工程科学学报
Chinese Journal of Engineering

# 基于DeepInsight和迁移学习的入侵检测技术

刘文琪 胡涛 闫洁 李煌 李诗佳 葛红娟

# Network intrusion detection technology based on DeepInsight and transfer learning

LIU Wenqi, HU Tao, YAN Jie, LI Huang, LI Shijia, GE Hongjuan

引用本文:
刘文琪, 胡涛, 闫洁, 李煌, 李诗佳, 葛红娟. 基于DeepInsight和迁移学习的入侵检测技术[J]. 北科大：工程科学学报, 2024, 46(12): 2238–2245. doi: 10.13374/j.issn2095–9389.2024.03.01.002
LIU Wenqi, HU Tao, YAN Jie, LI Huang, LI Shijia, GE Hongjuan. Network intrusion detection technology based on DeepInsight and transfer learning[J]. *Chinese Journal of Engineering*, 2024, 46(12): 2238–2245. doi: 10.13374/j.issn2095–9389.2024.03.01.002

在线阅读 View online: https://doi.org/10.13374/j.issn2095–9389.2024.03.01.002

工程科学学报, 第 46 卷, 第 12 期: 2238−2245, 2024 年 12 月
Chinese Journal of Engineering, Vol. 46, No. 12: 2238−2245, December 2024
https://doi.org/10.13374/j.issn2095-9389.2024.03.01.002; http://cje.ustb.edu.cn

# 基于 DeepInsight 和迁移学习的入侵检测技术

刘文琪[1)], 胡　涛[2)], 闫　洁[3)], 李　煌[1)], 李诗佳[1)], 葛红娟[1)]✉

1) 南京航空航天大学民航学院, 南京 211106　2) 中国航空综合技术研究所, 北京 100028　3) 中国电子科技集团公司第五十四研究所, 石家庄 050081

✉通信作者, E-mail: gehongjuan1101a@nuaa.edu.cn

**摘　要**　针对入侵检测研究中, 入侵检测训练样本较少、样本不平衡等问题, 本文提出一种基于 DeepInsight 和迁移学习的入侵检测方法 DI–TL–CNN (DeepInsight–transfer learning–convolutional neural network, DI–TL–CNN). 分析采用 DeepInsight 方法将入侵数据转换为适合 CNN 模型输入的图像数据集的过程; 研究基于 VGG16 模型的训练方法, 并进一步利用迁移学习开展目标域入侵检测的过程. 通过冻结和微调 CNN 模型中不同模块参数, 比较研究了 6 种迁移方案, 并基于数据集实验研究, 获得优化方案. 采用以 UNSW-NB15 为基础的不平衡数据集作为方法验证对象, 进行网络的入侵检测分析, 验证本文提出的 DI–TL–CNN 方法的正确性; 进一步实验比较研究本文提出的方法与其他方法的检测性能, 实验结果表明, DI–TL–CNN 方法更加适用于样本较小和不平衡数据情况下的入侵检测, 其准确率和召回率等性能指标均优于其他检测方法, 具有良好的应用前景.

**关键词**　入侵检测; DeepInsight; 迁移学习; 迁移方案; 卷积神经网络

**分类号**　TP309

# Network intrusion detection technology based on DeepInsight and transfer learning

*LIU Wenqi*[1)], *HU Tao*[2)], *YAN Jie*[3)], *LI Huang*[1)], *LI Shijia*[1)], *GE Hongjuan*[1)]✉

1) College of Civil Aviation, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China

2) Avic China Aero-Polytechnology Establishment, Beijing 100028, China

3) The 54th Research Institute of China Electronics Technology Group Corporation, Shijiazhuang 050081, China

✉Corresponding author, E-mail: gehongjuan1101a@nuaa.edu.cn

**ABSTRACT**　In the dynamic field of the internet in modern life, networks are increasingly vulnerable to a diverse range of cyberattacks. Conventional intrusion detection systems based on machine learning techniques require a large number of samples for training. However, in some scenarios, only a limited number of malicious samples can be collected. To address the issue of insufficient training samples and unbalanced sample classes for intrusion detection system in real network environments, this paper proposes an intrusion detection method named DeepInsight–transfer learning–convolutional neural network (DI–TL–CNN), which is based on DI and TL. First, the DI method is used to convert the intrusion dataset into an image form suitable for CNN model input. The DI method can transform text while maintaining the semantic relationships between data points, thereby providing high-quality images. In this step, we map the 1D feature vector representation of the input data onto the 2D image representation using T-SNE and construct 2D grayscale images. In the second step, we train and optimize the VGG16 model through TL and fine-tuning, enhancing the model's adaptability and performance. We propose six TL schemes by freezing and fine-tuning the parameters of different modules in the CNN model to enhance intrusion detection performance. In the TL process, the VGG16 model, pretrained on the ImageNet dataset, demonstrates promising results for generic image classification tasks. The bottom layers of CNN models often learn basic feature patterns that are applicable to

**收稿日期:** 2024−03−01

**基金项目:** 国家自然科学基金民航联合基金重点资助项目(U2133203, U2233205)

various tasks, while the features acquired by the top layers of the model are specific to the target domain intrusion dataset. Fine-tuning allows the model to adjust the pretrained architecture's higher-order features to better match the targeted dataset. During the training process, the bottom layers of the pretrained architecture are frozen, whereas the top layers are unfrozen for fine-tuning. The optimal intrusion detection model is determined through a comparison of the performance of the six TL schemes. Finally, the correctness and effectiveness of the proposed DI–TL–CNN method are validated on a dataset with insufficient training samples, using metrics such as accuracy, precision, recall, and F1-score. In the experiments, compared with existing state-of-the-art models for intrusion detection, the proposed method considerably enhances accuracy in the detection of network traffic data. The experimental results show that the DI–TL–CNN method is suitable for intrusion detection with small samples and unbalanced data, demonstrating the good application prospects of the method in complex networks.

**KEY WORDS** intrusion detection; DeepInsight; transfer learning; transfer learning schemes; convolutional neural network

随着互联网应用和网络服务的快速发展，网络入侵问题日益严重，入侵检测系统（IDS）作为一种主动防御机制受到越来越多的关注[1−2]. IDS 通过分析收集到的网络数据来识别网络中的入侵行为，可以自动向服务器系统和网络发出警报，来保障网络安全. 近年来，基于机器学习的 IDS 方法如支持向量机（SVM）[3]、逻辑回归方法（Logistic regression）[4]、随机森林（Random forest）[5] 等，因其适应性强、智能化程度高等特点在网络入侵检测领域得到了广泛的应用. Latah[6] 为了提高入侵检测系统的准确率，采用 k-近邻算法（kNN）和极限学习机，提出基于统计方式的混合入侵检测分类系统. 随着入侵数量和数据维数的增加，深度学习开始应用于入侵检测. 在深度学习方面，现有的研究中提出了多种模型[7−8]. Wang 等[9] 提出了一种基于 HAST-IDS 的入侵检测系统，该系统利用卷积神经网络（CNN）提取网络流量的空间特征，使用递归神经网络（RNN）提取网络流量的时间特征来提高检测效率.

然而，训练数据集的大小会影响模型的准确性. 机器学习和深度学习模型需要足够的训练数据，仅依靠少量的目标域数据训练一个有效的入侵检测模型十分困难. 当训练样本数量不足时，模型会出现严重的过拟合现象. 迁移学习（TL）[10−11] 是一种加速学习过程的理想技术，将源域的知识迁移到目标域，来解决训练数据太少的问题. 与其他学习方法相比，迁移学习可以重用现有知识，以加快模型训练速度，降低计算成本，提高模型的训练性能，达到更好的入侵检测效果. 在 Pan 等[12] 的研究中，根据不同的分类标准，将迁移学习分为归纳迁移学习、无监督学习、转导迁移学习等. Ning 等[13] 为了解决检测精度较低训练样本较少的问题，分别提出了基于半监督学习（SSL）、迁移学习和域自适应（DA）的三种方法. Mehedi 等[14] 提出一种基于深度迁移学习的 ResNet 模型，用于识别少量标记数据的正常和攻击场景. Yan 等[15] 提出一种基于迁移学习和集成学习的入侵检测系统 TL-CNN-IDS，使用 VGG16、Inception 和 Xception 三个 CNN 模型对网络入侵图像集进行训练. 在 Debicha 等[16] 提出的 DNN-IDS 方法中，使用三种迁移学习技术，特征提取（FE）、微调（FT）和复制微调（DFT）. Yang 等[17] 提出一种基于迁移学习和集成学习的车联网 IDS，利用五个 CNN 模型（VGG16, VGG19, Xception, Inception 和 InceptionResnet）构建基础学习器. 并提出一种数据转换方法，将车辆网络数据转换成图像，更容易区分各种网络攻击模式. 迁移学习因为其有效性及可用性，已广泛应用于入侵检测、图像分类、故障诊断等多个领域.

近年来，将数据转换成图像用于机器学习分类任务中得到了广泛的探索. Latif 等[18] 将网络流量数据转换为图像，以便更好地检测攻击和可疑流量，利用遗传算法对每个基础学习模型的超参数进行微调，并使用集成技术融合表现最好的模型输出. Dunmore 等[19] 基于 GAN 图像增强的卷积神经网络（MAGNETO），引入 DeepInsight 方法，提出了一种保留不同特征之间关系和语义信息的新型图像方法，对网络流量数据进行分类. DeepInsight 方法是 Sharma 等[20] 于 2019 年提出，将非图像数据转换成图像的方法，并将 CNN 模型对于图像处理的优势应用于非图像数据集. 该方法通过自动提取特征，能够减少对神经元的需求，从而更深入地训练模型.

本文针对入侵检测数据数量不足的情况，利用 DeepInsight 方法和迁移学习，提出一种新的入侵检测模型 DI–TL–CNN. 首先，利用 DeepInsight 方法对入侵检测数据集进行特征提取，获得特征矩阵，将网络流量数据转换为图像. 其次，利用迁移学习将 CNN 模型的底层参数迁移到目标域的入侵检测任务中，通过冻结和微调不同模块参数，研

究基于 VGG16 模型的训练方法, 提出 6 种迁移学习方案. 探讨不同迁移学习方案模型性能的变化, 并基于数据集实验研究, 获得优化方案. 论文采用以 UNSW-NB15 为基础的不平衡数据集作为方法验证对象, 在具有不平衡度不同的小样本数据集上开展比较研究, 验证本文所提方法在小样本数据集场景下检测结果的准确性.

## 1　DI–TL–CNN 模型框架

本文提出一种基于 DeepInsight 和迁移学习的 DI–TL–CNN 模型, 考虑网络入侵数据稀缺性的情况, 结合 DeepInsight 图像转换方法和迁移学习对模型的权值共享, 实现在较少的入侵样本下建立性能较好的入侵检测模型.

### 1.1　DI–TL–CNN 模型入侵检测方法

如图 1 所示, 为 DI–TL–CNN 模型的工作流程, 主要包含四部分: 源域数据 $\boldsymbol{D}_s=(\boldsymbol{X}_s, \boldsymbol{Y}_s)$, $\boldsymbol{X}_s$ 为源域特征, $\boldsymbol{Y}_s$ 为源域标签; 源任务 $\boldsymbol{T}_s=(\boldsymbol{Y}_s, \boldsymbol{f}_s)$, $\boldsymbol{f}_s$ 为源域数据检测模型; 目标域数据 $\boldsymbol{D}_t=(\boldsymbol{X}_t, \boldsymbol{Y}_t)$, $\boldsymbol{X}_t$ 为目标域特征, $\boldsymbol{Y}_t$ 为目标域标签; 目标任务 $\boldsymbol{T}_t=(\boldsymbol{Y}_t, \boldsymbol{f}_t)$, $\boldsymbol{f}_t$ 为目标域数据检测模型. 选择 ImageNet 图像集作为源域数据集, 使用 VGG16 模型[21] 作为源域模型, 为 CNN 模型提供底层特征的学习能力. 本文选择 UNSW-NB15 数据[22] 作为目标域数据集. 首先对目标域数据集 UNSW-NB15 进行数据预处理, 采用 DeepInsight 将网络数据集转换为目标域网络图像集, 基于二维 CNN 模型用于图像数据的训练, 将其最大程度发挥 CNN 模型提取非图像数据集特征的能力. 然后, 使用 VGG16 模型对生成的图像集进行训练. 利用迁移学习将 CNN 模型的底层参数迁移到目标任务模型中实现权值共享, 目标域网络图像集提供目标任务模型的顶层特征.

迁移学习在许多图像处理任务中得到应用, 在 CNN 模型的底层学习到的特征模式通常是适用于许多不同任务的通用模式. 基于 DI–TL–CNN 模型的方法利用迁移学习将 CNN 模型的底层参数迁移到目标任务模型中, 目标域网络图像集提供目标任务模型的顶层特征. 基于迁移学习的 DI–TL–CNN 模型方法的步骤:

(1) 对目标域网络数据集进行归一化、独热编码等预处理.

(2) 经过预处理后, 将数据集输入 DeepInsight 进行图像转换, 得到目标域网络图像集.

(3) 利用迁移学习方法, 将在源域 $\{\boldsymbol{D}_s, \boldsymbol{T}_s\}$ 上学习到的 CNN 模型隐藏层的权重 $\boldsymbol{W}_j$ 和偏置 $\boldsymbol{B}_j$ 以及超参数等迁移到目标域 $\{\boldsymbol{D}_t, \boldsymbol{T}_t\}$.

(4) 通过冻结和微调不同的模块参数, 训练 CNN 模型, 构建 DI–TL–CNN 模型, 进行网络入侵检测.

### 1.2　基于 DeepInsight 的数据-图像转换

基于二维 CNN 的模型用于图像数据的训练,

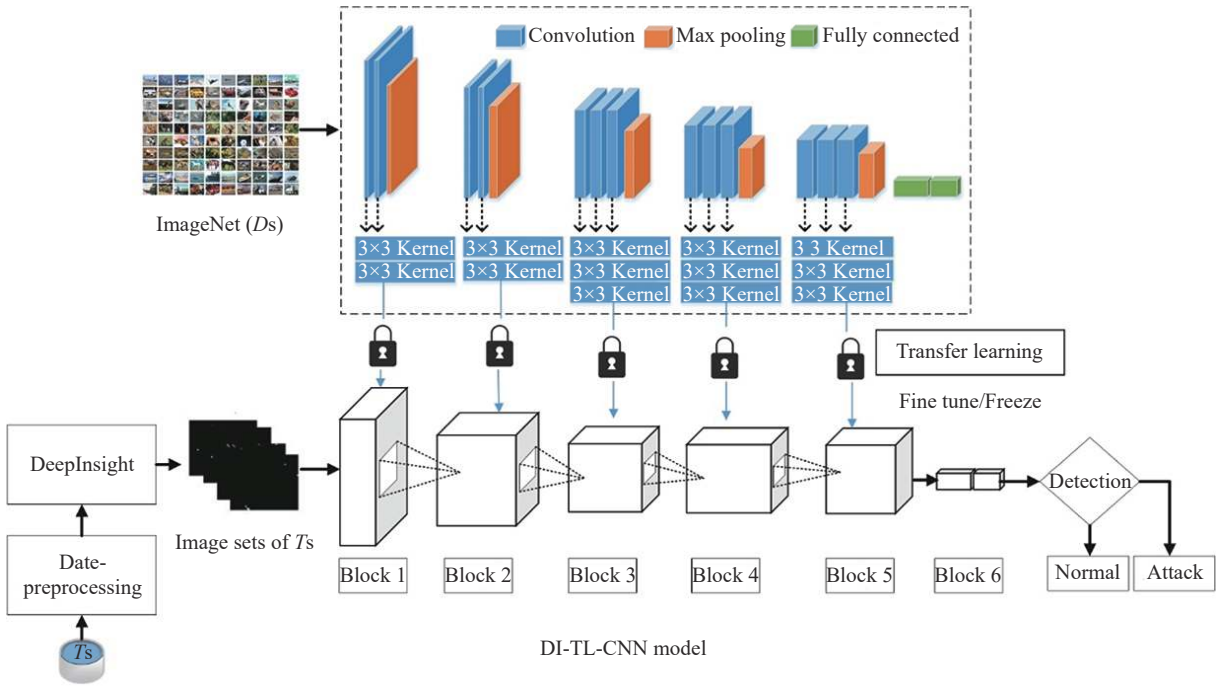


**图 1**　DI–TL–CNN 模型流程

**Fig.1**　Flowchart of the DI–TL–CNN model

因此, 本文利用 DeepInsight 方法对目标域数据进行图像转换, 转换过程如图 2 所示, 该方法步骤如下:

(1) 使用降维技术 t-SNE(t-distributed stochastic neighbor embedding)[23] 将输入特征向量转换为特征矩阵.

t-SNE 是一种非线性技术, 能将高维数据可视化在二维笛卡尔空间. 利用 t-SNE 将数据集 $\boldsymbol{T}(\boldsymbol{T}\in\mathbf{R}^{M\times N}$, $\boldsymbol{M}$ 表示训练样本, $\boldsymbol{N}$ 表示一维特征向量集合 $\boldsymbol{X}^{1D}$) 转置得到 $\boldsymbol{T}'(\boldsymbol{T}'\in\mathbf{R}^{N\times 2})$. t-SNE 算法在高维空间构建网络流量概率分布, 之后在低维空间构建网络流量的概率分布. 对任意两个网络流量特征 $\boldsymbol{X}_i$ 与 $\boldsymbol{X}_j$ 相似性的条件概率进行评价, 相似度高的特征赋予较高的概率, 否则赋予较低的概率, 用非对称 K–L 散度对相似性进行评价:

$$D_{\mathrm{KL}}(P\|Q)=\sum_{i\neq j}p(X_{ij})\cdot\log\frac{p(X_{ij})}{q(X_{ij})} \tag{1}$$

其中,

$$p(X_{ij})=\frac{\exp(-\|x_i-x_j\|^2/2\sigma^2)}{\sum_{k\neq l}\exp(-\|x_k-x_l\|^2)/2\sigma^2} \tag{2}$$

$$q(X_{ij})=\frac{(1+\|y_i-y_j\|^2)^{-1}}{\sum_{k\neq l}(1+\|y_i-y_j\|^2)^{-1}} \tag{3}$$

式中, $p(X_{ij})$为在高维空间中, 样本点 $x_i$ 和 $x_j$ 的相似性条件概率; $q(X_{ij})$为在低维空间中, 映射后的点 $y_i$ 和 $y_j$ 的相似性条件概率. $\sigma$ 是以样本点 $x_i$ 为中心的高斯分布的方差.

(2) 使用凸包算法找到包含所有元素的最小矩形, 并旋转对齐图像. 使用凸包算法找到包含通过 t-SNE 变换与 $\boldsymbol{X}^{1D}$特征相关联的点的最小边界矩形, 将其旋转到水平或垂直方向, 将特征点框架到二维笛卡尔平面上.

(3) 将元素值映射到像素位置. 将特征值与像素坐标关联起来, 构建特征向量的图像集. 旋转后的矩形可以用 min($x$)、max($x$)、min($y$) 和 max($y$) 表示(沿着二维笛卡尔坐标 $x$ 和 $y$ 轴的最小和最大坐标), 分割成一个二维网格 $m\times n$ 大小相等的矩形像素帧, 长度 $\frac{\max(x)-\min(x)}{m}$ 和宽度 $\frac{\max(y)-\min(y)}{n}$.

(4) 创建图像集. 每个样本的单个特征值被用作像素的灰度值. 将训练样本从一维特征向量形式转换为二维图像形式构建正常样本和攻击样本的二维灰度图, 生成图像集.

### 1.3 迁移学习和迁移方案

VGG16 模型是性能较好的图像分类模型, 既能作为源域数据集的分类器, 又能作为目标域模型的特征提取器. 本文选择 VGG16 模型作为迁移学习的模型, 通过迁移学习将源域模型的参数迁移到目标域模型. 如图 3 所示, 为 VGG16 模型网络架构图, 13 个卷积层、5 个池化层和 3 个全连接层. 将 VGG16 模型分为 6 个模块, 依次为 Block1-Block6.

为了得到性能更好的入侵检测模型, 分别比较可能的 6 种迁移方案, 进行训练微调, 其处理过程如下:

(1) 确定可能的多种迁移方案, 它们分别是:

迁移方案 1: 冻结全部卷积模块 Block1 ~ Block5, 训练微调模块 Block6 参数;

迁移方案 2: 冻结部分卷积模块 Block1 ~ Block4, 训练微调 Block5 ~ Block6 参数;

迁移方案 3: 冻结部分卷积模块 Block1 ~ Block3, 训练微调 Block4 ~ Block6 参数;

迁移方案 4: 冻结部分卷积模块 Block1 ~ Block2, 训练微调 Block3 ~ Block6 参数;

迁移方案 5: 冻结卷积模块 Block1, 训练微调 Block2 ~ Block6 参数;

迁移方案 6: 不冻结任何模块, 训练微调 Block1 ~ Block6 参数.

图 4(a) ~ (c) 分别给出了第 1 ~ 3 种迁移方案, 其他迁移方案以此类推.

(2) 针对 6 种迁移方案, 分别进行微调训练, 紫色模块表示冻结部分, 绿色模块表示微调训练部分.

冻结部分的参数利用迁移学习将源域模型的

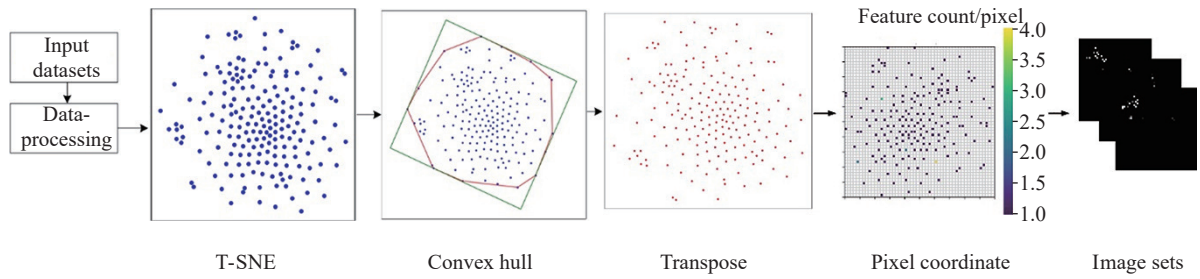


**图 2** DeepInsight 方法将数据转换成图像

**Fig.2** DeepInsight method converting the data into image sets

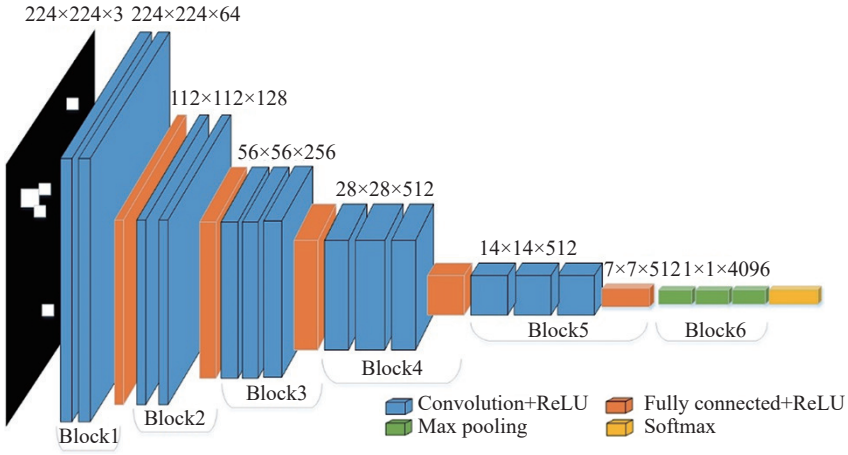


**图 3**　VGG16 网络架构

**Fig.3**　VGG16 network architecture

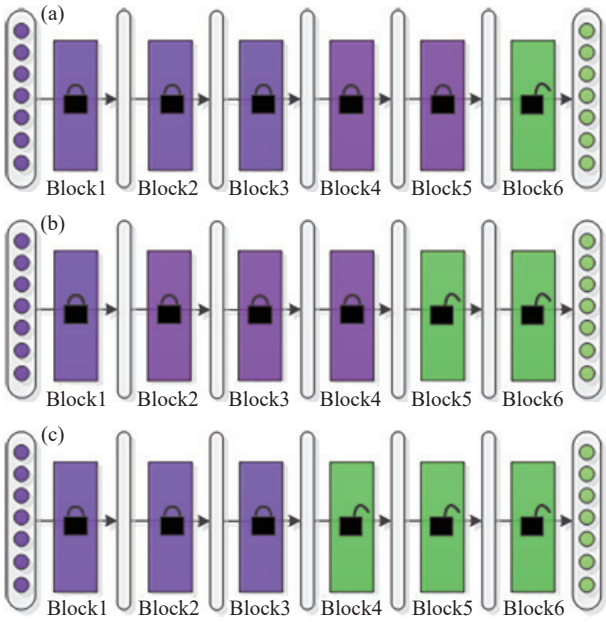


**图 4**　迁移方案 1 (a)、2 (b) 和 3 (c)

**Fig.4**　Transfer learning schemes 1 (a), 2 (b), and 3(c)

底层参数迁移到目标域模型, 输入目标域数据集训练微调模块的参数, 将其入侵检测模型更适用于目标域数据集.

(3)观察训练模型的评价指标 Accuracy, Recall, Precision 和 F1-score, 分析 6 种迁移方案的评价指标, 选择优化方案; 具体训练过程和结果见本文 3.1 节.

## 2　数据来源与评价指标

### 2.1　数据来源与预处理

本文的实验数据集采用 UNSW-NB15, UNSW-NB15 数据集由澳大利亚网络安全中心通过模拟实验平台获得, 模拟网络正常和攻击行为生成数据集. UNSW-NB15 数据集包含正常流量和 9 种最新攻击行为, 可以很好地反映现代网络流量模式. 为了验证模型在小样本和样本不平衡条件下的表现, 按照正常数据与攻击数据 2∶1 的比例构建小样本不平衡数据集 A; 按照正常数据与攻击数据 8∶2 构建小样本不均衡程度更高的数据集 B, 表 1 列出了数据集的基本情况.

**表 1**　UNSW-NB15 数据集

**Table 1**　UNSW-NB15 dataset

| Type | Dataset A | | Dataset B | |
|---|---|---|---|---|
| | Training set-A | Testing set-A | Training set-B | Testing set-B |
| Normal | 2000 | 1000 | 2000 | 1000 |
| Attack | 1000 | 500 | 500 | 250 |
| Total | 3000 | 1500 | 2500 | 1250 |

对 UNSW-NB15 数据集进行数据预处理, 步骤如下:

(1) 独热编码.

通过独热编码技术将符号特征转为二进制数字.

(2) 数据归一化.

为消除各特征间数量级的差异对结果产生的影响, 需要将特征值归一化到 [0, 1] 之间.

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{4}$$

式中, $x'$表示归一化后的值, $x$ 表示初始特征值, $x_{\min}$表示该属性中的最小特征值, $x_{\max}$表示该属性中的最大特征值.

### 2.2　评价指标

本文采用的评价指标, 包括准确率 (Accuracy)、召回率 (Recall)、精确度 (Precision) 和 F1-score. Accuracy 是指预测为正的样本占总样本数的比例, 反应模型的整体预测能力. Recall 是指预测为正的样

本占实际为正的样本的比例. Precision 是实际为正的样本占预测为正的样本的比例. 这两项可以反映模型在假阳性和假阴性方面的分类性能. F1-score 是召回率和准确率的调和平均值. 评价标准计算为如式 (5) ~ (8) 所示:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{5}$$

式中, TP 为真阳性, 表示实际和预测均为攻击的样本数; TN 为真阴性, 表示实际和预测均为正常的样本数; FP 为假阳性, 表示实际是正常而预测为攻击的样本数; FN 为假阴性, 表示实际为攻击而预测为正常的样本数.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{6}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{7}$$

$$\text{F1-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{8}$$

## 3 算法验证

### 3.1 基于 DI–TL–CNN 模型的不同迁移方案比较

如表 2 所示, 为本文针对 1.3 节提出的基于 DI–TL–CNN 模型的 6 种迁移方案的在数据集 A 的检测结果.

**表 2** 6 种迁移方案的检测结果

**Table 2** Results of the six transfer learning schemes

| Transfer scheme | Freeze block | Fine-tuning block | Accuracy/% | Recall/% | Precision/% | F1-score/% |
|---|---|---|---|---|---|---|
| 1 | 1, 2, 3, 4, 5 | 6 | 88.80 | 93.90 | 84.90 | 89.17 |
| 2 | 1, 2, 3, 4 | 5, 6 | 92.13 | 95.60 | 92.82 | 94.19 |
| 3 | 1, 2, 3 | 4, 5, 6 | 92.73 | 95.58 | 92. 61 | 94.64 |
| 4 | 1, 2 | 3, 4, 5, 6 | 93.40 | 96.00 | 94.12 | 95.10 |
| 5 | 1 | 2, 3, 4, 5, 6 | 94.47 | 96.10 | 94.30 | 95.11 |
| 6 | — | 1, 2, 3, 4, 5, 6 | 90.48 | 94.80 | 93.46 | 94.62 |

由表 2, 从迁移方案 1 至迁移方案 5 的检测结果可以看出, 随着微调模块数量的增多, 观察训练模型的评价指标 Accuracy、Recall、Precision、F1-score 都呈现上升的趋势, 这说明微调的参数越多, 越能很好地学习入侵数据集的特征, 模型的各方面的性能都得到了提升. 在迁移方案 6 中, Block1 ~ Block6 全部微调时, 比迁移方案 5 的准确率低 1.07%, 精确度低 0.18%. 说明 DI–TL–CNN 模型通过利用 DeepInsight 技术将入侵数据转换为图像寻找源域与目标域数据集之间的相似性, 将源域模型的一些能力和知识迁移到目标域模型, 迁移学习的实现优于随机权值初始化, 微调方法结合迁移学习能提高模型性能.

如图 5 所示, 为 6 种迁移方案准确率的比较. 迁移方案 5 比其他迁移方案的准确率更高, 收敛速度更快, 更稳定. DI–TL–CNN 模型微调方法结合迁移学习, 保留源域的底层特征, 随着微调模块数量的增多, 准确率呈现上升的趋势, 相比之下, 微调的模块越多, 准确率值越高.

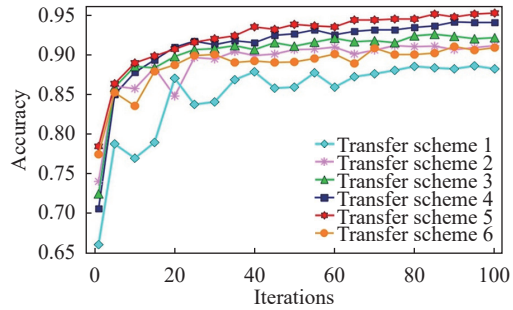


**图 5** 6 种迁移方案准确率比较

**Fig.5** Comparison of the accuracy of the six transfer schemes

如图 6 所示, 为 6 种迁移方案的 Loss 值比较. 整体而言, 6 种迁移方案的 Loss 值下降趋势相同, 随着 Epoch 数的增加, 6 种迁移方案的 Loss 值逐渐下降. 迁移方案 5 比其他几种迁移方案 Loss 值低, 性能最好. 这说明迁移方案 5 通过冻结 Block1, 微调 Block2 ~ Block6, 迁移学习的实现优于目标模型随机权值初始化, 仅需要较少的样本量就能训练出较好的入侵检测模型.

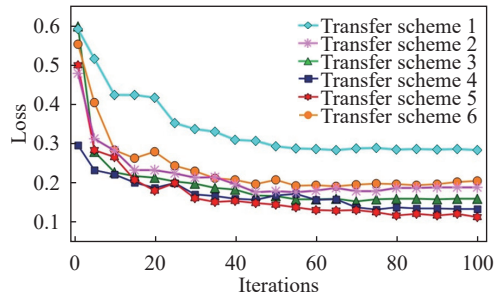


**图 6** 6 种迁移方案 loss 值比较

**Fig.6** Comparison of loss values of the six transfer schemes

以上分析表明, 迁移方案 5 通过冻结 Block1 模块, 微调 Block2 ~ Block6 模块得到的入侵检测模型综合评价指标最高, 性能最好, 选择该方案模型作为最后的方法.

### 3.2 DI–TL–CNN 模型性能验证

DI–TL–CNN 模型在数据集 A 和数据集 B 的性能, 如表 3 所示. DI–TL–CNN 模型在数据集 A、数

**表 3**　DI–TL–CNN 模型在不同数据集的比较

**Table 3**　Comparison of DI–TL–CNN performance across different datasets

| Dataset Type | Accuracy/% | Recall/% | Precision/% | F1-score/% |
|---|---|---|---|---|
| Dataset-A | 94.47 | 96.10 | 94.30 | 95.11 |
| Dataset-B | 94.25 | 95.12 | 94.57 | 94.84 |

据集 B 中的训练测试准确率和 Loss 值随迭代步数如图 7(a)、7(b) 所示.

本文在抽样设计的小样本数据集和小样本不均衡数据集上进行了对比实验, 在数据集 A 中, 基于 DI–TL–CNN 模型的检测准确率为 94.47%; 数据集 B 中, 该模型的检测准确率为 94.25%. 实验结果表明, 小样本不均衡程度较高的数据集上, 基于 DI–TL–CNN 模型依旧有较高的分类准确性. DeepInsight 方法既能转换文本又能保持数据点之间语义关系, 将数据转换成图像后, 为迁移学习提供处理良好的数据, 能较好的区分正常和攻击类别.

### 3.3　DI–TL–CNN 方法与其他方法比较

本文开展了 DI–TL–CNN 方法与其他方法的比较研究, 结果如表 4 所示. 与其他方法相比, DI–TL–CNN 方法的分类检测结果较好, 取得了比现阶段一些先进方法更好的实验结果, 具有更高的准确率. 在 Precision、Recall 和 F1-score 方面, DI–TL–CNN 模型与其他模型相比也表现较好. 由此表明, DI–TL–CNN 模型应用 DeepInsight 和迁移学习具有较大的优势: DeepInsight 方法保留原始数据的结构信息, 可以更好的捕捉图像的高级特征, 提供质量较高的图像; 迁移学习将参数迁移到目标域模型, 实现参数共享减少对训练样本的依赖. 总体而言, 在入侵检测样本量较小的情况下, 本文提出的 DI–TL–CNN 模型入侵检测性能更佳, 总体评价指标更均衡, 具有较高的应用价值.

## 4　结论

为提高入侵检测模型的性能, 本文提出一种基于 DeepInsight 方法和迁移学习的入侵检测模型 DI–TL–CNN.

(1) DI–TL–CNN 模型充分利用 DeepInsight 方法既能转换文本又能保持数据点之间语义关系的

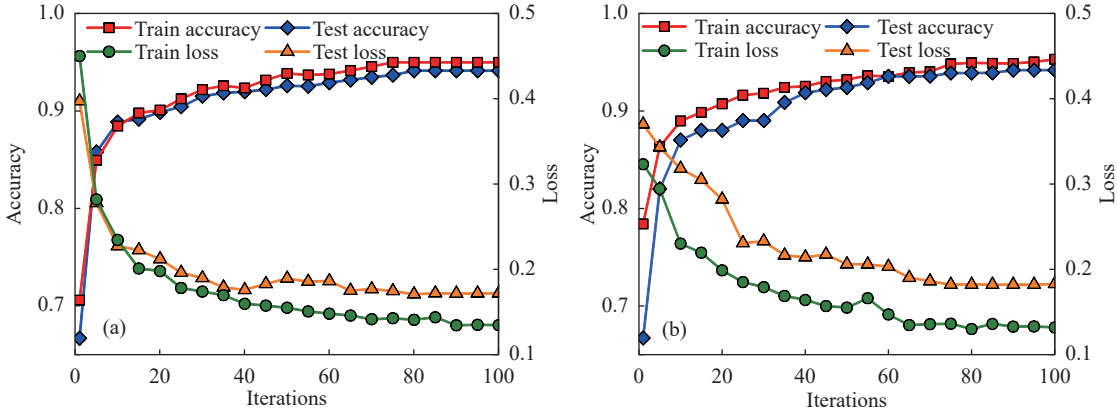


**图 7**　在不同数据集下 DI–TL–CNN 的性能. (a) 数据集 A; (b) 数据集 B

**Fig.7**　Performance of DI–TL–CNN on difference datasets: (a) dataset A; (b) dataset B

**表 4**　不同入侵检测方法的比较

**Table 4**　Comparison of different intrusion detection methods

| Method | Accuracy/% | Recall/% | Precision/% | F1-score/% |
|---|---|---|---|---|
| CNN | 90.84 | 89.25 | 90.95 | 92.07 |
| SVM | 88.75 | 69.52 | 74.56 | 71.90 |
| Logistic regression | 83.00 | 94.20 | 82.70 | 88.03 |
| kNN | 82.73 | 91.30 | 84.15 | 87.62 |
| Random forest | 92.07 | 93.40 | 94.63 | 94.01 |
| MAGNETO-GAN[24] | 89.73 | — | — | 91.97 |
| GMM-WGAN-IDS[25] | 87.70 | 87.70 | 88.46 | 85.40 |
| RF-RFE-ensemble[26] | 89.91 | 86.82 | 85.78 | 86.29 |
| **DI–TL–CNN** | **94.47** | **96.10** | **94.30** | **95.14** |

能力, 为迁移学习提供良好的输入图像集; 同时, 本文方法充分利用了 CNN 模型在图像分类领域的优势, 为提高网络入侵检测的准确率提供了另一种思路.

(2)通过微调方式训练 DI–TL–CNN 模型参数, 训练模型的大多数层被冻结, 保留模型的底层特征, 实现权重共享, 同时解冻顶层, 进行参数微调, 能够降低模型的计算损失, 提高迁移性能; 该方法可以扩展应用到其他研究领域的迁移学习中.

(3)实验结果表明, 基于 DI–TL–CNN 模型的入侵检测对较小样本和不平衡样本, 具有较高的分类准确性, 模型的综合性能较好. 本文提出的 DI–TL–CNN 模型训练和构建方法也适用于其他研究领域和场景.

## 参 考 文 献


[1] Lin Y D, Wang Z Y, Lin P C, et al. Multi-datasource machine learning in intrusion detection: Packet flows, system logs and host statistics. *J Inf Secur Appl*, 2022, 68: 103248

[2] Li H, Ge H J, Yang H Q, et al. An abnormal traffic detection model combined BiIndRNN with global attention. *IEEE Access*, 2022, 10: 30899

[3] Kalita D J, Singh V P, Kumar V. A novel adaptive optimization framework for SVM hyper-parameters tuning in non-stationary environment: A case study on intrusion detection system. *Expert Syst Appl*, 2023, 213: 119189

[4] Besharati E, Naderan M, Namjoo E. LR-HIDS: Logistic regression host-based intrusion detection system for cloud environments. *J Ambient Intell Humaniz Comput*, 2019, 10(9): 3669

[5] Pal M. Random forest classifier for remote sensing classification. *Int J Remote Sens*, 2005, 26(1): 217

[6] Latah M, Toker L. An efficient flow-based multi-level hybrid intrusion detection system for software-defined networks. *CCF Trans Netw*, 2020, 3(3): 261

[7] Cao C, Xie L, Li L P, et al. Intrusion detection techniques of variable-frequency vector control system. *Chin J Eng*, 2019, 41(8): 1074
(曹策, 解仑, 李连鹏, 等. 变频矢量控制系统入侵检测技术. 工程科学学报, 2019, 41(8): 1074)

[8] Li H, Sang Y Q, Ge H J, et al. Anomaly detection of aviation data bus based on SAE and IMD. *Comput Secur*, 2024, 137: 103619

[9] Wang W, Sheng Y Q, Wang J L, et al. HAST-IDS: Learning hierarchical spatial-temporal features using deep neural networks to improve intrusion detection. *IEEE Access*, 1806, 6: 1792

[10] Zhang W, Liu C, Fei H B, et al. Research on automatic speech recognition based on a DL-T and transfer learning. *Chin J Eng*, 2021, 43(3): 433
(张威, 刘晨, 费鸿博, 等. 基于 DL-T 及迁移学习的语音识别研究. 工程科学学报, 2021, 43(3): 433)

[11] Layeghy S, Baktashmotlagh M, Portmann M. DI-NIDS: Domain invariant network intrusion detection system. *Knowl-Based Syst*, 2023, 273: 110626

[12] Pan S J, Yang Q. A survey on transfer learning. *IEEE Trans Knowl Data Eng*, 2010, 22(10): 1345

[13] Ning J H, Gui G, Wang Y, et al. Malware traffic classification using domain adaptation and ladder network for secure industrial Internet of Things. *IEEE Internet Things J*, 2022, 9(18): 17058

[14] Mehedi S T, Anwar A, Rahman Z, et al. Dependable intrusion detection system for IoT: A deep transfer learning based approach. *IEEE Trans Ind Inform*, 2023, 19(1): 1006

[15] Yan F R, Zhang G H, Zhang D W, et al. TL-CNN-IDS: Transfer learning-based intrusion detection system using convolutional neural network. *J Supercomput*, 2023, 79(15): 17562

[16] Debicha I, Bauwens R, Debatty T, et al. TAD: Transfer learning-based multi-adversarial detection of evasion attacks against network intrusion detection systems. *Future Gener Comput Syst*, 2023, 138: 185

[17] Yang L, Shami A. A transfer learning and optimized CNN based intrusion detection system for Internet of vehicles // *IEEE International Conference on Communications* (*ICC* 2022). Seoul, 2022: 2774

[18] Latif S, Boulila W, Koubaa A, et al. DTL-IDS: An optimized intrusion detection framework using deep transfer learning and genetic algorithm, *J Netw Comput Appl*, 2024, 221, 103784

[19] Dunmore A, Dunning A, Jang-Jaccard J, et al. MAGNETO and DeepInsight: Extended image translation with semantic relationships for classifying attack data with machine learning models. *Electronics*, 2023, 12(16): 3463

[20] Sharma A, Vans E, Shigemizu D, et al. DeepInsight: A methodology to transform a non-image data to an image for convolution neural network architecture. *Sci Rep*, 2019, 9(1): 11399

[21] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition // *International Conference on Learning Representations*. Canada, 2014: 1409

[22] Moustafa N, Slay J. UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set) //*IEEE Military Communications and Information Systems Conference* (*MilCIS*). Canberra, 2015: 1

[23] Van der Maaten L, Hinton G. Viualizing data using T-SNE. *J Mach Learn Res*, 2008, 9: 2579

[24] Andresini G, Appice A, De Rose L, et al. GAN augmentation to deal with imbalance in imaging-based intrusion detection. *Future Gener Comput Syst*, 2021, 123: 108

[25] Cui J Y, Zong L S, Xie J H, et al. A novel multi-module integrated intrusion detection system for high-dimensional imbalanced data. *Appl Intell*, 2023, 53(1): 272

[26] Abbas Q, Hina S, Sajjad H, et al. Optimization of predictive performance of intrusion detection system using hybrid ensemble model for secure systems. *PeerJ Comput Sci*, 2023, 9: 1552