

本文引用格式：胡洋,肖明,康嘉文.改进掩码自编码器的多类工业图像异常检测方法[J].自动化与信息工程,2024,45(4):30-35.

HU Yang, XIAO Ming, KANG Jiawen. Improved mask autoencoder for multi-class industrial image anomaly detection method[J]. Automation & Information Engineering, 2024,45(4):30-35.

改进掩码自编码器的多类工业图像异常检测方法*

胡洋 肖明 康嘉文

(广东工业大学, 广东 广州 510006)

摘要: 针对不同异常图像数据间的差异以及深度模型泛化能力不足的问题, 提出一种改进掩码自编码器的多类工业图像异常检测方法。首先, 利用正常图像样本数据训练改进的掩码自编码器(MAE)模型, 使该模型具有正常图像重构的能力; 然后, 根据改进的 MAE 模型的重构图像与原图像的差异, 区分正常与异常图像数据; 最后, 利用该方法在公开的工业图像数据集上同时检测多个类别的异常图像数据, 平均 AUC 为 0.895, 相较于 MKD、U-Net、DAGAN, 其检测精度分别提高了 2.05%、9.28% 和 2.52%, 验证了该方法的有效性。

关键词: 编解码重构; 掩码自编码器; 异常检测; 工业图像

中图分类号: TP391.41

文献标志码: A

文章编号: 1674-2605(2024)04-0005-06

DOI: 10.3969/j.issn.1674-2605.2024.04.005

开放获取

Improved Mask Autoencoder for Multi-class Industrial Image Anomaly Detection Method

HU Yang XIAO Ming KANG Jiawen

(Guangdong University of Technology, Guangzhou 510006, China)

Abstract: Aiming at the differences between different abnormal image data and the insufficient generalization ability of deep models, an improved mask autoencoder based multi class industrial image anomaly detection method is proposed. Firstly, an improved Mask Autoencoder (MAE) model is trained using normal image sample data to enable the model to reconstruct normal images; Then, based on the difference between the reconstructed image of the improved MAE model and the original image, distinguish between normal and abnormal image data; Finally, this method was used to simultaneously detect multiple categories of abnormal image data on a publicly available industrial image dataset, with an average AUC of 0.895. Compared with MKD, U-Net, and DAGAN, the detection accuracy was improved by 2.05%, 9.28%, and 2.52%, respectively, verifying the effectiveness of this method.

Keywords: encoder-decoder reconstruction; masked autoencoder; anomaly detection; industrial image

0 引言

图像异常检测作为智能制造领域的重要技术^[1], 广泛应用于工业制造领域, 旨在识别产品的缺陷或异常。通过计算机视觉和深度学习技术, 图像异常检测能够自动分析产品的外观, 检测出细微的缺陷, 如裂纹、划痕、变形、污染等。

近年来, 随着深度学习技术的发展, 研究人员针对图像异常检测提出了许多基于深度学习的模型。RUFF 等^[2]利用多层感知机将正常图像样本映射到一

个超球体内, 异常图像样本映射到超球体外, 在公开的工业数据集 MVTecAD 上的曲线下面积 (area under curve, AUC) 为 0.857。VENKATARAMANAN 等^[3]以重构输入图像为目标训练变分自编码器来学习正常图像的特征分布, 并通过分析重构前后图像的差异进行图像异常检测, 在 MVTecAD 数据集上的 AUC 为 0.860。REISS 等^[4]利用预先训练的 Resnet 模型提取待测图像特征, 采用 K 近邻算法在正常图像和其他空间之间找到一个流形层, 通过该流形层来区分正常或异

常图像, 在 MVTecAD 数据集上的 AUC 为 0.865。LEE 等^[5]采用预先训练的 U-Net 模型重建正常图像块, 并作为参考图像块, 根据待测图像的重建图像块与参考图像块的差异来判断图像是否异常, 在 MVTecAD 数据集上 AUC 为 0.926。CHEN 等^[6]通过高斯插值法来训练编码器, 利用图像重构损失来训练解码器, 在 MVTecAD 数据集上 AUC 为 0.912。AKCAY 等^[7]利用生成对抗网络来检测图像异常, 在 MVTecAD 数据集上 AUC 为 0.782。LEE 等^[8]利用视觉变形器来检测图像异常, 在 MVTecAD 数据集上 AUC 为 0.780。DING^[9]提出一种学习多种异常的解耦表示方法, 引入基于潜在残差的异常学习模块, 通过图像正常、异常样本的特征差异来检测是否异常, 在 MVTecAD 数据集上 AUC 为 0.883。上述方法大多针对某一类图像数据进行异常检测, 当面对多类图像数据异常检测时, 这些模型面临过拟合风险。

掩码自编码器 (masked autoencoders, MAE) 是一种无监督的预训练模型^[10], 其通过一定比例的随机掩码进行编解码处理, 使编解码器能自动恢复被掩盖的部分。但 MAE 无法直接应用于图像异常检测任务, 因为若随机掩码没有掩盖缺陷, 就无法获取异常部分的重构损失, 可能导致图像异常检测失效。为此, 本文对 MAE 进行改进, 利用其特征捕捉能力来识别多类图像缺陷, 解决 MAE 无法直接应用于图像异常检测的问题。

1 图像异常检测方法

1.1 问题描述

图像异常检测任务描述如下: 给定一组只包含正常数据的图像集合 $x_{\text{train}} = \{a_1, \dots, a_n\}$ 和一组包含正常数据和异常数据的图像集合 $x_{\text{test}} = \{b_1, \dots, b_m\}$; 在图像重建模型的训练阶段, 通过潜在空间重建无缺陷的训练样本; 在测试阶段, 通过原始图像与重建图像的比较来判断图像是否异常。图像异常分数表示为

$$S_{\text{score}} = \frac{1}{N} \sum_{i=1}^N (x_i - \hat{x}_i)^2 \quad (1)$$

式中: N 为原始图像像素点的个数, x 和 \hat{x}_i 分别为图像重建模型重建前、后的像素值。

1.2 改进的 MAE 模型

MAE 模型训练时通过对输入的原始图像进行一定比例的随机掩码, 再经过编码和解码对掩码部分重建, 可提高编解码器的特征提取能力, 即 MAE 模型需要完成一定比例的区域重建任务。MAE 模型结构如图 1 所示。

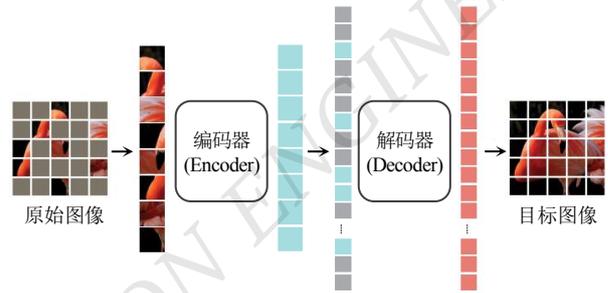


图 1 MAE 模型结构

图像异常检测需要重建图像的所有区域, 再通过对重建前后的图像来判断其是否存在异常。为重建整幅图像, 对 MAE 模型进行以下改进:

首先, 将输入的原始图像 x 分成 196 块, 待检测图像经过 n 次掩码变为 n 幅带掩码的图像, 设置掩码覆盖率为 r , 如需要完整恢复原始图像, 则掩码次数 $n = 1/(1-r)$, ($n \in N$), 其中每个掩码的补码和为 $\mathbf{1} \in \mathbb{R}^{196}$, 即

$$\overline{M}_1 + \overline{M}_2 + \dots + \overline{M}_n = \mathbf{1} \quad (2)$$

在训练过程中, 假设掩码覆盖率为 0.75, 则 1 幅待检测图像被改进的 MAE 模型处理为 4 幅图像, 一方面可使改进的 MAE 模型在一次前向过程中完成不同位置的像素恢复; 另一方面也可以通过扩大数据集, 提高模型的泛化能力, 加快模型的收敛速度。

然后, 改进的 MAE 模型从 4 幅图像中提取未被掩码的部分, 组成 1 幅图像, 输入到编码器中。

接着, 编码器进行位置信息编码和图像信息编码, 并生成 n 个特征向量 $f_{e1}, f_{e2}, \dots, f_{en}$ 。这些特征向量经过掩码的位置编码插值后, 生成 n 个带掩码信息的

特征向量 f_1, f_2, \dots, f_n ，经过解码器解码后生成图像块 y_1, y_2, \dots, y_n ，最终重建的目标图像为

$$y = \frac{1}{n-1} (M_1 \cdot y_1 + M_2 \cdot y_2 + \dots + M_n \cdot y_n) \quad (3)$$

式中： M_1, M_2, \dots, M_n 为 n 次掩码，用于只提取被掩码对应图像的重建块，因掩码覆盖率设置为 0.75，每块图像被遮盖的次数为 $4-1=3$ ，则目标图像 y 的系数为 $1/(4-1)$ 。

最后，检测阶段的掩码覆盖率与训练阶段的掩码覆盖率一致，并根据图像复原前后的差异计算图像的异常分数。

改进的 MAE 模型结构如图 2 所示。

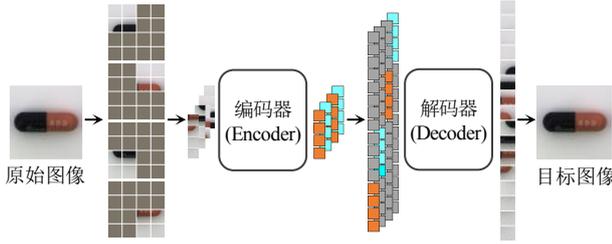


图 2 改进的 MAE 模型结构

1.3 损失函数

本文采用 $SSIM$ 、 l_1 和 $LPIPS$ 3 种损失函数来训练改进的 MAE 模型。其中， $SSIM$ 用于衡量图像的失真程度，其值越大代表图像越相似； l_1 为重建前后图像特征值的平均绝对误差，其值越大代表图像越不相似； $LPIPS$ 用于比较图像被同一个神经网络提取的特征差别，图像越不相似，其值越接近 1。损失函数为

$$\ell(x, y, \theta) = \omega_1 \cdot (1 - SSIM(x, y)) + \omega_2 \cdot l_1(x, y) + \omega_3 \cdot LPIPS(x, y) \quad (4)$$

式中： x 、 y 、 θ 分别为输入样本、重建样本和模型参数， ω_1 、 ω_2 、 ω_3 分别为 3 种损失函数的权重值。

$SSIM$ 的计算公式为

$$SSIM(x, y) = L(x, y)^\alpha \cdot C(x, y)^\beta \cdot S(x, y)^\gamma \quad (5)$$

$$\text{式中： } L(x, y) = \left(\frac{2\mu_x \mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \right)^\alpha \quad (6)$$

$$C(x, y) = \left(\frac{2\sigma_x \sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \right)^\beta \quad (7)$$

$$S(x, y) = \left(\frac{\sigma_{xy} + C_3}{\sigma_x \sigma_y + C_3} \right)^\gamma \quad (8)$$

$$u_x = \frac{1}{N} \sum_{i=1}^N x_i \quad u_y = \frac{1}{N} \sum_{i=1}^N y_i \quad (9)$$

$$\sigma_x = \left(\frac{1}{N-1} \sum_{i=1}^N (x_i - u_x)^2 \right)^{\frac{1}{2}} \quad (10)$$

$$\sigma_y = \left(\frac{1}{N-1} \sum_{i=1}^N (y_i - u_y)^2 \right)^{\frac{1}{2}} \quad (11)$$

式中： $L(x, y)$ 、 $C(x, y)$ 、 $S(x, y)$ 分别为亮度比较函数、对比度比较函数、结构比较函数， u_x 、 σ_x 分别为图像 x 中所有像素的平均值和标准差无偏估计， u_y 、 σ_y 分别为图像 y 中所有像素的平均值和标准差无偏估计， α 、 β 、 γ 分别为 3 个函数的权重值。

l_1 的计算公式为

$$l_1(x, y) = \frac{1}{N} \sum_{i=1}^N |x_i - y_i| \quad (12)$$

式中： N 为待检测图像像素点的个数， x 和 y 分别为图像重建前后的像素值。

$LPIPS(x, y)$ 的计算公式为

$$LPIPS(x, y) = \frac{1}{M} \sum_{i=1}^M \frac{1}{H_i W_i} (Model_i(x) - Model_i(y))^2 \quad (13)$$

式中： M 为神经网络的层数； i 为层索引； H 、 W 分别为模型输出特征图的高、宽； $Model$ 为神经网络，一般是 VGG 或 AlexNet，本文选取 AlexNet。

2 实验数据

2.1 数据集

为验证本文方法的有效性和准确性，在 MVTEC-AD^[11]数据集上进行定性和定量分析。该数据集包含

15 类 (5 类纹理图像和 10 类物体图像), 共 5 354 幅图像, 其中 3 629 幅图像为训练样本, 1 725 幅图像为测试样本。训练样本均为正常图像, 测试样本包含多种类别的缺陷图像和正常图像。

2.2 数据预处理

首先, 将待检测图像像素缩放成统一大小 $3 \times 224 \times 224$ (3 为通道数, 224×224 为图像的宽、高); 然后, 采用随机裁剪、随机旋转和随机反转进行数据增强, 使模型具有更强的泛化能力。

3 实验结果与分析

本实验在 4 张 GTX3080Ti GPU 上运行, 优化器采用 AdamW, 初始化学率设置为 1×10^{-3} , 训练迭代次数为 1 000 次, 训练批次为 16, warmup_epochs 为 10。

3.1 评价指标

本文采用 AUC 作为图像异常检测的评价指标。ROC 曲线由假阳率 (false positive rate, FPR) 和真阳率 (true positive rate, TPR) 绘制, AUC 为 ROC 曲线与坐标轴围成的面积。

F_{FPR} 和 T_{TPR} 的计算公式分别为

$$F_{FPR} = \frac{FP}{TN + FP} \quad (14)$$

$$T_{TPR} = \frac{TP}{TP + FN} \quad (15)$$

式中: FP 、 TP 、 FN 和 TN 分别为检测错误的正常图像数、检测正确的正常图像数、检测错误的异常图像数和检测正确的异常图像数。

3.2 实验对比与分析

为了验证本文方法的有效性, 将改进的 MAE 模型与 MKD^[12]、U-Net^[13]、DAGAN^[14]3 种异常检测模型进行对比实验。以上 3 种异常检测模型都是利用某一类的图像训练对应的异常检测模型, 因此针对 15 类数据类别有 15 种相应的模型权重。本文在训练改进的 MAE 模型过程中, 对所有训练数据进行训练, 不仅可以满足模型的数据需求量, 提高模型的重建能力,

还可以加大模型的实用性, 仅训练一次便可以检测多类图像数据的异常。本文改进的 MAE 模型与 3 种异常检测模型在测试样本上的 AUC 指标如表 1 所示。

表 1 4 种不同模型的异常检测性能比较

数据类别	AUC			
	MKD	U-Net	DAGAN	本文模型
Bottle	0.994	0.863	0.983	0.993
Cable	0.892	0.636	0.665	0.966
Capsule	0.805	0.673	0.687	0.819
Carpet	0.792	0.774	0.903	0.729
Grid	0.781	0.857	0.867	0.840
Hazelnut	0.984	0.996	1.000	0.969
Leather	0.951	0.870	0.944	0.954
Metal nut	0.736	0.676	0.815	0.851
Pill	0.827	0.781	0.768	0.863
Screw	0.833	1.000	1.000	0.804
Tile	0.916	0.964	0.961	0.926
Toothbrush	0.921	0.811	0.950	0.938
Transistor	0.855	0.674	0.794	0.843
Wood	0.943	0.958	0.979	0.984
Zipper	0.932	0.750	0.781	0.951
Average	0.877	0.819	0.873	0.895

由表 1 可知: 与 MKD 模型相比, 本文改进的 MAE 模型平均的 AUC 提高了 0.022; 改进的 MAE 模型在 15 类数据中有 7 类数据比其他模型表现得更好, 证明了改进的 MAE 模型具有较好的异常检测能力。

对编解码块数和掩码覆盖率进行消融实验, 如表 2 所示。其中, E12D8 代表在改进的 MAE 模型中 Encoder 和 Decoder 的数目分别为 12 和 8, 掩码覆盖率代表图像覆盖的部分占整幅图像面积的比例。

表 2 消融实验

掩码覆盖率	平均 AUC 指标					
	E2D2	E4D2	E6D6	E12D8	E8D12	E2D4
0.50	0.829	0.807	0.800	0.635	0.635	0.860
0.80	0.832	0.852	0.770	0.626	0.604	0.872
0.90	0.842	0.794	0.783	0.628	0.612	0.880
0.75	0.832	0.826	0.796	0.632	0.622	0.895

由表 2 可知, 当掩码覆盖率为 0.75, Encoder 和 Decoder 分别为 2 和 4 时, 改进的 MAE 模型检测效果最佳。

改进的 MAE 模型, 编解码器过少, 模型学习能力弱, 易发生欠拟合; 编解码器过多, 模型易发生过拟合; 掩码覆盖率越大, 模型的重建能力越强, 但掩盖了图像的大部分特征, 可利用的数据量变少, 降低了模型的泛化能力; 掩码覆盖率越小, 掩盖的图像特征越少, 可利用的数据量越多, 但降低了模型的重建能力。

此外, 本文还分析了测试集异常分数, MVTecAD 数据集异常分数统计表如图 3 所示。其中, 左斜线柱状条、右斜线柱状条、交叉线柱状条分别表示在异常分数区间段正常图像数、异常图像数、两类图像重叠的图像数。

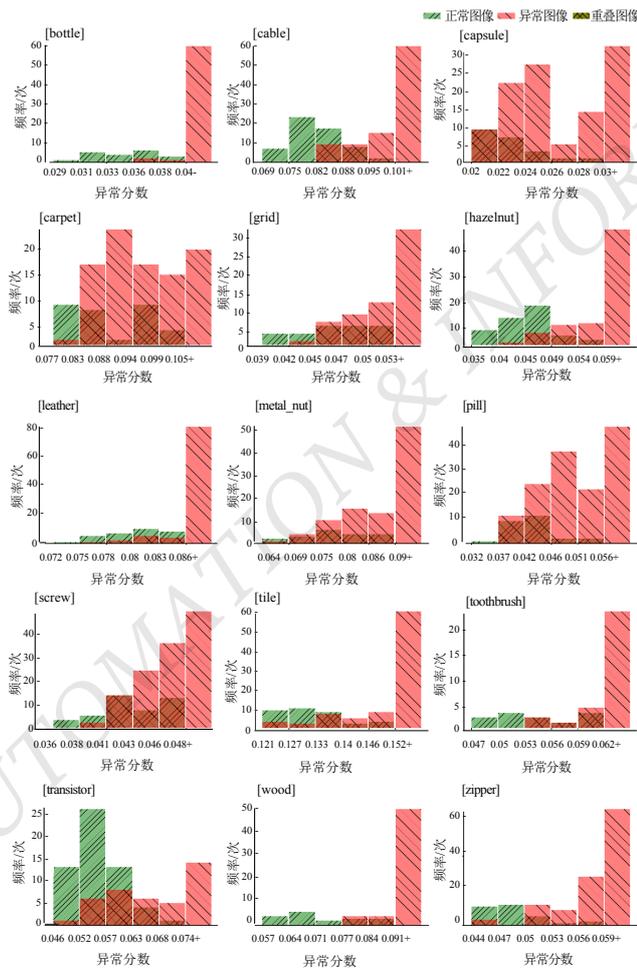


图 3 MVTecAD 数据集异常分数统计表

由图 3 可知, 15 类数据集中除了 leather 和 transistor 两类数据, 其余数据基本都是异常数据高分数频率 (子图中最右边的柱状条) 最高, 直观地反映了本文改进的 MAE 模型有效性。

4 结论

针对多类图像异常检测存在的重建能力欠佳, 无法将 MAE 模型直接应用于图像异常检测任务等问题, 本文对 MAE 模型进行改进。更改模型掩码机制, 使模型能够重建输入图像的每个像素点, 并根据图像重建前后的像素点对比进行异常图像检测。经实验验证, 改进的 MAE 模型能够适应异常检测任务, 通过 1 次训练能够同时检测 15 类数据集, 且检测效果较好, 有效缓解了现有的异常检测模型检测精度及部署效率低的状况。

©The author(s) 2024. This is an open access article under the CC BY-NC-ND 4.0 License (<https://creativecommons.org/licenses/by-nc-nd/4.0/>)

参考文献

- [1] 吕承侃, 沈飞, 张正涛, 等. 图像异常检测研究现状综述[J]. 自动化学报, 2022, 48(6): 1402-1428.
- [2] RUFF L, VANDERMEULEN R, GOERNITZ N, et al. Deep one-class classification[C]//International conference on machine learning. PMLR, 2018: 4393-4402.
- [3] VENKATARAMANAN S, PENG K C, SINGH R V, et al. Attention guided anomaly localization in images[C]//European Conference on Computer Vision. Cham: Springer International Publishing, 2020: 485-503.
- [4] REISS T, COHEN N, BERGMAN L, et al. Panda: Adapting pretrained features for anomaly detection and segmentation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 2806-2814.
- [5] LEE H, KIM N W, LEE J G, et al. Patch-level operation with adaptive patch control for improving anomaly localization[J]. IEEE Access, 2021, 9: 90727-90737.
- [6] CHEN Y, TIAN Y, PANG G, et al. Deep one-class classification via interpolated gaussian descriptor[C]//Proceedings of the AAAI Conference on Artificial Intelligence, 2022, 36(1): 383-392.

- [7] AKCAY S, ATAPOUR-ABARGHOUEI A, BRECKON T P. Ganomaly: Semi-supervised anomaly detection via adversarial training[C]//Computer Vision-ACCV 2018: 14th Asian Conference on Computer Vision, Perth, Australia, December 2-6, 2018, Revised Selected Papers, Part III 14. Springer International Publishing, 2019:622-637.
- [8] LEE Y, KANG P. AnoViT: Unsupervised anomaly detection and localization with vision transformer-based encoder-decoder [J]. IEEE Access, 2022,10:46717-46724.
- [9] DING C, PANG G, SHEN C. Catching both gray and black swans: Open-set supervised anomaly detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022:7388-7398.
- [10] HE K, CHEN X, XIE S, et al. Masked autoencoders are scalable vision learners[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022:16000-16009.
- [11] BERGMANN P, FAUSER M, SATTLEGGER D, et al. MVTec AD-A comprehensive real-world dataset for unsupervised anomaly detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 9592-9600.
- [12] SALEHI M, SADJADI N, BASELIZADEH S, et al. Multiresolution knowledge distillation for anomaly detection [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021:14902-14912.
- [13] RONNEBERGER O, FISCHER P, BROX T. U-net: Convolutional networks for biomedical image segmentation[C]// Medical Image Computing and Computer-Assisted Intervention-MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18. Springer International Publishing, 2015:234-241.
- [14] TANG T W, KUO W H, LAN J H, et al. Anomaly detection neural network with dual auto-encoders GAN and its industrial inspection applications[J]. Sensors, 2020,20(12):3336.

作者简介:

胡洋, 男, 1999 年生, 硕士研究生, 主要研究方向: 异常检测、可解释人工智能。E-mail: huyangokde@163.com

肖明, 男, 1968 年生, 博士研究生, 教授, 主要研究方向: 人工智能、物联网、大数据。E-mail: xiaoming1968@qq.com

康嘉文, 男, 1989 年生, 博士研究生, 教授, 主要研究方向: 区块链、元宇宙、人工智能。E-mail: kavinkang@gdut.edu.cn

(上接第 29 页)

- [19] CHANG S, YANG Z, YOU Y, et al. Dssnet: A deep sequential sleep network for self-supervised representation learning based on single-channel EEG[J]. IEEE Signal Processing Letters, 2022, 29: 2143-2147.
- [20] QIN X, ZHANG Z, HUANG C, et al. U2-Net: Going deeper with nested U-structure for salient object detection[J]. Pattern Recognition, 2020, 106: 107404.
- [21] RONNEBERGER O, FISCHER P, BROX T. U-net: Convolutional networks for biomedical image segmentation[C]// Medical Image Computing and Computer-Assisted Intervention-MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18. Springer International Publishing, 2015:234-241.
- [22] JIA Z, LIN Y, WANG J, et al. SalientSleepNet: Multimodal salient wave detection network for sleep staging[J]. arXiv preprint arXiv:2105.13864. 2021.
- [23] GLOROT X, BORDES A, BENGIO Y. Deep sparse rectifier neural networks[C]. Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics. JMLR Workshop and Conference Proceedings, 2011: 315-323.
- [24] GOLDBERGER A L, AMARAL L A N, GLASS L, et al. PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals[J]. Circulation, 2000, 101(23): e215-e220.
- [25] KINGMA D P, BA J. Adam: A method for stochastic optimization[J]. arXiv preprint arXiv:1412.6980, 2014.

作者简介:

陈庆端, 男, 1999 年生, 硕士研究生, 主要研究方向: 深度学习、生物信号处理。E-mail: 544689020@qq.com