Computer Science and Application 计算机科学与应用, 2021, 11(3), 682-689
Published Online March 2021 in Hans. http://www.hanspub.org/journal/csa
https://doi.org/10.12677/csa.2021.113070

Hans汉斯

基于深度自编码器核概率密度的异常检测模型

吕　鹏

烟台大学计算机与控制工程学院，山东　烟台
Email: 1246170471@qq.com

收稿日期：2021年2月25日；录用日期：2021年3月19日；发布日期：2021年3月30日

摘　要

无监督技术通常依靠数据的概率密度分布来检测异常数据，在该类异常监测模型中，具有低概率密度的对象被认为是异常对象。然而，对高维数据的密度分布建模是困难的，这使得从高维数据中检测异常数据的问题变得极具挑战性。最先进的方法被称为‘两步走’框架，该框架首先对数据应用降维技术进行降维，然后在低维空间进行异常检测来解决此问题。不幸的是，低维空间不一定保留原始高维数据的密度分布，这损害了异常检测的有效性。在这项工作中，本文提出了一种新颖的高维数据异常检测方法，称为AEDE (AutoEncoding kernel Density Estimation model)。核心思想是结合核密度估计(KDE)的密度估计能力和深度自编码器的表示学习能力，以便可以学习能够有效分离异常数据的概率密度分布。通过在自编码器的训练过程中使用概率密度策略，AEDE成功地整合了两部分的优势，即深度自编码器和概率密度模型。本文使用四个公开数据集进行的实验表明，在检测异常方面，AEDE模型明显优于最新方法，F_1得分提高了30%。

关键词

异常检测，深度自编码器，核密度估计，深度学习

Deep Autoencoding Kernel Density Estimation Model for Anomaly Detection

Peng Lv

School of Computer and Control Engineering, Yantai University, Yantai Shandong
Email: 1246170471@qq.com

Received: Feb. 25th, 2021; accepted: Mar. 19th, 2021; published: Mar. 30th, 2021

Abstract

Unsupervised techniques typically rely on the probability density distribution of the data to detect

文章引用: 吕鹏. 基于深度自编码器核概率密度的异常检测模型[J]. 计算机科学与应用, 2021, 11(3): 682-689.
DOI: 10.12677/csa.2021.113070

anomalies, where objects with low probability density are considered to be abnormal. However, modeling the density distribution of high dimensional data is known to be hard, making the problem of detecting anomalies from high-dimensional data challenging. The state-of-the-art methods solve this problem by first applying dimension reduction techniques to the data and then detecting anomalies in the low dimensional space. Unfortunately, the low dimensional space does not necessarily preserve the density distribution of the original high dimensional data. This jeopardizes the effectiveness of anomaly detection. In this work, we propose a novel high dimensional anomaly detection method called AEDE. The key idea is to unify the representation learning capacity of deep autoencoder with the density estimation power of kernel density estimation (Auto Encoding kernel Density Estimation model, KDE) such that a probability density distribution of the high dimensional data can be learned that is able to effectively separate the anomalies out. AEDE successfully consolidates the merits of the two worlds, namely variational autoencoder and KDE by using a probability density-aware strategy in the training process of the autoencoder. Our extensive experiments using four benchmark datasets demonstrate that our method significantly outperforms the state-of-the-art methods in detecting anomalies, achieves up to 30% improvement in F_1 score.

Keywords

Anomaly Detection, Deep Autoencoder, Kernel Density Estimation, Deep Learning

1. 引言

无论是在网络安全领域[1]，机器制造业领域[2]，信息系统管理领域[3]和医学领域[4]，异常检测都是一个值得研究的基本问题。无论是高维数据还是多维数据，异常检测的核心都是密度估计。通常情况下，正常数据的数量较大，并且符合一定的分布，异常数据的数量较少且离散分布，所以普遍认为异常数据处于低密度区域。

在过去的几十年中，异常检测已经取得了极大的进展和突破[5]，但是随着大数据时代的到来，数据的维度不断增加，由于维度诅咒的存在，对高维数据进行异常检测仍然是一个挑战。随着数据维度的增加，传统的方法越来越难以在数据的原始空间进行密度估计，并且随着维度的增加，噪声和无关因素也越来越多，这对检测结果的负面影响也越来越大。

目前解决该类问题最好的是一个叫做两步走[6]的框架，该框架先对数据进行降维，将高维数据降低为低维数据，然后在低维空间进行密度估计[7] [8]，但是，仅当正常实例和异常实例在数据的较低维空间中是可分离的时，它们才有用，因此要确定正确包含用于区分异常与正常数据的特征的低维表示仍然很困难。近来，深度学习在异常检测领域取得了巨大的成功[9]，流模型[10]，循环神经网络[11]，自编码器[10]及其一系列变体广泛的应用于高维数据异常检测领域，例如对抗性自编码器[12]，变分自编码器[13]等等，并取得了一系列的成果。这些方法的核心思想是将输入数据编码为低维表示，然后通过最小化重构误差将低维表示解码回原始数据空间。这些模型旨在通过训练深度神经网络来获得潜在数据空间中原始数据的核心特征，而不会产生噪音和无关特征。最近的几项研究已将此结构应用于实际问题，例如ALAD [14]，AnoGAN [15]等，但该类模型仍存在很大的探索空间。AnoGan 使用对抗性自编码器来检测图像数据中的异常，但是该模型仅利用重构误差进行异常检测，而没有充分利用低维表示。ALAD 同时

考虑了基于双向 GAN 的数据分布和重构误差，从而得出了异常检测任务的对抗性学习特征，尽管如此，ALAD 仍然仅使用基于对抗性学习特征的重建误差来确定数据样本是否异常。但是，现实世界的数据不仅可能具有高维，而且缺乏清晰的预定义分布(例如 GMM)。DAGMM [16]在异常检测中结合了深度自编码器和高斯混合模型(GMM)，在对输入数据的密度分布进行建模时，GMM 中还需要手动调整参数，这对检测性能会产生严重影响。

在本文中，提出了一种深度自编码核密度估计模型(AEDE)，这是一种深度学习框架，可解决高维数据集异常检测中遇到的上述挑战。一方面，AEDE 使用深度自编码器来获得数据的低维表示。由于变分自动编码器同时考虑了重构误差和潜在数据空间中数据的分布，因此在低维空间中保留了高维数据的密度分布。具体来说，AEDE 仅利用正常数据来训练深度自编码器，因此潜在空间中的数据分布仅针对正常数据，可以将其与异常对象区分开。另一方面，AEDE 使用核密度估计模型[17]训练数据的概率密度分布。与 DAGMM 需要手动指定混合高斯模型的数量不同，AEDE 可以对任意分布数据进行建模。当自编码器将输入数据编码为低维表示时，同时又将输入数据的关键特征保留在潜在数据空间中时，具有高密度值的数据更有可能是正常对象，而低密度值的数据则被认为是异常对象。

2. 模型设计

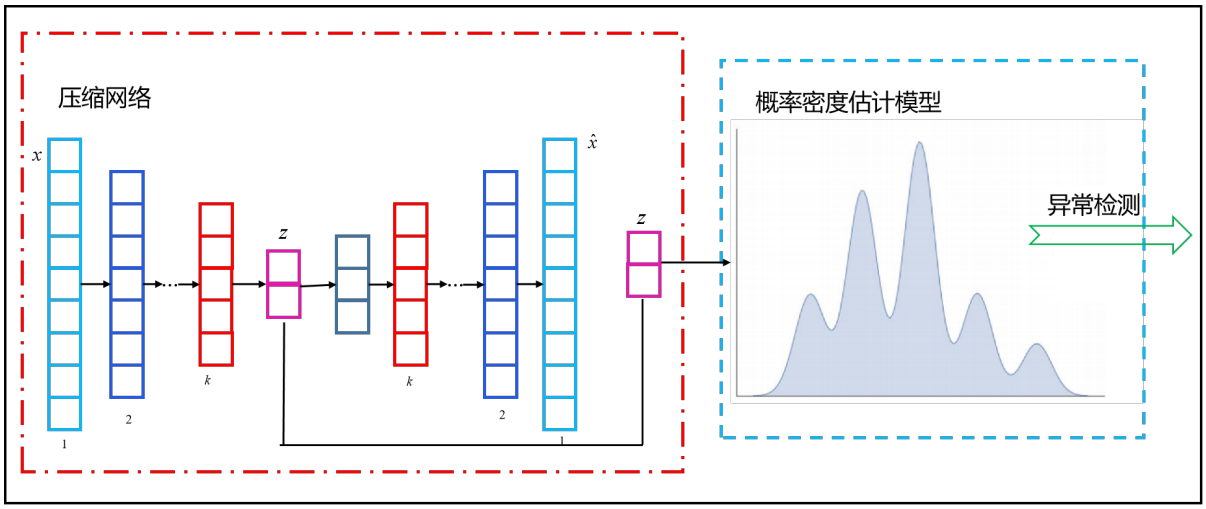

Figure 1. Overview
图 1. 模型框架

本文所提出的深度自编码器核密度估计模型的体系结构如图 1 所示。AEDE 主要由两部分组成：压缩网络和概率密度估计模型。第一部分是压缩网络，在压缩网络中，AEDE 通过深度自编码器对输入数据进行压缩，以获得其在潜在数据空间中的低维表示和重建误差，并将这些低维表示和重建误差一起提供给概率密度估计模型；第二部分是概率密度估计模型，概率估计模型采用高斯核密度估计来获取数据的低维表示，并学习数据概率密度分布。

2.1. 深度自编码器

原始数据在潜在数据空间中的低维表示是通过深度自编码器得出的，自编码器包括编码器和解码器两部分。编码器是具有权重和偏差 θ 的神经网络，它的输入是高维数据x，其输出是低维表示 z。本文将编码器表示为 $q_\theta(z|x)$。解码器是另一个具有权重和偏差 ϕ 为神经网络，将解码器表示为 $p_\phi(x|z)$。编码器通

过非线性转换，将高维数据转化为低维数据，而解码器通过这个逆过程，将低维数据转化为高维数据。

对于给定的输入数据x，深度自编码器按以下方式计算其低维表示z：

$$z = q(x,\theta), \tag{1}$$

$$\hat{x} = p(z,\phi), \tag{2}$$

这里的$\hat{x}$是原始数据的重建数据。

深度自编码器的损失函数如下：

$$l(\theta,\phi) = \|x - \hat{x}\|^2 \tag{3}$$

该损失函数被称为重建误差，用来衡量原始数据和重建数据的相似程度，相似程度越高，重建误差越小。

2.2. 概率估计模型

在概率密度估计模型中，本文使用深度自编码器学习到的低维表示来对输入数据的概率密度分布进行建模。对于概率估计模型的选择，本文选择核密度估计模型。

对于n个输入数据$\hat{x}_1,\hat{x}_2,\cdots,\hat{x}_n$，通过深度自编码器模型来获得输入数据的低维表示$\hat{z}^1,\hat{z}^2,\cdots,\hat{z}^n$，对于自编码器的输入数据，本文通过核密度估计模型计算得到的概率密度估计函数如下：

$$f_h(S) = \frac{1}{n}\sum_{i=1}^{n} K_h(s - z_i) = \frac{1}{nh}\sum_{i-1}^{n} K\left(\frac{s - z_i}{h}\right) \tag{4}$$

其中 s 是变量，而 K 是核函数，而 h ($h > 0$)是一个称为带宽的平滑参数。在核密度估计模型中本文采用高斯核函数。

2.3. 训练策略

本节主要介绍压缩网络和概率密度估计模型的训练过程，见表 1，训练样本由 $x_i\ (i = 0,1,2,\cdots,n)$ 表示。

Table 1. ADAE training process
表 1. ADAE 模型训练

输入：训练数据 $x_i(x_1,x_2,\cdots,x_3)$。
输出：ADAE 模型。
1. θ，φ←编码器解码器参数。
2. **for** ***i*** from 1 to *epochs* **do** ▷压缩网络：
3. $z_i = q(x_i,\theta)$；
4. $\hat{x}_i = p(z_i,\phi)$。
5. *loss function* $= \|x_i - \hat{x}_i\|^2$。
6. θ，ϕ←采用 *SGD* 更新参数。
7. 固定参数 θ，ϕ
8. for i from 1 to n do ▷训练概率密度模型：
9. $z_i = q(x_i,\theta)$。
10. $f_h(S) = \frac{1}{n}\sum_{i=1}^{n} K_h(s - z_i) = \frac{1}{nh}\sum_{i-1}^{n} K\left(\frac{s - z_i}{h}\right)$。

θ 和 ϕ 是编码器和解码器的参数，*epochs* 是训练轮数，z 是低维表示，$\hat{x}_i$ 是通过解码器获得数据的重构，使用随机梯度下降(*SGD*)更新参数 θ 和 ϕ，最后，获得概率密度分布函数 $f_h(S)$，以模拟训练数据在潜在空间中的分布。第一行初始网络参数，2~5 行训练神经网络，得到低维表示，第 6 行采用随机梯度下降进行优化参数，然后固定参数，8~10 行训练数据的概率密度分布。

2.4. 测试策略

本节主要介绍模型测试过程，见表 2，测试样本由 $y_j\left(j=1,2,3,\cdots,m\right)$ 表示。

Table 2. ADAE testing process
表 2. ADAE 训练过程

输入：测试样本 $y_j\left(j=1,2,3,\cdots,m\right)$, 异常比例 α。
输出：异常数据。

1. **for** j from 1 to m **do:**
2. $z_j=q\left(y_j,\theta\right)$。
3. **for** j from 1 to m **do:**
4. $f\left(z_j\right)=f_h\left(z_j\right)$。
5. $thr=sort_{thr}\left(f\left(c\right),\alpha\right)$
6. **for** j **from 1** ***to*** m **do:**
7. **if** $f\left(z_j\right)<thr$ **then:**
8. y_j **是异常。**
9. **else:**
10. y_j **不是异常。**

在测试过程中，通过训练好的密度估计模型计算每个测试数据的概率密度，通过异常比例确定阈值，高于阈值的判定为正常，低于阈值的判定为异常。1~2 行通过自编码器获得数据的低维表示，3~4 行是通过训练样本得到概率密度分布，计算每个输入数据的密度值看，第 5 行是对密度值进行排序，得到阈值，6-10 行是根据阈值判断数据是否异常。

3. 实验结果

在本节中，通过在四个公共数据集上与五个目前最先进的算法进行比较，评估本文提出的模型在异常检测中的有效性和鲁棒性。

3.1. 数据集

这四个著名的公共数据集是：KDDCUP，Thyroid，Arrhythmia 和 KDDCUP-Rev。六个对比算法分别是：OC-SVM [18]，DSEBM，DSEBM，AnoGAN 和 ALAD。实验采用的四个数据集的详细信息如表 3 所示。文中异常数据的比例如表 3 所示。

Table 3. Data sets description
表 3. 数据集描述

数据集	数据对象数量	数据维度	异常比例(α)
KDDCUP	494,021	118	0.2
Thyroid	3772	36	0.025
Arrhythmia	432	274	0.15
KDDCUP-Rev	121,597	118	0.2

3.2. 评估策略

本文使用平均精度，召回率和 F_1 分数来量化结果。精度和召回率定义如下：$\textbf{Precision}=\frac{|G|\bigcap|R|}{|R|}$ 和

$\mathbf{Recall} = \frac{|\boldsymbol{G}| \bigcap |\boldsymbol{R}|}{|\boldsymbol{G}|}$，其中 $\boldsymbol{G}$ 表示数据集中的真正的异常集，$\boldsymbol{R}$ 表示方法报告的异常集。$\boldsymbol{F_1}$ 分数定义如下：$\boldsymbol{F}_1 = \frac{2 * \mathbf{Precision} * \mathbf{Recall}}{\mathbf{Precision} + \mathbf{Recall}}$。基于异常率 $\boldsymbol{\alpha}$，可以确定阈值以识别异常对象。平均准确率是预测对的样本数占样本总数的比例，召回率是样本中的正例有多少被预测正确了，而 F_1 分数是通过准确率和召回率计算得到的。

3.3. 有效性评估

该章节是 AEAD 模型在四个公开数据集上与最新的算法进行性能评估实验，表 4 显示了 AEAD 和所有对比算法运行 20 次后，取得的平均精度，召回率和 $\boldsymbol{F_1}$ 得分。

Table 4. Effectiveness evaluation
表 4. 有效性评估

Method	KDDCUP			Thyroid		
	Precision	Recall	F_1	Precision	Recall	F_1
OC-SVM	0.7457	0.8523	0.7954	0.3639	0.4239	0.3887
DSEBM-r	0.8744	0.8414	0.8575	0.0400	0.0403	0.0403
DSEBM-e	0.2151	0.2180	0.2170	0.1319	0.1319	0.1319
DAGMM	0.9297	0.9442	0.9369	0.4766	0.4834	0.4782
AnoGAN	0.8786	0.8297	0.8865	0.0412	0.0430	0.0421
ALAD	0.9427	0.9577	0.9501	0.3196	0.3333	0.3263
AEDE	**0.9840**	**0.9655**	**0.9710**	**0.7934**	**0.7849**	**0.7891**
Method	**Arrhythmia**			**KDDCUP-Rev**		
	Precision	Recall	F_1	Recall	Precision	F_1
OC-SVM	0.5397	0.4082	0.4581	0.7148	0.9940	0.8316
DSEBM-r	0.4286	0.5000	0.4615	0.2036	0.2036	0.2036
DSEBM-e	0.4643	0.4645	0.4643	0.2212	0.2213	0.2213
DAGMM	0.4909	0.5078	0.4983	0.9370	0.9390	0.9380
AnoGAN	0.4118	0.4375	0.4242	0.8422	0.8305	0.8363
ALAD	0.5000	0.5313	0.5152	0.9547	0.9678	0.9612
AEDE	**0.8461**	**0.8333**	**0.8396**	**0.9890**	**0.9889**	**0.9890**

该章节是 AEAD 模型在四个公开数据集上与最近的算法进行性能评估实验，表 4 显示了 AEAD 和所有对比算法运行 20 次后，取得的平均精度，召回率和$\boldsymbol{F_1}$得分。本文遵循 ALAD(用于训练的整个官方数据集的 80%)和 DAGMM(占整个正常数据集的 50%)中的设置。在 KDDCUP 和 KDDCUP-Rev 实验中，AEAD 通过随机抽样获取 50%的数据用于训练，其余 50%保留用于测试，并且仅将“正常”数据中的数据样本用于训练模型。在 Thyroid 和 Arrhythmia 的实验中，随机抽样获取 80％的数据用于训练，其余 20%保留用于测试，并且仅将正常类别的数据样本用于训练模型。

从表 4 可以看到，在四个数据集上，AEAD 的平均精度，召回率和$\boldsymbol{F_1}$得分均明显优于所有对比方法。在 KDDCUP 数据集上与最新的 ALAD 相比，AEAD 的标准$\boldsymbol{F_1}$得分提高了 2.09%，在所有平均精度，召

回率和F_1方面均超过了所有对比算法。此外，AEAD 的效果明显优于最新的 DSEBM，DAGMM 和 ALAD 方法，Thyroid 和 Arrhythmia 的标准F_1得分分别提高了 31.09%和 32.44%。对于高维数据中的异常检测，OC-SVM 无法获得良好的结果。这是因为 OC-SVM 的核心思想是通过使用正常数据在高维空间中找到决策边界，但是当数据的属性过多时，不相关的冗余属性可能会对 OC-SVM 的结果产生很大的负面影响。尽管 DSEBM 考虑了重构误差和能量误差，但它忽略了潜在表示，这可能是本文提出的模型和 DAGMM 性能优于 DSEBM 的主要原因。AEAD 优于 DAGMM 的原因可能归因于 AEAD 采用核密度估计来建模数据的概率密度分布，而不是高斯混合模型。核密度估计模型优于高斯混合模型，因为高斯混合模型是参数估计，而 KDE 是非参数估计，它允许拟合数据的函数形式为在没有任何理论指导或约束的情况下获得数据的分布。此外，高斯混合模型还需要手动选择混合高斯模型的数量，这在缺乏领域知识的情况下非常棘手。对于 AnoGAN，它采用对抗自编码器来恢复每个输入数据的潜在表示，并使用重构误差和判别分量作为异常准则，但 AnoGAN 并未充分利用低维表示。尽管 ALAD 可以模拟分布，当实验数据足够多时，它的数据很好，但它也忽略了潜在表示的考虑。AEAD 优于所有对比算法的另一个潜在原因是，AEAD 采用了一种新颖的概率密度感知策略，该策略仅相对于正常训练样本的学习概率密度分布来估计每个输入数据的密度值，这种策略有助于有效地分离出潜在数据空间中密集分布的异常。

4. 总结

在未来的工作中，我们计划探索基于深度自编码器的有效半监督异常检测方法，我们还研究了针对大型高维数据的更有效的无监督异常检测。本文所提方法适用于全局异常检测，当知道异常的比率时，可以取得良好的结果。我们的下一步是探索如何改善模型从而适应高维数据的局部异常检测，使其不仅能够在海量数据的情况下，可以很好地模拟数据的分布，并且在小规模数据下，也能取得很好的效果。

基金项目

烟台大学研究生科技创新基金(YDZD2021)资助。

参考文献

[1] Tan, S.C., Ting, K.M. and Liu, T.F. (2011) Fast Anomaly Detection for Streaming Data. *Proceedings of the* 22*nd International Joint Conference on Artificial Intelligence*, Barcelona, 16-22 July 2011, 1511-1516.

[2] Liu, F.T., Ting, K.M. and Zhou, Z.-H. (2008) Isolation Forest. 2008 8*th IEEE International Conference on Data Mining*, Pisa, 15-19 December 2008, 413-422. https://doi.org/10.1109/ICDM.2008.17

[3] Keller, F., Muller, E. and Bohm, K. (2012) HiCS: High Contrast Subspaces for Density-Based Outlier Ranking. 2012 *IEEE* 28*th International Conference on Data Engineering*, Arlington, 1-5 April 2012, 1037-1048. https://doi.org/10.1109/ICDE.2012.88

[4] 陈科谚, 余蕙君, 张瑚, 等. 唇腭裂在胎儿期发育异常的染色体核型和微阵列分析[J]. 广东医学, 2019, 40(20): 2880-2885.

[5] 卓琳, 赵厚宇, 詹思延. 异常检测方法及其应用综述[J]. 计算机应用研究, 2020(S1): 9-15.

[6] Chandola, V., Banerjee, A. and Kumar, V. (2009) Anomaly Detection: A Survey. *ACM Computing Surveys* (*CSUR*), **41**, 1-58. https://doi.org/10.1145/1541880.1541882

[7] Idé, T. and Kashima, H. (2004) Eigenspace-Based Anomaly Detection in Computer Systems. *Proceedings of the* 10*th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, August 2004, 440-449. https://doi.org/10.1145/1014052.1014102

[8] Yu, W., Aggarwal, C.C., Ma, S. and Wang, H. (2013) On Anomalous Hotspot Discovery in Graph Streams. 2013 *IEEE* 13*th International Conference on Data Mining*, Dallas, 7-10 December 2013, 1271-1276. https://doi.org/10.1109/ICDM.2013.32

[9] Chalapathy, R. and Chawla, S. (2019) Deep Learning for Anomaly Detection: A Survey. *Proceedings of the* 26*th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, August 2020, 3507-3508. https://doi.org/10.1145/3394486.3406704

[10] Kingma, D.P. and Dhariwal, P. (2018) Glow: Generative Flow with Invertible 1x1 Convolutions. *Proceedings of the Advances in Neural Information Processing Systems*, NeurIPS, 10215-10224.

[11] 李锋, 王泽南. 基于 RNN 的心电信号异常检测研究[J]. 智慧健康, 2018, 4(31): 10-13.

[12] Ravanbakhsh, M., Nabi, M., Mousavi, H., *et al.* (2018) Plug-and-Play CNN for Crowd Motion Analysis: An Application in Abnormal Event Detection. *IEEE Winter Conference on Applications of Computer Vision* (*WACV*), Lake Tahoe, 12-15 March 2018, 1689-1698. https://doi.org/10.1109/WACV.2018.00188

[13] An, J. and Cho, S. (2015) Variational Autoencoder Based Anomaly Detection Using Reconstruction Probability. *Special Lecture on IE*, **2**, 1-18.

[14] Zenati, H., Romain, M., Foo, C.-S., *et al.* (2018) Adversarially Learned Anomaly Detection. *IEEE International Conference on Data Mining* (*ICDM*), Singapore, 17-20 November 2018, 727-736. https://doi.org/10.1109/ICDM.2018.00088

[15] Schlegl, T., Seeböck, P., Waldstein, S.M., *et al.* (2019) f-AnoGAN: Fast Unsupervised Anomaly Detection with Generative Adversarial Networks. *Medical Image Analysis*, **54**, 30-44. https://doi.org/10.1016/j.media.2019.01.010

[16] Zong, B., Song, Q., Min, M.R., *et al.* (2018) Deep Autoencoding Gaussian Mixture Model for Unsupervised Anomaly Detection. *Proceedings of the International Conference on Learning Representations*, Vancouver.

[17] Günter, S., Schraudolph, N.N. and Vishwanathan, S.V.N. (2007) Fast Iterative Kernel Principal Component Analysis. *The Journal of Machine Learning Research*, **8**, 1893-1918.

[18] Chen, Y., Zhou, X.S. and Huang, T.S. (2001) One-Class SVM for Learning in Image Retrieval. *Proceedings* 2001 *International Conference on Image Processing*, Thessaloniki, 7-10 October 2001, 34-37.