网络结构

实验的数据集分为训练集、验证集和测试集。在训练集中,选取了比较新的 NSL-KDD 数据文件,得到了比较平衡的数据类型分布。这个数据集是通过从 DARPA-KDD 中移除重复记录而获得的入侵检测基准。该数据集包含 38 种攻击类型,其中 24 种用于训练集,14 种用于测试集。这 38 种攻击类型为 DoS、Probe、U2R 和 R2L。这个新的数据集包含 148 517 条流量记录,每个记录包含 169 维的统计网络数据特征。训练集包含 58 000 条攻击数据和 60 000 条正常数据。验证集包含 630 条攻击数据和 7 343 条正常数据。测试集合包含 12 833 条攻击数据和 9 711 条正常数据。为了最大限度保持各个特征的有效信息,在转换成图像之前对所有的数据做鲁棒缩放器归一化处理。在鲁棒缩放中,从给定特征的值中减去中位数,然后除以四分位数范围,转换为分布在[0, 1]之间的数值。预测时需要将所有的数值根据同样的鲁棒缩放器转换为对应的[0, 1]之间的数值。

通过对残差块的组合可以构建二维的 CNN 分类的深度学习网络[19-20]。为了提取数据中高度复杂的特征,采用多个残差块(Block)中的多个卷积层、跳跃连接和输出来避免梯度消失。具体模型结构见表 1、表 2。

表 1　教师网络模型结构表

Table 1　Teacher network model structure

教师网络各层类型	输出形状	参数数量
Conv2d-1	[-1, 8, 12, 12]	224
BatchNorm2d-2	[-1, 8, 12, 12]	16
ReLU-3	[-1, 8, 12, 12]	0
Conv2d-4	[-1, 8, 12, 12]	584
BatchNorm2d-5	[-1, 8, 12, 12]	16
Conv2d-6	[-1, 8, 12, 12]	32
ReLU-7	[-1, 8, 12, 12]	0
ResidualBlock-8	[-1, 8, 12, 12]	0
Conv2d-9	[-1, 8, 12, 12]	584
BatchNorm2d-10	[-1, 8, 12, 12]	16
ReLU-11	[-1, 8, 12, 12]	0
Conv2d-12	[-1, 8, 12, 12]	584
BatchNorm2d-13	[-1, 8, 12, 12]	16
ReLU-14	[-1, 8, 12, 12]	0
ResidualBlock-15	[-1, 8, 12, 12]	0
Conv2d-16	[-1, 8, 12, 12]	584
BatchNorm2d-17	[-1, 8, 12, 12]	16
ReLU-18	[-1, 8, 12, 12]	0
Conv2d-19	[-1, 8, 12, 12]	584
BatchNorm2d-20	[-1, 8, 12, 12]	16
ReLU-21	[-1, 8, 12, 12]	0
ResidualBlock-22	[-1, 8, 12, 12]	0
Conv2d-23	[-1, 8, 12, 12]	584
BatchNorm2d-24	[-1, 8, 12, 12]	16
ReLU-25	[-1, 8, 12, 12]	0
Conv2d-26	[-1, 8, 12, 12]	584
BatchNorm2d-27	[-1, 8, 12, 12]	16
ReLU-28	[-1, 8, 12, 12]	0
ResidualBlock-29	[-1, 8, 12, 12]	0
Linear-30	[-1, 2]	2 306

注：总参数为 6 778

表 2　学生网络模型结构表

Table 2　Student network model structure

学生网络各层类型	输出形状	参数数量
Conv2d-1	[-1, 1, 12, 12]	28
BatchNorm2d-2	[-1, 1, 12, 12]	2
ReLU-3	[-1, 1, 12, 12]	0
Conv2d-4	[-1, 1, 12, 12]	10
BatchNorm2d-5	[-1, 1, 12, 12]	2
Conv2d-6	[-1, 1, 12, 12]	4
ReLU-7	[-1, 1, 12, 12]	0
ResidualBlock-8	[-1, 1, 12, 12]	0
Conv2d-9	[-1, 1, 12, 12]	10
BatchNorm2d-10	[-1, 1, 12, 12]	2
ReLU-11	[-1, 1, 12, 12]	0
Conv2d-12	[-1, 1, 12, 12]	10
BatchNorm2d-13	[-1, 1, 12, 12]	2
ReLU-14	[-1, 1, 12, 12]	0
ResidualBlock-15	[-1, 1, 12, 12]	0
Conv2d-16	[-1, 1, 12, 12]	10
BatchNorm2d-17	[-1, 1, 12, 12]	2
ReLU-18	[-1, 1, 12, 12]	0
Conv2d-19	[-1, 1, 12, 12]	10
BatchNorm2d-20	[-1, 1, 12, 12]	2
ReLU-21	[-1, 1, 12, 12]	0
ResidualBlock-22	[-1, 1, 12, 12]	0
Conv2d-23	[-1, 1, 12, 12]	10
BatchNorm2d-24	[-1, 1, 12, 12]	2
ReLU-25	[-1, 1, 12, 12]	0
Conv2d-26	[-1, 1, 12, 12]	10
BatchNorm2d-27	[-1, 1, 12, 12]	2
ReLU-28	[-1, 1, 12, 12]	0
ResidualBlock-29	[-1, 1, 12, 12]	0
Linear-30	[-1, 2]	290

注：总参数为 408

表 1 中展示的教师网络每个卷积层中都设置了 8 个隐层。表 2 中展示的学生网络在每个卷积层中都设置了 1 个隐层。从参数总量可以看出，学生网络在最简单的时候可以达到教师模型总参数量的 6%。在离线状态下，从训练集中随机抽取 50 000 条数据训练教师网络，然后使用剩余的训练集数据在教师网络的教导下离线蒸馏学生网络，可对比学生网络参数在训练达到 30 次之后，隐层数对训练准确率的影响。

学生网络不同参数量的准确性对比如图 2 所示。当通过调整学生网络的卷积层、隐层数来获得不同的参数数量时，即使使用最小的参数数量也能获得很好的训练结果。使用学生网络训练结果对测试集进行测试。在涉及入侵检测任务的情况下，所有结果对应以下 4 种类别：

（1）真正例（True Positive，TP）：恶意样本被正确地分类为恶意。

（2）真反例（True Negative，TN）：正常样本被正

确地分类为正常。

(3)假正例(False Positive,FP):正常样本被错误地分类为恶意。

(4)假反例(False Negative,FN):恶意样本被错误地分类为正常。定义准确率(Acc)、精确度(Pre)、召回率(Rec)、$F1$ 分数的数学公式如下:

$$\mathrm{Acc} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}} \tag{1}$$

$$\mathrm{Pre} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{2}$$

$$\mathrm{Rec} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{3}$$

$$F1 = \frac{2 \cdot \mathrm{Pre} \cdot \mathrm{Rec}}{\mathrm{Pre} + \mathrm{Rec}} \tag{4}$$

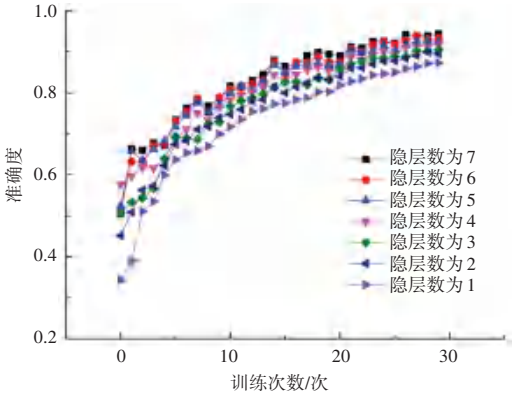

图 2 学生网络不同参数量的准确性对比图

Fig. 2 Accuracy comparison of the student network with different parameter sizes

接下来,研究得到的学生网络参数数量对预测性能影响见表 3。观察到随着隐层数的下降,学生网络的判别能力下降,但是仍然能够保持在较高的水平。

表 3 学生网络参数数量对预测性能影响表

Table 3 Impact of student network parameter number on prediction performance %

隐层数	准确率	召回率	精确率	$F1$ 分数
7	91.75	87.50	98.67	92.75
6	91.80	87.77	98.85	92.98
5	91.60	87.36	98.62	92.65
4	91.50	87.20	98.58	92.54
3	91.40	88.04	98.00	92.75
2	91.35	87.05	98.50	92.42
1	91.35	86.82	96.67	91.48

从表 3 中可以看出,在参数量减少的过程中,预测精度没有出现显著降低,说明在预测准确率达到 88%以上的情况下就可以将学生网络的参数调整至最低隐层数量。在入侵检测领域,达到如上的准确率仍然不足以较好地应对网络安全威胁。在保持轻量化的同时,必须对学生网络的蒸馏模型进行后续优化与有效的改进,以便于进一步提高判断准确性。

3 模型改进及实验

根据深度神经网络在入侵检测网络流量识别过程中的特点,加入了在线蒸馏和对比学习。与现有的入侵检测算法不同,本文提出了一种基于在线知识蒸馏的联合训练方法。其中,多隐层的卷积网络充当教师网络,较少隐层数量的学生网络充当学生网络,实现了教师网络和学生网络的联合训练和优化。本文将这种联合训练优化的教师-学生网络称为在线知识蒸馏。

3.1 在线知识蒸馏

为了体现端到端学习的互补优势,算法通过不同隐层数量的学生网络生成了的初步判断的特征集,根据输出的特征集之间的差异就可以得到特征损失。通过将蒸馏损失、对比损失、特征损失相结合,整体上实现了在线知识蒸馏。

给定网络数据样本集 $x = [X_1, X_2, \cdots, X_n]$,其中 X_i 属于 d 维空间,以及所对应标签集 $y = [Y_1, Y_2, \cdots, Y_n]$,这里 Y_i 属于 $\{0,1\}$ 集合,为了获得更好的预测性能,通常需要使用预定义的损失函数对模型进行优化,目前交叉熵损失函数是最广泛使用的损失函数。在线知识蒸馏网络结构如图 3 所示,在线知识蒸馏网络分为 2 个部分:学生网络和教师网络。

图 3 在线知识蒸馏网络结构图

Fig. 3 Online knowledge distillation network architecture diagram

在学生分支中,教师网络和学生网络对比,每一个块都形成了特征损失,并传递给了学生网络进行修正。给定一个 Logits 向量 Z 作为深度模型最后一个残差块经过全连接层的输出,基于响应的知识蒸馏的教师损失 L 可以表述为:

$$L_T = L_{\mathrm{CE}}[y, \sigma(Z_T; T = t)] + L_C \tag{5}$$

其中,Z_T 表示教师网络的 Logits;L_{CE} 表示教师

网络 y 标签的交叉熵损失；σ 表示加入温度参数 T 后的 Softmax 函数；t 表示自适应温度；L_C 表示对比损失。

在分类任务中，基于响应的知识蒸馏算法的重点是学生学习教师模型输出的软目标知识，与硬目标知识相比，教师的软目标可以提供给学生更多的类间知识。神经网络通常通过使用 Softmax 函数输出层产生二分类概率 q，对于一个网络的输出 Logits，一种网络流量的分类概率值由以下公式计算得出：

$$q(\mathrm{Nor}) = \frac{\exp[Z_S(\mathrm{Nor})/T]}{\exp[Z_S(\mathrm{Nor})/T] + \exp[Z_S(\mathrm{Att})/T]} \quad (6)$$

$$q(\mathrm{Att}) = \frac{\exp[Z_S(\mathrm{Att})/T]}{\exp[Z_S(\mathrm{Nor})/T] + \exp[Z_S(\mathrm{Att})/T]} \quad (7)$$

其中，Z_S 表示学生网络输出的 Logits；Nor 表示正常；Att 表示攻击；T 值表示蒸馏的温度。当 T 值设置为 1 时，为标准的 Softmax 函数；当 T 上升时，Softmax 函数产生的概率分布变得更软，可以给知识蒸馏过程中的学生网络提供更多的正常和攻击概率知识。

对于训练集中的真实标签，在训练学生网络时，同时使用学生模型的真实标签和教师模型生成的软标签能够产生更好的效果。因此，教师训练学生的损失函数由学生自身的损失函数和蒸馏损失函数结合构成，在线知识蒸馏算法的学生网络总损失值 L_S 定义为：

$$L_S(x, \boldsymbol{W}) = (1 - 2\lambda) L_{\mathrm{CE}}[y, \sigma(Z_S; T = 1)] + \lambda L_{\mathrm{CE}}[\sigma(Z_T; T = t), \sigma(Z_S; T = t)] + \lambda L_{\mathrm{CE}}(\mathrm{feature}_T, \mathrm{feature}_S) \quad (8)$$

其中，x 表示输入值；$\boldsymbol{W}$ 表示学生模型的参数；λ 表示权重系数；L_{CE} 表示交叉熵损失；y 表示真实标签；σ 表示加入温度参数 T 后的 Softmax 函数；t 表示自适应蒸馏温度；Z_S 和 Z_T 分别表示学生模型和教师模型的输出 Logits；$\mathrm{feature}_T$ 和 $\mathrm{feature}_S$ 分别表示教师每个残差块的输出特征。设置蒸馏损失和特征损失为同一个权重系数 λ 的原因是在此设置下，网络中的每个模块都能得到及时的矫正。教师网络的总损失函数需要使用到对比学习，因此在下节介绍。

3.2　对比学习

对比学习根据数据样本及其增强视图之间的内在关系构建对比任务。如果考虑样本之间的标签关系，则通过优化监督对比损失，就可以使正样本之间的距离达到最小化，而负样本之间的距离达到最大化。本文采用了随机序列遮蔽方法来构建增强视图，结合样本标签，生成正负样本对。对比交叉熵损失用于以端到端的方式支持对比分类模型的训练，能够同时优化对比损失和分类损失。

内部类多样性与跨类相似性的存在，为识别攻击流量带来了重大挑战。以网络流量样本 X 和标签 Y 为基础，输出的 Logits 为 Z，可以通过同一类别之间样本的平均距离来简单表示内部类距，对此可以表示为：

$$D_{\mathrm{intra}} = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} d(Z_i, Z_j) \ \text{where} \ (y_i = y_j) \quad (9)$$

通过计算不同类别之间样本的平均距离，来衡量类间距离，具体数学公式如下：

$$D_{\mathrm{inter}} = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} d(Z_i, Z_j) \ \text{where} \ (y_i \neq y_j) \quad (10)$$

为了获得更佳的检测结果，不仅要专注于预测标签，还需要关注样本空间本身的自然属性。在对比学习中，温度是一个用于调整相似度分布的超参数。在对比损失函数中，温度影响不同样本之间的相似度计算。当温度较高时，相似度值相对平滑，所有样本的相似度差异变小；当温度较低时，相似度值变得更加尖锐，模型将更加关注最相似和最不相似的样本。适当的温度可以帮助模型更好地收敛，避免梯度消失或模型过于敏感的问题，因此提出对比学习的动态调整的温度选择，推得的公式如下：

$$T_C = \frac{1}{1 + D_{\mathrm{intra}}} \quad (11)$$

其中，T_C 表示对比温度。整体对比损失可以表示为相似样本的损失和不相似样本的损失之和，数学公式定义为：

$$L_C = \sum_{i=1}^{n} [Y_i D_{\mathrm{intra}}^2 + (1 - Y_i)(M - D_{\mathrm{intra}})^2] \quad (12)$$

其中，M 表示样本边界排斥区域，即一个最小的距离差异，即要求在训练过程中，正样本的距离应该小于负样本的距离加上 M。

同时，蒸馏的损失计算中也需要引入一个自适应温度函数。通过调整蒸馏温度 T，可以控制输出概率分布的平滑程度。较高的温度会使得概率分布更加平滑，从而使得模型学习到更为丰富的知识。反之，较低的温度则会使概率分布更加尖锐，强调最有可能的类别。自适应温度方法旨在根据训练过程中学生和教师的 Logits 的余弦相似度来动态调整温度 T。其中，余弦相似度的计算公式为：

$$\mathrm{CosSim}(Z_{\mathrm{Si}}, Z_{\mathrm{Ti}}) = \frac{Z_{\mathrm{Si}} \times Z_{\mathrm{Ti}}}{\| Z_{\mathrm{Ti}} \| \times \| Z_{\mathrm{Si}} \|} \quad (13)$$

其中，Z_{Ti} 和 Z_{Si} 分别表示教师、学生网络在第 i 步训练产生的 Logits。动态温度 T 可由如下公式计算求得：

$$T = G \cdot [\mathrm{CosSim}(Z_{Si}, Z_{Ti}) + 1] \quad (14)$$

其中，G 是经验参数，一般可以选取 5~10 之间的数值。这里，可以使用 5 作为实验的温度。

3.3 消融实验

为验证所提在线知识蒸馏和对比学习算法的有效性，本文进行了消融研究来评估以下关键模块的性能：自适应温度、联合训练、对比学习。

本文在实验中构建了以下变体：变体 A，未采用自适应温度（本文将其替换为固定温度 $T = 10$）；变体 B，未采用联合训练（本文将其替换为教师先训练、然后和学生进行知识蒸馏）；变体 C，未采用对比学习（本文将其替换为随机顺序的训练样本）。

本文在 NSL-KDD 的训练集上对这些变体进行了 30 回合的训练，并在测试数据集上对训练后的变体进行了测试，通过量化比较来评估每个变体的性能，结果见表 4。表 4 中，本体表示未替换任何模块的网络结构。

表 4 各类网络性能影响表

Table 4 Performance impact table of various networks %

类型	准确率	召回率	精确率	$F1$ 分数
本体	92.35	88.16	98.89	93.22
变体 A	91.89	87.65	98.73	92.86
变体 B	91.75	87.50	98.67	92.75
变体 C	91.60	87.36	98.62	92.65

图 4 中，使用 ROC 曲线（Receiver Operating Characteristic Curve，ROC）评估本文所提出的模型的二分类性能。分析可知，AUC（Area Under Curve）达到 0.93，具备较好的二分类能力。与离线蒸馏的学生网络相比，准确率稳定提高了约 1%，其余指标无明显下降。

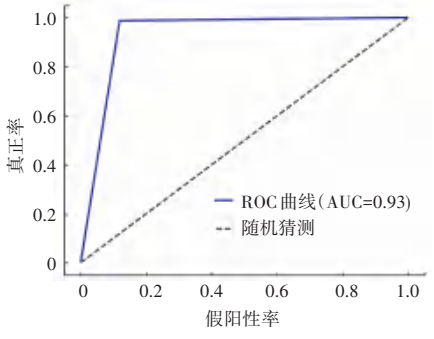

图 4 所提模型的 ROC 曲线

Fig. 4 ROC curve of the proposed model

4 结束语

根据某集团公司“网络安全运营能力建设工程”的网络安全整体规划，落实公安部“提升安全监测预警和应急处置能力”要求，积极响应国家“2024 年专项行动新形势”要求，满足数字化转型和智能化发展需求，解决数字化业务面临系统安全、数据泄露等安全风险，需增强常态化安全防护措施。

在物联网的安全防护中，为解决长期存在的边缘端防护难题，引入云边协同模式。针对边缘侧能耗受限的问题，本文提出了对比学习和在线知识蒸馏的入侵检测识别方法。主要创新有：

（1）使用鲁棒缩放器对转换图像前的数据做归一化处理，将数据降维后统一转换为二维的灰度图像。

（2）引入包含多个残差块的 CNN 网络构建入侵检测模型，使用自适应温度进行训练；用裁剪隐层数量的方式获得学生网络。

（3）使用在线学习和对比学习结合的方法获得合格的学生网络。

通过数值仿真实验证明，本文所提方法在参数量仅为原深度模型 6%的情况下，实现了学生网络在 NSL-KDD 数据集上保持 92%准确性和 98%召回率的性能，可以满足边缘端低功耗设备的部署要求。

参考文献

[1] 冯云霞，王西贤. 基于椭圆曲线加密算法的工业物联网数据隐私保护方案[J]. 智能计算机与应用，2022，12(12)：110-121.

[2] 罗滔. 基于多级鉴权的物联网数据接口管理技术研究[J]. 智能计算机与应用，2024，14(12)：176-179.

[3] 徐新林，邓异. 无线自组网舰载通信数据安全传输技术研究[J]. 智能计算机与应用，2023，13(3)：83-92.

[4] 王壮壮，陈宏松，杨丽敏，等. 联邦学习与数据安全研究综述[J]. 智能计算机与应用，2021，11(1)：126-133.

[5] MEADEM N，VERBIERS N，ZOLFAGHAR K，et al. Exploring preprocessing techniques for prediction of risk of readmission for congestive heart failure patients [C]//Proceedings of the International Conference on Knowledge Discovery and Data Mining (KDD). New York：ACM，2013：247-252.

[6] KIM J，HYUN M，CHUNG I，et al. Feature fusion for online mutual knowledge distillation [C]// Proceedings of the 25th International Conference on Pattern Recognition (ICPR). Piscataway，NJ：IEEE，2021：4619-4625.

[7] BELARBI O，KHAN A，CARNELLI P，et al. An intrusion detection system based on deep belief networks [C]// Proceedings of the 4th International Conference on Science of Cyber Security (SciSec 2022). New York：ACM，2022：377-392.

[8] ZHONG Wei，YU Ning，AI Chunyu，Applying big data based deep learning system to intrusion detection [J]. Big Data Mining

and Analytics，2020，3(3)：181-195.
[9] 张和伟，王奉章. 基于被动分簇算法的即时通信网络安全漏洞检测方法[J]. 智能计算机与应用，2023，13(7)：119-122.
[10] KIM G，LEE S，KIM S，A novel hybrid intrusion detection method integrating anomaly detection with misuse detection [J]. Expert Systems with Applications，2014，41(4)：1690-1700.
[11] KUNANG Y N，NURMAINI S，STIAWAN D，et al. Attack classification of an intrusion detection system using deep learning and hyperparameter optimization [J]. Journal of Information Security and Applications 2021，58：102804.
[12] WANG Wei，SHENG Yiqiang，WANG Jinlin，et al. HAST-IDS：Learning hierarchical spatial-temporal features using deep neural networks to improve intrusion detection [J]. IEEE Access，2017，6：1792-1806.
[13] KANNA P R，SANTHI P. Unified deep learning approach for efficient intrusion detection system using integrated spatial-temporal features [J]. Knowledge-Based Systems，2021，226，107132.
[14] FENG Yebo，LI Jun，JIAO Lei，et al. Towards learning-based，content agnostic detection of social bot traffic [J]. IEEE Transactions on Dependable and Secure Computing，2020，18：2149-2163.
[15] LI Yanmiao，XU Yingying，LIU Zhi，et al. Robust detection for network intrusion of industrial IoT based on multi-CNN fusion [J]. Measurement，2020，154：107450.
[16] DORIGUZZI C R，MILLAR S，SCOTT H S，et al. LUCID：A practical，lightweight deep learning solution for DDoS attack detection [J]. IEEE Transactions on Network and Service Management，2020，17(2)：876-889.
[17] VU L，BUI C T，AND NGUYEN Q U. A deep learning based method for handling imbalanced problem in network traffic classification [C]// Proceedings of the Eighth International Symposium on Information and Communication Technology. New York：ACM，2017：333-339.
[18] LOPEZ M M，SANCHEZ E A，ARRIBAS J I，et al. Supervised contrastive learning over prototype-label embeddings for network intrusion detection [J]. Information Fusion，2022，79：200-228.
[19] AL-QATF M，LASHEN Y，AL-HABIB M，et al. Deep learning approach combining sparse autoencoder with SVM for network intrusion detection[J]. IEEE Access，2018，6：52843-52856.
[20] ATEFINI R，AHMADI M. Network intrusion detection using multi-architectural modular deep neural network [J]. Journal of Supercomputing，2021，77(4)：3571-3593.