

基于 Mamba 的跨数据集情绪识别方法研究

黄嘉扬¹ 翁湖钦¹ 韩伟业¹ 秦 伟^{1*}

HUANG Jiayang WENG Huqin HAN Weiye QIN Wei

摘 要

近年来, 基于脑电信号的情绪识别取得了显著进展, 但研究仍以单一数据集情绪识别为主。随着公开脑电数据集的增加, 研究逐渐转向跨数据集情绪识别。然而, 由于个体差异、环境条件和数据采集设备等因素导致数据分布存在较大差异, 使得跨数据集情绪识别面临较大挑战。文章针对跨数据集情绪识别任务, 基于 Mamba 模型提出了 CD-Mamba 模型, 并在 THU-EP 和 SEED 数据集上进行实验。为了更客观、公平地评估跨数据集情绪识别性能, 通过引入新的验证方法, 避免相同视频片段对情绪识别结果产生偏差。此外, 通过挑选不同数据集的公共通道, 确保跨数据集任务的可行性, 同时提升了模型的泛化能力及准确性。实验结果表明, CD-Mamba 模型在跨数据集情绪识别任务中取得了良好效果, 为跨数据集情绪识别研究提供了新的研究思路。

关键词

脑电信号; 情绪识别; 跨数据集; Mamba

doi: 10.3969/j.issn.1672-9528.2025.08.024

0 引言

情感脑机接口 (affective brain-computer interfaces, aBCIs) 能够让机器监测和调节人类的情绪状态, 从而促进人类的认知、沟通和决策能力。在脑机接口 (BCI) 中, 采用了多种模式, 包括立体脑电图 (stereoelectroencephalography, SEEG)、功能磁共振成像 (functional magnetic resonance imaging, fMRI) 和脑电图 (electroencephalogram, EEG)。其中, EEG 由于其非侵入性、抗干扰性及数据采集的便捷性, 在研究和实际应用中受到越来越多的关注。近年来, 基于脑电信号的情绪识别已成为一个备受关注的研究领域, 并取得了显著的研究进展^[1-5]。

目前许多情绪识别方法主要使用监督机器学习技术, 并且是基于单一数据集的跨被试情绪识别研究。跨被试情绪识别是从训练被试的数据中建立的情绪识别模型推广应用到未知被试身上, 从而实现对新被试情绪状态的准确预测。跨被试情绪识别的难点在于个体差异显著, 不同被试的脑电信号特征往往表现出明显的个性化差异。

近年来, 研究人员普遍采用深度学习模型解决跨被试情绪的问题。例如, 卷积神经网络 (convolutional neural networks, CNN) 能够从原始的脑电信号数据中提取有效的时间信息。Lawhern 等人^[6]提出了一种紧凑的卷积神经网络架构 EEGNet, 用于 EEG 信号的脑机接口深度模型, 该模型通过

引入深度卷积和可分离卷积, 显著减少了模型参数数量的同时, 并在多种 BCI 范式中表现出良好的跨范式泛化能力。然而, CNN 主要侧重于局部时间特征的提取, 难以捕捉全局信息。

为学习全局的信息, 得到更加泛化的效果, 研究者开始关注到 Transformer 架构^[7]。Song 等人^[8]提出了一种名为 EEG Conformer 模型, 该模型结合卷积神经网络和自注意力机制, 用于统一提取 EEG 信号的局部和全局特征, 显著提升了运动想象和情绪识别等 EEG 解码任务的性能, 并提供了可视化策略以增强模型的解释性。Si 等人^[9]提出了一个混合注意卷积网络与 Transformer 网络 (mixed attention-based convolution and transformer network, MACTN) 的模型, 通过用 CNN 提取局部的时间信息, 并使用通道注意力机制选取与任务相关的通道, 再由 Transformer 架构提取全局的信息, 并在多个数据集上衡量模型的情感识别性能。

上述提到的工作都是基于单数据集的验证方式, 随着公开脑电数据集的增加, 研究重点逐渐向跨数据集情绪识别方向发展。然而由于个体差异、环境条件和数据采集设备等因素导致数据分布存在较大差异, 跨数据集的泛化是脑电信号研究的一个关键挑战。因此本文是基于跨数据集情绪识别问题进行研究。

1 数据集

本文选取了两个情感识别数据集进行跨数据集情绪识别实验。

1. 五邑大学 广东江门 529020

(1) SEED^[10-11]数据集由上海交通大学教授吕宝粮于2015年10月发布,是一个基于62导联EEG信号的情感计算数据集。该数据集包含15名被试(7名男性、8名女性)的脑电图数据,通过选取15部中国电影片段来激发被试的情绪反应。每位被试均体验了3类情绪类型:积极、中性和消极。数据集采集的采样率为1 000 Hz。

(2) THU-EP^[12]数据集是2022年世界机器人大赛提供的情感计算数据集,包含80名被试(50名男性、30名女性)的32导联EEG数据。该数据集采用28段视频片段来激发情绪,覆盖了9类情绪类型:愤怒、厌恶、恐惧、悲伤、快乐、愉悦、激励、温情和中性。数据集采集的原始采样率为250 Hz。

2 数据预处理

在进行跨数据集情绪识别中的深度学习方法研究中,首先对多个数据集的脑电数据进行预处理。为保证不同数据集的样本维度相同并提高脑电数据的质量,将读取数据集的完整数据并进行统一预处理。

根据文献[9]的数据预处理,首先对脑电数据进行重参考。对给定的脑电信号片段,通过匹配不同数据集中的相同通道,确保数据能够输入统一模型。同时,对数据进行相邻通道重参考,以提高信号质量并减少噪声,即相邻通道的信号会进行相减处理,如图1所示。接下来,对脑电信号片段进行带通滤波(0.5~45 Hz)和带阻滤波(48~52 Hz),保留有用的频率信息,以有效进行情绪分类。由于两个数据集的原始采样率不同,将其统一调整为125 Hz。最后,为消除同一设备和不同设备之间的个体差异,对每个样本的电极数据进行Z-score标准化。

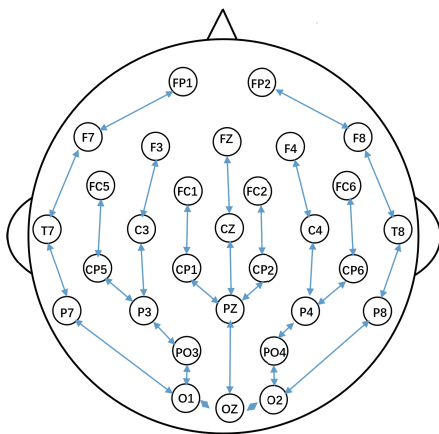


图1 相邻通道重参考

此外,为解决两个数据集的标签异构问题,将THU-EP数据集的9类转为SEED数据集对应的3类标签,即积极标签包含快乐、愉悦、激励和温情,消极标签包含愤怒、厌恶、恐惧、悲伤,而中性标签相对应,具体如表1所示。

表1 两个数据集的标签转换

SEED数据集	THU-EP数据集
积极	快乐
	愉悦
	激励
	温情
中性	中性
消极	愤怒
	厌恶
	恐惧
	悲伤

对两个数据集的视频片段的数据进行按6 s一个样本,时间窗口的步幅为4 s,即一个样本的维度是(30, 750),其中,30指30个电极通道,而750是6 s乘以125 Hz得到的样本长度。由此可得,SEED的样本数是37 575,THU-EP的样本数是34 617。由于THU-EP数据集的标签不均衡,将按最小的标签数进行对齐,最终得到的样本数是8 880。

3 CD-Mamba模型

在EEG信号处理中,时间序列信息在脑电信号中具有关键作用。EEG信号作为一种时序数据,包含了人脑活动的动态变化模式,不仅反映了大脑在特定时刻的活动状态,还揭示了不同时间点之间的依赖关系,这种特性对情绪、运动、记忆等复杂大脑活动的解析尤为重要。因此,在脑电信号特征提取中,捕捉和建模这些时间序列的依赖关系,可以更准确地描绘脑信号的时序特征,从而提升解码精度。

本文引入Mamba模块^[13],Mamba模型在计算机视觉领域具有出色的长序列建模性能,能够增强模型在时间维度上的处理效果。Mamba模块可以对时间序列进行深度特征提取,帮助模型捕捉EEG信号中复杂的跨时间依赖关系。结合CNN的局部时空特征提取能力和Mamba模块对时间序列的全局捕获,设计了CD-Mamba(cross dataset mamba)模型,该模型从不同的数据集挑选公共通道后实现了跨数据集情绪识别,如图2所示。

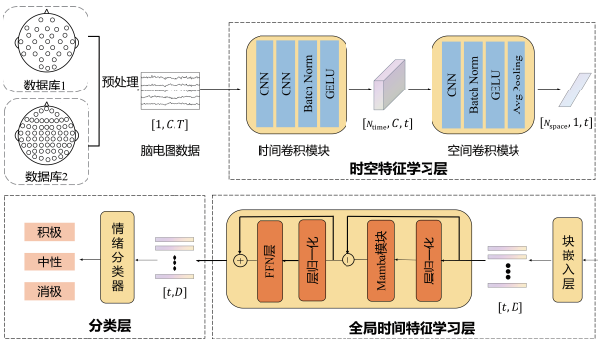


图2 CD-Mamba模型的总体架构

由于模型处理的是跨数据集问题,将评估模型性能的数

数据集定义为目标域数据集，目标域数据集划分为训练集和测试集，其中测试集用于评估跨数据集情绪识别性能；另一个数据集作为训练集，用于辅助训练，定义为源域数据集。将两个数据集合并为一个大的脑电信号数据集，对于一个脑电信号样本可以定义为 $\mathbf{x} \in \mathbf{R}^{1 \times C \times T}$ ，其中， C 表示公共电极数， T 表示该样本的时间长度对应的采样点数。所提模型框架包含 3 大主要部分：时空特征学习层、全局时间特性学习层和分类层。在时空特征学习层，使用卷积神经网络提取局部时间信息和通道间的关系；接着，在全局时间特征学习层，通过 Mamba 机制捕捉与情绪相关的全局时间信息；最后，在分类层进行情绪预测。

3.1 时空特征学习层

在情绪识别模型中，时空特征学习层旨在从 EEG 数据中有效提取局部时间信息和全局空间信息。该模块由时间卷积和空间卷积两部分组成，分别负责从时间域和空间域提取脑电信号特征。以下分别介绍时间卷积模块和空间卷积模块的设计。

时间卷积模块的主要任务是捕获 EEG 信号中的局部的时间信息。EEG 信号具有明显的时间依赖性，不同时间点的信号变化包含了丰富的情绪相关信息。通过时间卷积，可以有效提取这些时间信息，从而提升模型对情绪状态的泛化能力。具体公式为：

$$\tilde{\mathbf{x}} = \text{Conv2d}(\mathbf{x}; k_{\text{time}}, N_{\text{time}}) \quad (1)$$

式中： k_{time} 是时间卷积核的大小，表示为一维卷积核的尺寸； N_{time} 是卷积核的个数； $\tilde{\mathbf{x}} \in \mathbf{R}^{N_{\text{time}} \times C \times t}$ 是经过时间卷积后的结果； t 是卷积后的时间特征维度大小。

EEG 信号具有强烈的空间依赖性，不同电极通道之间的交互信息对情绪识别尤为重要。因此，空间卷积模块用于捕捉通道间关系，同时实现数据降维。通过这种方式，可以提高模型的计算效率，同时保留关键特征。将特征 $\tilde{\mathbf{x}}$ 作为输入，经过空间卷积层后，得到新特征，具体公式为：

$$\hat{\mathbf{x}} = \text{Conv2d}(\tilde{\mathbf{x}}; k_{\text{space}}, N_{\text{space}}) \quad (2)$$

式中： k_{space} 是空间卷积核的大小； N_{space} 是空间卷积核的个数。通过卷积操作，输出的特征维度变为 $\hat{\mathbf{x}} \in \mathbf{R}^{N_{\text{space}} \times 1 \times t}$ ，有效整合了空间维度的信息。

3.2 全局时间特征学习层

为捕获 EEG 信号中的全局时序信息，设计了全局时间特征学习层，全局时间特征学习层包含块嵌入层（Patch Embedding）和 Mamba 提取层。这一设计受到文献 [7] 的启发，旨在增强模型对全局时间特征的学习能力。

首先将特征 $\hat{\mathbf{x}}$ 变换一下维度为 $\bar{\mathbf{x}} \in \mathbf{R}^{t \times k_{\text{space}}}$ ，通过线性层 $E \in \mathbf{R}^{k_{\text{space}} \times D}$ 把每个时间点转化为嵌入块，紧接着输入到残差

结构的 Mamba 模块中学习全局时间信息，具体公式分别为：

$$\mathbf{z}_0 = [\bar{\mathbf{x}}_1 E; \bar{\mathbf{x}}_2 E; \dots; \bar{\mathbf{x}}_t E] \quad (3)$$

$$\bar{\mathbf{z}}_l = \text{Mamba}(\text{LN}(\mathbf{z}_{l-1})) + \mathbf{z}_{l-1}, l = 1, 2, \dots, L \quad (4)$$

$$\mathbf{z}_l = \text{FFN}(\text{LN}(\bar{\mathbf{z}}_l)) + \bar{\mathbf{z}}_l, l = 1, 2, \dots, L \quad (5)$$

Q 经过线性变换之后，得到的特征 $\mathbf{z}_0 \in \mathbf{R}^{t \times D}$ ， D 是线性变换后的嵌入块的维度；LN 是层归一化（Layer Normalization），FFN 是由多层感知机（MLP）组成，对学习后的特征进行线性变换。定义 Mamba 提取层有 L 层，最后一层的输出特征为 $\mathbf{z}_L \in \mathbf{R}^{t \times D}$ ，在全局时间特征学习层的特征输入和输出维度是不会改变的。

3.3 分类层

最后，分类层主要由一层线性层组成，具体公式为：

$$\mathbf{h} = \text{BatchNorm}(\text{AvgPool}(\mathbf{z}_L)) \quad (6)$$

$$\tilde{\mathbf{y}} = \mathbf{h} \mathbf{w} + \mathbf{b} \quad (7)$$

式中：BatchNorm 是批标准化；AvgPool 是平均池化，用于降维，得到特征向量 $\mathbf{h} \in \mathbf{R}^{1 \times D}$ ； $\mathbf{w} \in \mathbf{R}^{D \times N_{\text{class}}}$ 和 $\mathbf{b} \in \mathbf{R}^{1 \times N_{\text{class}}}$ 是线性层的参数，用于预测分类； N_{class} 是指标签类别数。

4 实验

本文以 SEED 数据集作为目标域，将其划分为训练集和测试集，使用测试集评估模型的情绪识别性能；同时，将 THU-EP 数据集作为源域，用于训练集的部分数据。

4.1 实验细节

本实验的训练过程中，采用了 SGD（stochastic gradient descent）优化器，训练的轮次（epoch）统一用 50，批次大小（batch size）设置为 64，学习率是 0.001，损失采用交叉熵损失函数，而 CD-Mamba 模型的超参数如表 2 所示。

表 2 CD-Mamba 模型参数

超参数	值
D	80
L	4
N_{time}	40
k_{time}	(1, 25)
N_{space}	40
k_{space}	(30, 1)

文献 [14] 对 SEED 数据集进行了进一步划分，以确保训练集和测试集使用的视频刺激不一致，从而减少相同刺激视频对模型性能的影响，如图 3 所示。此外，为了更合理、客观地评估模型，采用了基于跨被试的五折交叉验证方法。由于 SEED 数据集的 3 个会话中使用的是相同的视频片段，为避免相同视频对实验结果的干扰，根据文献 [14] 提出的验证方法，将 SEED 数据集中的 15 个刺激视频进行划分：训练集

中被试的前 9 个视频片段的脑电数据作为训练集，测试集中被试的后 6 个视频片段的脑电数据作为测试集。这一划分方式有助于消除视频内容带来的影响，从而更准确地评估模型的性能。

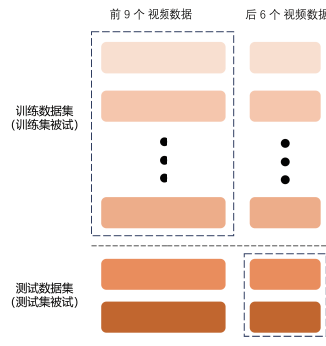


图 3 新的验证方法

4.2 对比方法

使用以下方法对比所提模型，结果为：

- (1) EEGNet^[6]：通过深度卷积和可分离卷积减少模型参数量，具有良好的跨任务泛化能力和对有限数据的适应性。
- (2) DeepConvNet^[15]：采用浅层和深层卷积神经网络对 EEG 数据集进行分类，从而提升了分类准确性。
- (3) TSception^[16]：通过动态时间层、非对称空间层及高阶融合层学习 EEG 的空间非对称性和时间信息，从而实现情绪状态的分类。

表 3 不同模型在 SEED 数据集及跨数据集下的平均准确率

模型	通道数	SEED 数据集			
		Session 1	Session 2	Session 3	平均
EEGNet	30	55.82	52.16	53.53	53.84
DeepConvNet		46.55	47.59	48.90	47.68
TSception		55.82	49.52	54.74	53.36
CD-Mamba		59.92	52.90	56.09	56.30
EEGNet	62	55.97	52.94	54.88	54.60
DeepConvNet		55.86	56.27	57.77	56.63
TSception		56.62	55.31	56.68	56.20
CD-Mamba		57.99	55.91	56.29	56.73

模型	通道数	THU-EP->SEED 数据集			
		Session 1	Session 2	Session 3	平均
EEGNet	30	56.34	50.85	53.78	53.66
DeepConvNet		53.45	50.67	53.24	52.45
TSception		56.31	50.93	54.62	53.95
CD-Mamba		61.66	56.15	58.00	58.60

表 3 展示了 3 次 Session 的实验结果，平均结果表明，所提模型在跨数据集使用下，CD-Mamba 模型的准确率比单数据集 62 通道的 CD-Mamba 模型高出 1.87%，比单数据集

30 通道的 CD-Mamba 模型高出 2.3%。这表明所提模型充分利用了其他数据集的数据信息。此外，模型比其他 3 个模型最高准确率都高，相比 EEGNet 模型，所提模型高 4%；相比 DeepConvNet 模型，所提模型高 1.97%；比 TSception 模型高 2.4%，说明所提模型在跨数据集的情况下有显著提升。

模型的混淆矩阵如图 4 所示，通过对 Session 1、Session 2 和 Session 3 的混淆矩阵进行分析可以看出，模型在识别积极情绪方面表现稳定且良好，正确率在 3 个 Session 上的结果分别为 75%、71% 和 71%，显示出对积极类型的情绪有较强的分类能力。同时，模型对中性情绪的识别也表现出较好的性能，整体趋势仍较为稳定。对于消极情绪，模型的识别结果稍有起伏。总体而言，模型在识别积极情绪上表现优异，对中性和消极情绪的分类也具备一定准确性，未来可以通过进一步优化中性和消极情绪间的细微差异来提升整体性能。

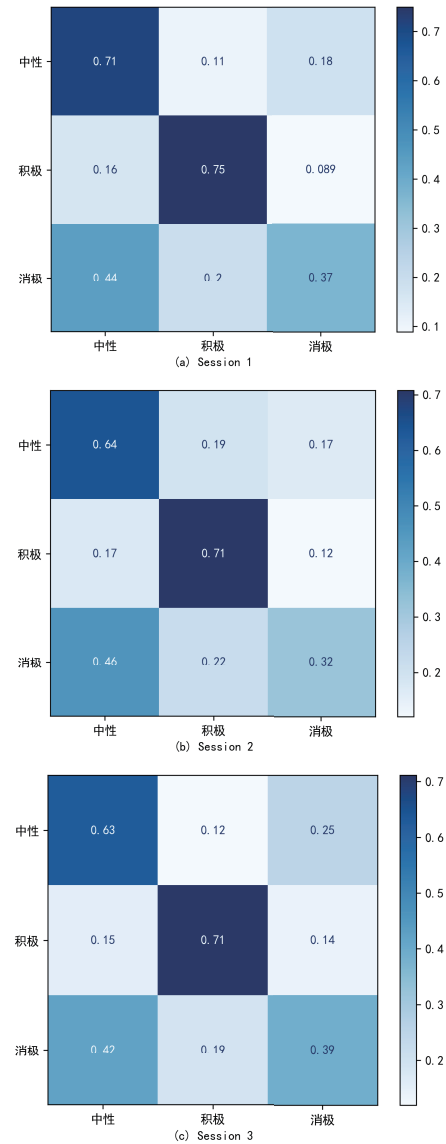


图 4 CD-Mamba 的混淆矩阵

5 结论

本文通过采用公共通道和公共标签映射的方法,解决了跨数据集情绪识别中标签异构和通道异构的问题。同时,通过统一的预处理流程,确保不同通道数的数据集能够在同一模型下进行训练。此外,为避免训练集和测试集的刺激视频存在相同视频片段作为情绪刺激引发潜在干扰,在跨被试实验的基础上,采用新的数据集划分策略。实验结果表明,在跨数据集场景下,模型性能得到了提升。该研究对解决跨数据集情绪识别问题具有重要的参考价值。

参考文献:

- [1] ALI M, MOSA A H, AL MACHOT F, et al. EEG-based emotion recognition approach for e-healthcare applications[C]//2016 eighth international conference on ubiquitous and future networks (ICUFN). Piscataway:IEEE,2016: 946-950.
- [2] XU J H, ZHONG B C. Review on portable EEG technology in educational research[J]. Computers in human behavior, 2018, 81: 340-349.
- [3] ZHANG J H, YIN Z, CHEN P, et al. Emotion recognition using multi-modal data and machine learning techniques: a tutorial and review[J]. Information fusion,2020,59: 103-126.
- [4] CHEN C Q, LI Z C, KOU K I, et al. Comprehensive multisource learning network for cross-subject multimodal emotion recognition[J].IEEE transactions on emerging topics in computational intelligence, 2025,9(1):365-380.
- [5] WU D R, LU B L, HU B, et al. Affective brain-computer interfaces (aBCIs): a tutorial[J]. Proceedings of the IEEE, 2023, 111(10): 1314-1332.
- [6] LAWHERN V J, SOLON A J, WAYTOWICH N R, et al. EEGNet: a compact convolutional neural network for EEG-based brain-computer interfaces[J].Journal of neural engineering, 2018,15(5): 56013.
- [7] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16x16 words: transformers for image recognition at scale[EB/OL].(2021-06-03)[2025-02-11].<https://doi.org/10.48550/arXiv.2010.11929>.
- [8] SONG Y H, ZHENG Q Q, LIU B C, et al. EEG Conformer: convolutional transformer for EEG decoding and visualization[J].IEEE transactions on neural systems and rehabilitation engineering, 2022, 31: 710-719.
- [9] SI X P, HUANG D, LIANG Z, et al. Temporal aware mixed attention-based convolution and transformer network for cross-subject EEG emotion recognition[J]. Computers in biology and medicine,2024, 181: 108973.
- [10] DUAN R N, ZHU J Y, LU B L. Differential entropy feature for EEG-based emotion classification[C]//2013 6th International IEEE/EMBS Conference on Neural Engineering (NER). Piscataway:IEEE,2013: 81-84.
- [11] ZHENG W L, LU B L. Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks[J].IEEE transactions on autonomous mental development, 2015, 7(3): 162-175.
- [12] HU X, WANG F, ZHANG D. Similar brains blend emotion in similar ways: neural representations of individual difference in emotion profiles[J]. Neuroimage, 2022, 247: 118819.
- [13] GU A, DAO T. Mamba: linear-time sequence modeling with selective state spaces[EB/OL].(2024-05-31)[2024-11-26]. <https://doi.org/10.48550/arXiv.2312.00752>.
- [14] HU H Y, ZHAO L M, LIU Y Z, et al. A novel experiment setting for cross-subject emotion recognition[C]//2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). Piscataway: IEEE, 2021: 6416-6419.
- [15] SCHIRRMESTER R T, SPRINGENBERG J T, FIEDERER L D J, et al. Deep learning with convolutional neural networks for EEG decoding and visualization[J].Human brain mapping, 2017, 38(11): 5391-5420.
- [16] DING Y, ROBINSON N, ZHANG S, et al. TSception: capturing temporal dynamics and spatial asymmetry from EEG for emotion recognition[J].IEEE transactions on affective computing, 2023,14(3): 2238-2250.

【作者简介】

黄嘉扬(1996—),男,广东汕尾人,硕士研究生,研究方向:类脑计算。

翁湖钦(1998—),男,广东汕头人,硕士研究生,研究方向:类脑计算。

韩伟业(2001—),男,广东汕尾人,硕士研究生,研究方向:类脑计算。

秦伟(1979—),通信作者(email: qinw05080@gmail.com),男,湖南益阳人,博士,研究方向:机器学习、图像识别、多源信息融合等。

(收稿日期: 2025-02-19 修回日期: 2025-06-30)