文章编号:1671-4598(2022)03-0259-07 DOI:10.16526/j.cnki.11-4762/tp.2022.03.043 中图分类号:TP391 文献标识码:A

卷积自编码器融合核近似技术的异常检测模型

武玉坤[1,2], 李 伟[2], 陈沅涛[3]
(1. 浙江邮电职业技术学院, 浙江 绍兴 312366; 2. 浙江工业大学 计算机科学与技术学院, 杭州 310023;
3. 长沙理工大学 计算机与通信工程学院, 长沙 410114)

摘要: 图像中的异常检测是计算机视觉中非常重要的研究主题, 它可以定义为单分类问题; 针对图像数据集的规模大, 维度高等特性, 一种新的深度卷积自编码器 (CAE, convolutional autoencoder) 与核近似单分类支持向量机 (OCSVM, one class support vector machine) 相结合的异常检测模型 CAE-OCSVM 被提出; 模型中的深度卷积自编码器负责学习图像的本质特征表示, 然后使用随机傅里叶特征对卷积自编码器学习到的本质特征进行核近似, 核近似后输入线性单类支持向量机进行图像异常检测; 核近似技术克服了核学习技术时间复杂度高的问题; 同时深度卷积自编码器与核近似单类支持向量机通过梯度下降法实现了端到端的学习; 模型的 AUC 性能在 4 个公开的图像基准数据集上进行了实验验证, 同时模型与其它常用的异常检测模型在不同的异常率的情况下进行了性能对比; 实验结果证实 CAE-OCSVM 模型在 4 个公开图像数据集上的性能都优于其它异常检测模型, 表明了 CAE-OCSVM 模型更适合大规模高维数据集的异常检测。

关键词: 异常检测; 卷积自编码器; 核近似; 单分类支持向量机; 随机傅里叶特征

Anomaly Detection Model of Convolutional Autoencoder Combined with Kernel Approximation Technology

WU Yukun[1,2], LI Wei[2], CHEN Yuantao[3]
(1. Zhejiang Post and Telecommunication College, Shaoxing 312366, China;
2. College of Computer Science and Technology, Zhejiang University of Technology, Hangzhou 310023, China;
3. School of Computer and Communication Engineering, Changsha University of Science and Technology, Changsha 410114, China)

Abstract: Anomaly detection in images is a very important research topic in computer vision. It can be defined as singal classification problem; for the large-scale and high-dimensional characteristics of image data sets, a novel anomaly detection model for Convolutional Autoencoder and One Class Support Vector Machine (CAE-OCSVM) is proposed, which is a combination of deep CAE and a kernel approximate OCSVM. The deep CAE in the model is responsible for learning the essential feature representation of the image, and then random Fourier features are used to perform kernel approximation to the essential features learned by the CAE. After the kernel approximation, the linear OCSVM performs anomaly detection on the image, and the kernel approximation technology overcomes the problem of high time complexity of kernel learning technology. Meanwhile the CAE and the kernel approximation OCSVM achieve the end-to-end learning through the gradient descent method. The AUC performance of the model is tested and verified on four public image benchmark data sets. At the same time, AUC performance is compared with other commonly used anomaly detection models under different anomaly rates. Experimental results show that the performance of the CAE-OCSVM model on the four public image data sets is better than that of other anomaly detection models, indicating that the CAE-OCSVM model is more suitable for large-scale high-dimensional data set anomaly detection.

Keywords: anomaly detection; convolutional autoencoder; kernel approximation; singal class support vector machine; random fourier feature

0 引言

异常检测是模式识别领域中众所周知的问题, 在正常或预期模式已知的情况下, 分类器要识别的异常模式都是在训练集中稀缺或不存在的模式数据[1]。它可以定义为一类分类 (OCC, one-class classification) 问题, 图像中异常行

收稿日期:2022-01-04; **修回日期**:2022-02-11。
基金项目:国家自然科学基金(61502422,61972056),浙江省自然科学基金(LY18F020028);浙江省科技厅公益项目(2017C33108;浙江省教育厅一般科研项目(Y202044619)。
作者简介:武玉坤(1980-),男,山东临沂人,博士研究生,副教授,主要从事机器学习,数据挖掘等方向的研究。
李 伟(1958-),男,吉林长春人,博士,教授,博士生导师,国家"千人计划"专家,主要从事大数据、智慧城市方向的研究。
陈沅涛(1980-),男,湖南邵阳人,博士,副教授,硕士生导师,主要从事人工智能、图像处理方向的研究。
引用格式:武玉坤,李 伟,陈沅涛.卷积自编码器融合核近似技术的异常检测模型[J].计算机测量与控制,2022,30(3):259-265,276.

为的分类一直是一个非常有趣的话题，近年来许多研究都集中在检测异常图像和自动视频监控中的事件[2-3]。

通常，异常检测算法是通过学习数据的特征表示，然后把那些远离这个框架模型的数据视为异常；具有大量实例（训练集）的已知类别（正常模式）被称为正类或目标类。大量方法已被广泛用于异常检测问题。例如基于邻居的模型、基于统计的模型和基于深度学习的模型。Xu 和 Ricci 使用单分类支持向量机来预测视频中的异常帧[4]。Kim 和 Grauman 设计了一种用于检测异常活动的单类分类方法，称为时空马尔可夫随机场模型[5]。在文献［6］中，为 OCC 学习了一个概率模型，分类器是无监督的，不需要标记训练数据。在文献［7］中，训练无监督的深度信念网络（DBN，deep belief network）在相对低维空间中提取一组特征，并用 DBN 学习的特征训练一类分类器。一般来说，一类分类器对于大型和高维数据集中的决策表面建模效率低下，但是通过将分类器与 DBN 相结合，可以减少冗余特征并提高 OCC 的性能。尽管在过去几年异常检测取得了丰硕的进展，但随着数据维度的增长，异常检测就越困难，原因在于任何一个异常样本都可能是一个罕见的实例，被观察到的概率也就越低。在没有人工监督的情况下对多维或高维数据进行稳健的异常检测仍然是一项具有挑战性的任务。因此需要有效的方法来检测大规模高维数据中的异常，并要对不同级别的数据噪声具有很强的鲁棒性。解决该问题广泛采用的方案是降低输入的维数，然后在相应的低维空间中进行检测异常。特征降维的方法有很多，比如主成分分析（PCA，principal component analysis）、核 PCA[8]、自编码器（AE，auto-encoder）、非负矩阵分解（NMF，non-negative matrix factorization）、遗传编程、DBN 等降维方法。其中，PCA 和 AE 是两种常用的方法。基于 PCA 的降维方法保留特征值较大的数据信息，丢弃数据信息中具有较小的特征值。AE 是一种基于深度学习的数据降维方法，虽然基于 AE 的降维方法有更多的计算成本，但它已被证明在异常检测中进行特征降维非常成功。深度学习模型已经在图像和视频的监督分类中实现了最优的目标识别性能；之所以有如此高的性能的主要是因为他们可以自动提取特征，并具有卓越的图像表示判别能力。文献［9］中堆叠降噪自动编码器（SDAE，stacked denoising auto-encoders）仅使用正类进行训练单类分类器取得了良好的性能。然而，当直接应用于 OCC 问题时，SDAE 可能效率低下，因为自编码器重构层的特征空间映射可能是稀疏的，它不能保证重构层中数据的紧凑表示，在训练过程中，SDAE 的重构层是最小化所有实例的重构误差，而不是试图在特征空间中实现紧凑表示，这个问题阻碍了正类与负类的界定并且减弱了分类性能。

OCSVM 是广泛使用的有效用于识别异常的无监督技术。但是 OCSVM 在复杂的高维数据集上的性能不是最优的，所以很多文献已经提出了使用特征选择和特征提取方法来处理复杂的高维数据，以便与 OCSVM 相结合来实现异常检测。在深度学习取得前所未有的成功之后，许多结合使用深度特征提取学习和 OCSVM 的混合模型已经出现。使用自动编码器作为特征提取器，其中隐藏层表示作为传统异常检测算法（如 OCSVM）的输入。由于混合模型使用自动编码器提取深度特征，然后将其提供给单独的异常检测方法，它们无法影响隐藏层中的表征学习。从某种意义上说，特征学习与异常检测任务无关，并且不是为异常检测而定制的。也就是说深度学习的表示学习与 OCSVM 的异常检测是解耦的，不能实现端到端的学习。另外一个限制是由于 OCSVM 是一种核学习方法，常用的核函数是高斯核函数，由于核矩阵（Gram）的计算，其计算复杂度随样本量的增加呈平方关系增长。针对此限制的解决方案是采用核近似方法，常用的核近似方法有 Nystrom 方法[10]与随机傅里叶核近似方法（RFF，random fourier features）[11]。Nystrom 方法是得到核矩阵的一个低秩矩阵；随机傅里叶核近似方法是使用随机特征映射显式地将数据映射到一个欧氏内积空间。在平移不变核的情况下，可以使用随机傅里叶特征来设计大规模的核机器学习方法。在文献［11］中，作者提出了利用随机特征来逼近核的方法，而不是使用隐式特征映射。其思想是使用随机特征映射明确地将数据映射到一个欧氏内积空间，就可以使用欧氏内积来进行核近似。Huang 等[12]表明，在语音识别中，用随机傅里叶特征进行核近似，浅层核机器学习与深度学习的性能相匹配，同时具有较高的计算效率。

随着当前大数据时代的到来，各种数据集规模越来越大，未知的异常类型越来越多样化，异常检测模型的准确性、鲁棒性以及高效性仍是首要考虑的因素。本文提出了基于深度卷积自编码器与核近似单分类支持向量机相结合的端到端的图像异常检测模型，深度卷积自编码器主要用于表示学习以及降维，深度自编码器学习到的有效特征使用随机傅里叶特征进行核近似，再输入线性单类支持向量机，使提出的模型具有更高的准确率与鲁棒性。

提出的模型在于将深度网络提取丰富特征表示的能力与 OCSVM 目标函数相结合，获得超平面以将所有正常数据点与原点分开。

本文的其余部分安排如下：第 2 节回顾了 CAE 与 OCSVM 和核近似的背景；第 3 节详细阐述了本文提出的混合模型；第 4 节介绍了所用的数据集、实验设置等；第 5 节对试验结果进行了分析与讨论。最后对本文进行了总结。

1 相关技术

1.1 深度卷积自编码器

自动编码器模型由 Rumelhart 等人提出[13]，AE（auto-encoder）被认为是一种无监督的全连接单层神经网络，可以从未标记的数据集中学习特征表示。AE 的核心思想是重建数据的原始模式，从而获得数据的降维表示。

最原始的自编码器网络是一个三层的前馈神经网络结构，由输入层、隐藏层、输出层（重构层）构成，如图 1 所

示。整个自编码器由下面两个操作组成。

一是编码，数据从输入层到隐藏层的过程，其过程如式（1）所示：

$$h(x) = f(W^T x + b) \tag{1}$$

其中：f 是激活函数，$x = (x_1, x_2, \cdots, x_n) \in R^n$ 是输入向量，n 表示输入层神经元的个数，$h = (h_1, h_2 \cdots, h_k) \in R^k$ 是隐藏层的向量，表示输入向量的低维压缩表示，k 表示隐藏层神经元的数量。$W \in R^{n \times k}$ 是输入层与隐藏层连接的权重矩阵，$b \in R^{n \times 1}$ 是输入层的偏置向量。

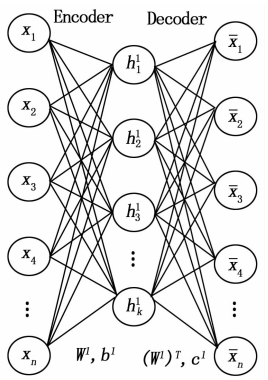

图 1　自编码器基本结构

二是解码：数据从隐藏层到输出层的过程，其计算过程如式（2）所示：

$$\bar{x} = f((W^*)^T h(x) + c) \tag{2}$$

$\bar{x} = (\bar{x_1}, \bar{x_2}, \cdots, \bar{x_n}) \in R^n$ 是重构层输出向量，输出层的神经元个数与输入层的神经元个数相同。其中，$W^* = W^T$，$c \in R^{k \times 1}$ 是隐含层的偏置向量。

自编码器是一种无监督的神经网络模型，其核心作用是能够学习到输入数据的深层表示。将输入层数据 x 通过转换得到其隐藏层的表示，然后由隐藏层重构，还原出新的输入数据 $\bar{x}$。AE 的训练目标就是使得重构后的数据 $\bar{x}$ 能够尽量还原输入层数据 x。AE 的损失函数通常采用均方误差来定义，如式子（3）所示：

$$L(x, \bar{x}, W) = \frac{1}{2n} \sum_{k=1}^{n} \| x_k - \bar{x_k} \|_2^2 \tag{3}$$

AE 和 DAE（deep auto-encoder）的主要局限性在于它们无法捕获图像和视频序列中的二维结构，这样的特性将导致网络参数的冗余，并且去除了可以从图像中提取的局部信息，而这在异常检测场景中特别重要，因为异常位于图像中。为了解决这个问题，Masci 等人提出了 CAE 体系结构[14]。CAE 与普通 AE 类似，卷积自编码器（CAE）将卷积神经网络（CNN，convolutional neural network）中的卷积滤波的优点与自动编码器的无监督预训练结合在一起。但它们之间的区别在于，在 CAE 中，权重在输入的所有位置之间共享，从而保留了二维空间的局部性，类似 CNN[15]。损失函数类似于 AE，如式（4）所示。

$$L(x, \bar{x}, W) = \frac{1}{2n} \sum_{k=1}^{n} \| x_k - \bar{x_k} \|_2^2 + \lambda \| W \|_2^2 \tag{4}$$

其中：λ 是正则项 $\| W \|_2^2$ 的系数，类似于 CNN，CAE 也包括卷积层，池化层及全连接层，反卷积层，反池化层。如图 2 所示。

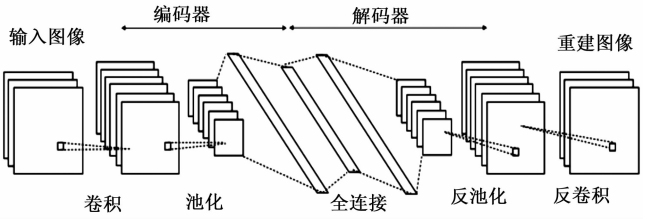

图 2　卷积自编码器体系结构

即编码器和解码器分别由全连接改成卷积与反卷积操作，通过外层的池化与反池化将整个网络贯穿起来。CAE 的编码过程包括卷积操作与池化操作，卷积操作表示为：

$$h^k = \sigma(x * W^k + b^k) \tag{5}$$

其中：h^k 表示第 k 个卷积面，σ 表示激活函数，W^k 和 b^k 分别表示第 k 个权值和偏置，$*$ 表示二维卷积操作。卷积操作完成后进行池化，池化的方法包括最大池化，平均池化与随机池化，本文选择最大池化。CAE 的解码过程包括反池化层，反卷积层；反卷积层通过反卷积执行卷积层的逆运算。反卷积层中的学习滤波器用作重构输入形状的基础，并考虑了所需的输出维度，如式（6）所示：

$$o = \sigma(\sum_{k \in H} h^k * \widetilde{W}^k + c) \tag{6}$$

其中：H 表示卷积面的指标集，c 表示输入通道的偏置，$\widetilde{W}^k$ 表示对权值 W^k 进行两个维数方向翻转后得到的权值，o 表示重构结果。卷积和反卷积层可以堆叠以构建 CAE 的深层架构。卷积层第一层（和反卷积层中的最后一层）中的滤波器提取低层特征，而后续层可以提取高级特征。

1.2　单分类支持向量机

Scholkopf 等人[16]提出了单分类 SVM，把 SVM 从二分类扩展到了单分类。给定一组训练向量 $x_i \in R^n, i = 1, \cdots, N$，其中所有的训练向量都属于同一个类。OCSVM 构建了一个超平面，该超平面基本上将所有目标类数据点与原点分开，并使该超平面与原点的距离最大化。这主要是通过解决式（7）所示的优化问题来完成的。

$$\begin{aligned} &\min_{\omega, \rho, \xi} \quad \frac{1}{2}\omega^T\omega - \rho + \frac{1}{\upsilon N}\sum_{i=1}^{N}\xi_i \\ &s.t. \quad \omega^T\varphi(x_i) \geqslant \rho - \xi_i \quad i = 0, \cdots N, \\ &\qquad \xi_i \geqslant 0, \quad i = 0, \cdots N \end{aligned} \tag{7}$$

决策函数表示为：

$$f(x) = sign(\sum_{i=1}^{N} a_i \boldsymbol{K}(x_i, x) - \rho) \tag{8}$$

如果 x_i 属于目标类，取值为 1，否则取值为 0。其中，ω 是权重系数，$\varphi(\cdot)$ 是特征空间中的映射，$\boldsymbol{K}$ 是核矩阵，a_i 是拉格朗日乘数，N 是训练样本总数，υ 是一个控制训练集中离群点比例的正则参数，ρ 是偏差项。决策函数中的 a_i 可以通过公式（9）所示的对偶问题来获得。

$$\min_{a} \frac{1}{2} a^T Q a$$

$$s.t. \quad 0 \leqslant a_i \leqslant \frac{1}{\upsilon N} \quad i = 0, \cdots, N,$$

$$\sum_{i}^{N} a_i = 1 \tag{9}$$

其中：$Q_{ij} = K(x_i, x_j) = \varphi(x_i)^T \varphi(x_j)$。

从上述描述可以看出，为了寻找最优超平面的问题，即 OCSVM 的决策边界，超参数选择的至关重要[17]。为了解决这个问题，本文使用铰链损失函数来取代（7）式中的 ξ_i。则式（7）转换为一个无约束的优化问题，如下式（10）所示：

$$\min_{\omega, \rho} \quad \frac{1}{2}\omega^T\omega - \rho + \frac{1}{\upsilon N}\sum_{i=1}^{N} \max(0, \rho - \omega^T h(x_i)) \tag{10}$$

其中：$h(x_i)$ 表示卷积自编码器的最后隐含层的输出。

1.3 随机傅里叶特征

为了解决核机器学习的可扩展性问题，核近似算法被广泛应用，由于随机傅里叶核近似方法具有较低的复杂度，且不需要预训练，因此本文主要采用 RFF。Bochner 定理[11]保证，只要核函数满足平移不变性，连续性和正定性，就存在一个概率分布 $p(\cdot)$ 的傅里叶变换与 $k(\cdot)$ 对应，即：

$$k(x, y) = \int_{R^d} p(w) e^{jw^T(x-y)} \mathrm{d}w \tag{11}$$

其中：$p(w)$ 是 w 的概率密度函数，对于高斯核函数，通过 $k(x, y)$ 的傅里叶逆变换计算 $p(w)$，可得 $w \sim N(0, 2\gamma I)$，其中 I 表示单位矩阵。利用标准蒙特卡洛近似积分方法逼近高斯核，D 维独立同分布的权重向量 $(w_1, w_2, \cdots w_D)$ 可以从分布 $p(\cdot)$ 中进行采样，核函数的偏移余弦映射如式（12）所示。

$$z(x) = \sqrt{\frac{1}{D}}[\cos(w_1^T x) \cdots \cos(w_D^T x) \quad \sin(w_1^T x) \cdots \sin(w_D^T x)]^T \tag{12}$$

其中：$w_i \in R^d$ 服从正态分布 $N(0, 2\gamma I)$。应用核近似映射，非线性单分类支持向量机的无约束目标函数变换为：

$$\min_{\omega, \rho} \quad \frac{1}{2}\omega^T\omega - \rho + \frac{1}{\upsilon N}\sum_{i=1}^{N} \max(0, \rho - \omega^T z(x_i)) \tag{13}$$

这样训练集就被转换为将 $\{(z(x_i), y_i)\}_{i=1}^{l}$，作为线性 OCSVM 的输入，从而可将线性 OCSVM 的高效求解算法应用到非线性 OCSVM 上，目标函数的优化求解将会变得更加简单，虽然 R^D 的维度高于 R^d 的维度。

2 提出的模型

这部分详细阐述基于 CAE 与 OCSVM 的组合模型，即深度卷积自动编码—单分类支持向量机（CAE-OCSVM）模型，用于高维和大数据集的异常检测的任务。该模型由两个主要组件组成，如图 3 所示。

深度卷积自编码器主要用于降维，获得输入图像的压缩表示。单分类支持向量机向量机主要用于异常检测，为了使 OCSVM 也适用大规模高维数据集，本文使用随机傅里叶特征进行核近似。模型中卷积自编码器的瓶颈层的特征经核近似后直接作为线性 OCSVM 的输入，同时进行训练以优化两个模型的变量参数，这将会使得卷积自编码学习到的特征更有利于 OCSVM 把正常数据与异常数据区分开来。

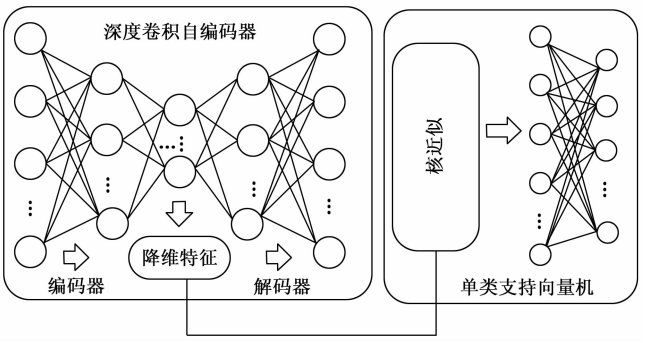

图 3 提出的 CAE-OCSVM 框架模型

假设 x 表示深度卷积自编码器的输入，$\bar{x}$ 表示重构输入，即自编码器瓶颈层的特征表示。θ 表示 CAE 参数集合，模型的目标函数表示如式（14）所示：

$$Q(\theta, w, \rho) = \alpha L(x, \bar{x}) + \frac{1}{2}\omega^T\omega - \rho + \frac{1}{\upsilon N}\sum_{i=1}^{N} \max(0, \rho - \omega^T z(x_i)) \tag{14}$$

其中：

$$L(x, \bar{x}) = \frac{1}{2n}\sum_{k=1}^{n} \| x_k - \bar{x_k} \|_2^2 + \lambda \| W \|_2^2 \tag{15}$$

式（15）表示卷积自编码器的重构误差，z 是如公式（12）所示的随机傅里叶映射，α 是平衡卷积自编码器重构损失与 OCSVM 间隔优化的超参数。从公式（14）的目标函数可以看出，在模型训练过程中，不仅考虑了单分类 SVM 的损失函数的最小化，而且也考虑了深度卷积神经网络的重构误差的最小化优化，在两者的作用下，将会学习到数据更加本质的特征表示，这将便于提高异常检测的准确性。目标函数使用梯度下降法进行优化，具体如算法如表 1 所示。

表 1 提出的算法流程

算法 1：提出模型的算法流程	
初始化	Step1：给出训练集 X，CAE 的网络结构，迭代次数 T；
	Step2：利用 Xavier 算法初始化 CAE 网络参数，同时使用网格搜索算法初始化 OCSVM 的参数
迭代训练	Step3：使用公式(3)、(4)计算 CAE 的隐含层输出，用公式(12)对 CAE 的隐含层输出进行核近似
	Step4：使用梯度下降法对公式(14)进行优化求解，更新模型参数
输出	Step5：模型收敛，输出模型参数

3 试验配置

本文的方法以无监督的方式从高维数据中检测异常。在 4 个公开数据集上进行了试验，并与主流无监督方法进行了比较。4 个公开基准数据集分别是 MNIST，Fashion

MNIST，CIFAR-10，STL-10，详细描述如表 2 所示。在某些应用中，异常情况具有通用性，例如入侵检测[18]，从搜索引擎检索图像[19]等。针对这些应用，从每个数据集中选择一个类别作为普通类别，而将其余类别视为数据集中的异常类别。本文使用 AUC（Area Under Curve）作为模型的性能评价指标。

表 2　实验数据集

MNIST	784(28 * 28)	70,000
Fashion MNIST	784(28 * 28)	70,000
CIFAR-10	3072(32 * 32 * 3)	60,000
STL-10	27,648(96 * 96 * 3)	13,000

3.1　数据集介绍

MNIST：它包含 70 000 张黑白手写数字图片，类别从 0～9，本文试验中选择某一类别作为异常类，样本是 28 * 28 的二维灰白图像。

Fashion MNIST：它是一些流行服饰的数据集，包含 10 个类别，跟 MNIST 类似，也是 70 000 张灰白图像，格式也是 28 * 28 的二维灰白图像。

CIFAR-10：它包含 10 个普适物体类别，每个类别包含 6 000 张彩色图像，图像的尺寸是 32 * 32 * 3。

STL-10：它与 CIFAR-10 类似，也是 10 个类别，每个类别包含 1 300 张彩色图像，800 张作为训练集，500 张作为测试集，图像的尺寸是 96 * 96 * 3。

3.2　基准方法

CAE-OCSVM 模型与文献［20］的模型单类支持向量机（OCSVM），局部离群因子（LOF，local outlier factor），隔离森林（IF，isolation forest），深度卷积自编码器（DCAE，deep convolutional autoencoder）和深度支持向量数据描述（Deep SVDD，deep support vector data description）、深度嵌入聚类（DEC，deep embedded cluster）以及 DAESVM 的异常检测方法进行了比较。

3.3　数据预处理与参数设置

3.3.1　预处理

根据式（16），利用 max-min 归一化将所有输入图像的数值映射到［0，1］范围，进行归一化处理。

$$x_i = \frac{x_i - \min}{\max - \min} \tag{16}$$

其中：x_i 为数据的属性值，min 为该数据属性的最小值，max 为最大值。

3.3.2　模型配置

本文采用的深度卷积自编码器的网络结构如图 4 所示，这个结构包含 3 个卷积层，3 个反卷积层与 1 个稠密层，在稠密层上使用 L2 正则化。在编码器部分的过滤器的尺寸分别为 5 * 5，5 * 5，3 * 3，卷积操作的步长为 2 * 2，使用 ReLu 激活函数。稠密层的神经元数量是 10，这样可以得到一个 10 维的压缩特征。本文采用 Adam 优化算法来进行神经网络的训练，所有数据集的学习率设置为 0.001，每个数据集的最小训练批次与训练次数如表 3 所示。

表 3　训练参数设置

数据集	训练批次	训练迭代次数
MNIST	256	200
Fashion MNIST	256	400
CIFAR-10	256	1 000
STL-10	64	300

4　结果与讨论

为了评估该方法的总体性能，在 4 个公开可用的数据集上进行了实验，并与异常检测的 5 个常用方法的 AUC 进行了比较。此外，分析了异常率的影响并将其与其他方法进行了比较。

4.1　4 个数据集上的 AUC 比较

在 4 个高维数据集上评估了本文提出的模型，这些数据集的描述如表 2 所示。表 4 至表 7 显示了所提模型在 MNIST，FASHION MNIST，CIFAR-10 和 STL-10 数据集上的实验结果。每个结果都是运行五次的平均值。从以下表格可以看出，CAE-OCSVM 模型优于其他经典异常检测方法，4 个数据集的平均 AUC 为 0.930、0.859、0.656、0.672。在表 4 的 MNIST 数据集中，尽管当正常类别分别为数字 5、6 和 7 时，LOF 的 AUC 略高于所提模型，但所提模型总体上显示出更好的性能。在 Fashion MNIST 数据集上，从表 5 中可以看出，当正常类别为 Dress 与 Shirt 时，LOF 模型的 AUC 分别为 0.883 和 0.787，要高于本文模型的 0.841 和 0.689，在其余类别以及平均 AUC 方面，本文模型的性能会更好。CIFAR-10 的结果列于表 6，当正常类别为 Airplane，Cat，Dog 和 Ship 时，CAE-OCSVM 表现出良好的性能。尽管对于其他正常类别，本文模型表现并不特别令人满意，但 CAE-OCSVM 的总体 AUC 达到了最佳，平均 AUC 为 0.656。表 7 显示了数据集维度为 27，648 的 STL-10 的结果，当正常类别为 Deer，Dog，Truck 时，CAE-OCSVM 的 AUC 性能略低于方法 IF，特别是在类别 Truck 中，模型的性能低于 IF、DCAE 和 OCSVM，但整体上可获得最佳性能。其平均 AUC 为 0.672，大于其它方法的 AUC。

表 4　模型在 MNIST 数据集异常率为 0.1 情况下的平均 AUCs

类别	LOF	IF	DCAE	OCSVM	DAESVM	CAE-OCSVM
0	0.868	0.880	0.631	0.890	0.922	0.942
1	0.972	0.993	0.977	0.978	0.973	0.985
2	0.841	0.697	0.596	0.731	0.840	0.875
3	0.890	0.772	0.546	0.794	0.908	0.932
4	0.887	0.867	0.700	0.879	0.885	0.903
5	0.922	0.737	0.557	0.772	0.887	0.912
6	0.976	0.883	0.746	0.859	0.910	0.942
7	0.951	0.903	0.820	0.908	0.909	0.928
8	0.853	0.721	0.489	0.806	0.936	0.942
9	0.956	0.875	0.757	0.886	0.930	0.961
平均值	0.912	0.833	0.682	0.847	0.910	0.930

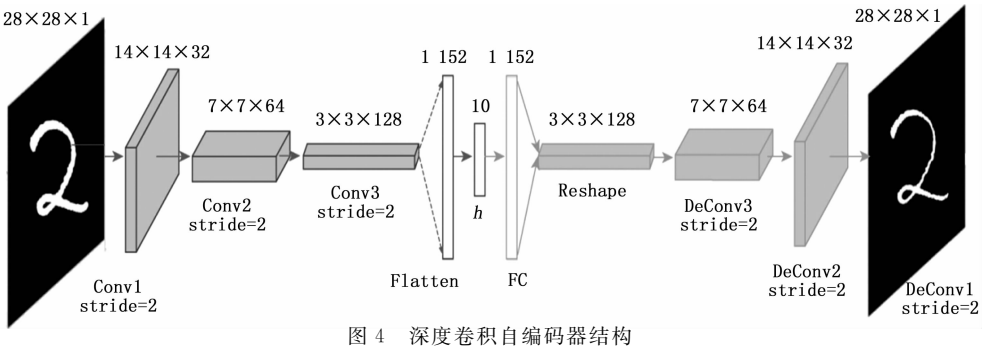

图 4 深度卷积自编码器结构

表 5 模型在 Fashion MNIST 数据集异常率 0.1 情况下的平均 AUCs

类别	LOF	IF	DCAE	OCSVM	DAESVM	CAE-OCSVM
T-shirt	0.788	0.791	0.633	0.834	0.803	0.862
Trouser	0.870	0.908	0.949	0.891	0.949	0.979
Pullover	0.860	0.706	0.643	0.808	0.840	0.872
Dress	0.883	0.847	0.691	0.826	0.815	0.841
Coat	0.870	0.774	0.579	0.820	0.848	0.911
Sandal	0.568	0.824	0.489	0.721	0.799	0.848
Shirt	0.787	0.627	0.587	0.784	0.788	0.689
Sneaker	0.704	0.911	0.699	0.924	0.923	0.935
Bag	0.741	0.645	0.417	0.679	0.842	0.754
Ankle boot	0.679	0.797	0.660	0.840	0.876	0.894
平均值	0.775	0.783	0.635	0.813	0.849	0.859

表 6 模型在 CIFAR-10 数据集异常率 0.2 情况下的平均 AUCs

类别	LOF	IF	DCAE	OCSVM	DAESVM	CAE-OCSVM
Airplane	0.633	0.633	0.612	0.596	0.721	0.731
Automobile	0.380	0.413	0.358	0.653	0.597	0.610
Bird	0.673	0.659	0.638	0.658	0.654	0.658
Cat	0.472	0.481	0.540	0.590	0.568	0.610
Deer	0.705	0.732	0.657	0.719	0.683	0.693
Dog	0.469	0.499	0.500	0.524	0.598	0.632
Frog	0.673	0.707	0.558	0.681	0.711	0.692
Horse	0.488	0.529	0.440	0.705	0.578	0.591
Ship	0.613	0.668	0.630	0.615	0.718	0.725
Truck	0.408	0.468	0.319	0.551	0.637	0.612
平均值	0.551	0.579	0.525	0.629	0.647	0.656

4.2 数据集中异常率的影响

本文为了验证所提出模型的鲁棒性，设置异常率为 0.1 到 0.5，试验结果如图 5 所示。从图 5 可以看出，本文模型的 AUC 值在不同的数据集以及不同的异常率中都是最大的。此外，从图 5 也可以看出，随着的异常率的变化，本文所提模型的 AUC 值的波动是最小的，而其他模型的 AUC 值却随着异常率的增加而下降较大，这充分验证了所提模型具有更强的鲁棒性与泛化能力。从图中结果可以看出，随着异常率的增大，各个模型的 AUC 性能基本上是呈下降趋势的，这也证明了模型对样本中存在少量异常点的异常检测能力。从图中也反映出了深度学习与 OCSVM 结合的模型在异常率较小的时候，效果显著，DAESVM 与 CAE-OCSVM 在异常率 0.1～0.3 之间时，两个模型的性能是最优的，这也体现了这种混合模型适合具有少量异常样本数据集的异常检测。

表 7 模型在 STL-10 数据集异常率 0.2 情况下的平均 AUCs

	LOF	IF	DCAE	OCSVM	DAESVM	CAE-OCSVM
	0.543	0.430	0.457	0.378	0.648	0.725
Bird	0.641	0.687	0.685	0.652	0.674	0.702
Car	0.578	0.583	类别	0.598	0.605	0.621
Cat	0.544	0.507	Airplane	0.423	0.704	0.782
Deer	0.620	0.624	0.610	0.628	0.619	0.621
Dog	0.658	0.747	0.698	0.766	0.708	0.719
Horse	0.733	0.562	0.537	0.570	0.604	0.615
Monkey	0.536	0.565	0.466	0.552	0.611	0.621
Ship	0.594	0.636	0.545	0.603	0.660	0.660
Truck	0.622	0.680	0.699	0.665	0.627	0.652
平均值	0.607	0.602	0.562	0.584	0.646	0.672

4.3 模型复杂度分析

根据提出的模型可知，模型主要包含深度卷积自编码器和单类支持向量机，模型的时间复杂度主要是两者时间复杂度之和，由文献 [21] 可知，卷积自编码器的时间复杂度为：

$$T_{CAE} \sim O\left(\left(\sum_{i=1}^{k} M_k^2 \cdot S_k^2 \cdot C_{k-1} \cdot C_k\right)\right)$$

其中：K 为卷积自编码器的卷积层数，M^2 为特征图面积，S^2 为卷积核面积，k 表示第 k 个卷积层，C_k 为第 k 个卷积层的卷积核个数。

而提出的模型的优势在于卷积自编码器器输出后进行核近似，然后通过线性支持向量机进行异常检测，故重点分析一下核近似支持向量机的时间复杂度的提升。假定训练集样本数为 n，d 为深度卷积自编码器的输出维度，也就是样本降维后的维度，D 为随机傅里叶特征的空间维度，由文献 [22] 可知，使用 SMO 算法训练高斯核 SVM 的时间复杂度为 $O(n^2)$，而求解线性支持向量机的时间复杂度为 $O(nD)$，随机傅里叶特征映射的时间复杂度为 $O(nDd)$；在单类支持向量机的参数设置中，本文使用 k 折交叉验证来决定高斯核的参数 γ 和正则化参数 ν，令 α，β 分别表示搜索网格的长度，则高斯核 SVM 的时间复杂度为 $t_0 = O(k\alpha\beta n^2)$ 线性 SVM 的时间复杂度为 $t_1 = O(k\alpha\beta nD)$，随机特征映射的时间复杂度

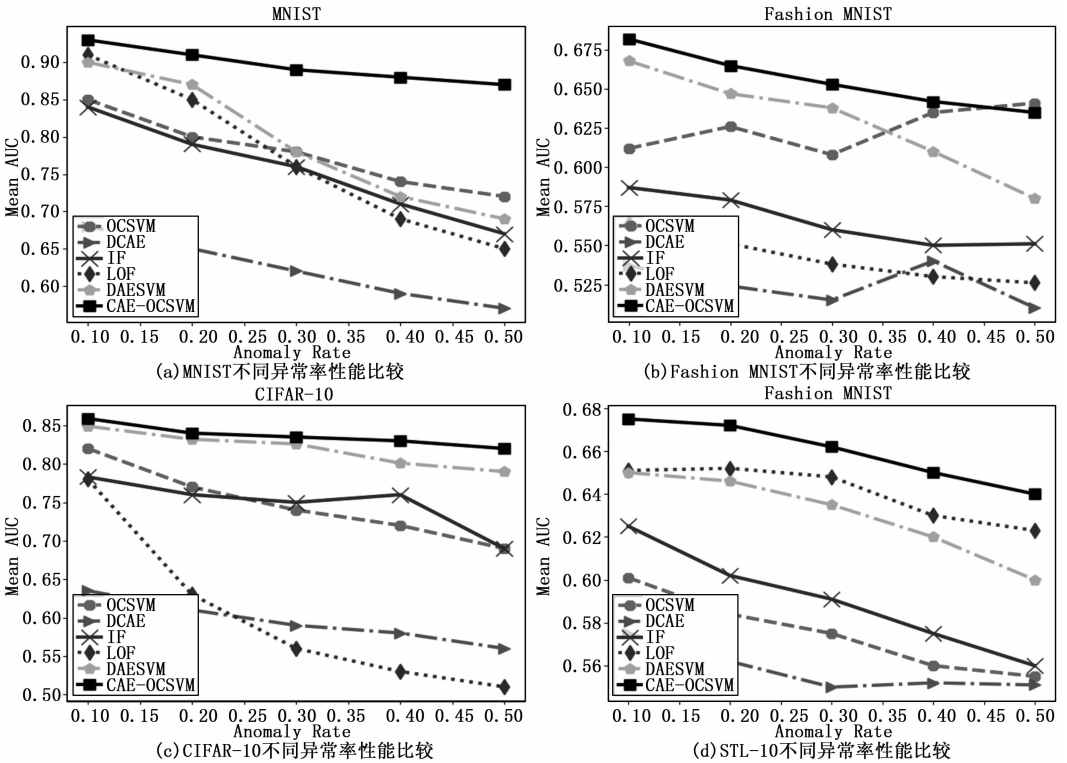

图 5　4 个数据集上不同异常率的性能分析

为变为 $t_2=O(\alpha nDd)$，那么从核近似到训练线性支持向量机的总的时间为随机特征映射的过程所花费的时间与训练单类支持向量机的时间之和，即 $T_{\text{OCSVM}}\sim t_1+t_2$，故所提模型的时间复杂度为 $T=T_{\text{CAE}}+T_{\text{OCSVM}}$。

由分析可知，高斯核 OCSVM 的时间复杂度与样本数量的平方成正比，而核近似之后，经线性 OCSVM 训练的时间复杂度与样本的数量成线性关系，远远小于核支持向量机的训练时间。这充分证明了所提模型的高效性。

5　结束语

本文提出了一个新的图像异常检测模型，模型联合深度卷积自编码器与单分类支持向量机，深度卷积自编码器负责图像降维以及特征表示学习，单分类支持向量机负责异常检测。同时使用随机傅里叶特征对单分类支持向量机的核函数进行核近似，把 3 种技术联合起来，构建了一个新的模型目标函数，目标函数能够学习到数据集的本质特征，实现了一个端到端的训练模型。模型在 4 个图像数据集上进行了实验验证，并且与不同的异常检测模型进行了性能对比，从实验结果来看，提出的模型在整体性能方面优于其它异常检测模型。模型还在不同的异常率的情况下进行了实验，结果证实，提出的模型具有更好的鲁棒性与泛化能力。下一步将重点在多核近似方面进行进一步的研究，来进一步提高模型的泛化能力，同时将在更多的数据集上进行实验验证。

参考文献：

[1] PIMENTEL M A F, CLIFTON D A, CLIFTON L, et al. A review of novelty detection [J]. Signal Processing, 2014, 99: 215-249.

[2] YUAN Y, FANG J, WANG Q. Online anomaly detection in crowd scenes via structure analysis [J]. IEEE transactions on cybernetics, 2014, 45 (3): 548-561.

[3] CHEN C, SHAO Y, BI X. Detection of anomalous crowd behavior based on the acceleration feature [J]. IEEE sensors journal, 2015, 15 (12): 7252-7261.

[4] XU D, RICCI E, YAN Y, et al. Learning deep representations of appearance and motion for anomalous event detection [J]. arXiv preprint arXiv: 1510.01553, 2015.

[5] KIM J, GRAUMAN K. Observe locally, infer globally: a space-time MRF for detecting abnormal activities with incremental updates [C] //2009 IEEE conference on computer vision and pattern recognition. IEEE, 2009: 2921-2928.

[6] XIAO T, ZHANG C, ZHA H B. Learning to detect anomalies in surveillance video [J]. IEEE Signal Processing Letters, 2015, 22 (9): 1477-1481.

[7] ERFANI S M, RAJASEGARAR S, KARUNASEKERA S, et al. High-dimensional and large-scale anomaly detection using a linear one-class SVM with deep learning [J]. Pattern Recognition, 2016, 58: 121-134.

[8] 李　权，周兴社. 基于 KPCA 的多变量时间序列数据异常检测方法研究 [J]. 计算机测量与控制，2011，19 (4)：822-825.

[9] HASAN M, CHOI J, NEUMANN J, et al. Learning temporal regularity in video sequences [C] //Proceedings of the IEEE conference on computer vision and pattern recognition, 2016: 733-742.

(下转第 276 页)