

文章编号:1671-6833(2024)05-0069-08

# 基于改进 StackCNN 网络和集成学习的脑电信号 视觉分类算法

杨青<sup>1,2,3</sup>, 王亚群<sup>1,2,3</sup>, 文斗<sup>1,2,3</sup>, 王莹<sup>1,2,3</sup>, 王翔宇<sup>1,2,3</sup>

(1. 华中师范大学 人工智能与智慧学习湖北省重点实验室,湖北 武汉 430079;2. 华中师范大学 计算机学院,湖北 武汉 430079;3. 华中师范大学 国家语言资源监测与研究网络媒体中心,湖北 武汉 430079)

**摘要:** 针对直接使用图像诱发的脑电信号进行视觉分类的现有研究少,并且视觉分类的平均准确率低等问题,设计了一种卷积神经网络(CNN)和集成学习相结合的方法,用于学习脑电信号相关的视觉特征表示。通过在 StackCNN 网络中加入 K-max 池化方法,解决在提取脑电特征时信息丢失的问题,并结合 Bagging 算法增强网络的泛化能力,该方法称为 StackCNN-B。采用基于残差神经网络(ResNet)回归对图像进行分类,验证 StackCNN-B 方法在图像分类上的性能。消融实验及与现有研究对比实验的结果表明:所提方法识别准确率较高,在学习脑电信号的视觉特征表示上的平均准确率达到 99.78%,在图像分类上的平均准确率达到 96.45%,与 Bi-LSTM-AttGW 方法相比,平均提高了 0.28 个百分点和 2.97 个百分点。研究结果验证了脑电信号可以有效地解码与视觉识别相关的人类大脑活动,也表明所提出 StackCNN-B 模型的优越性。

**关键词:** 脑电图;视觉分类;卷积神经网络;Bagging 算法;ResNet 网络

中图分类号:TP399 文献标志码:A doi:10.13705/j.issn.1671-6833.2024.02.009

人类在视觉识别方面有很出色的表现,这是机器无法做到的。尽管卷积神经网络(convolutional neural networks, CNN)的出现和大规模应用使得机器视觉分类的性能有了明显提高,但它们的泛化能力仍无法达到人类水平。

脑电图(electroencephalogram, EEG)是从人脑中记录的电信号。目前大多数研究集中在脑电信号处理上,它可以帮助我们理解脑电活动与电信号之间的关系。研究人员已经发现大脑活动记录包含有关视觉对象类别的信息<sup>[1-3]</sup>。因此,通过 EEG 数据和神经成像技术(如核磁共振成像)可以识别人类用于视觉分类的特征空间。近些年的研究很少直接使用图像诱发的 EEG 信号进行视觉分类。分析 EEG 数据可以使用深度学习方法从复杂信号中分析、学习和提取复杂模式。El-lone 等<sup>[4]</sup>发表了关于使用密集脑电图数据进行视觉识别的研究。刺激数据集包含 400 幅图像,使用 256 个电极的密集脑电

图系统记录大脑信号,平均分类率为 82.70%。Parekh 等<sup>[5]</sup>提出了一种 CNN 结构驱动的脑电图像标注系统。用 2 500 幅图像作为训练刺激源,F1 得分为 0.88。Spampinato 等<sup>[6]</sup>训练递归神经网络(recurrent neural networks, RNN)学习脑电波视觉表示,并且使用基于 CNN 的回归器将学习能力转移到机器上,获得 83% 的准确率。Zheng 等<sup>[7]</sup>提出 LSTM 结合集成学习的模型,使用 40 个类别的 2 000 个图像进行特征提取,特征提取准确率为 97.13%。Zheng 等<sup>[8]</sup>提出了一种基于端到端的注意力机制的 Bi-LSTM 方法,在 Bi-LSTM 框架中引入了两种注意力策略,获得了 99.50% 的准确率。在最近的研究中,Mishra 等<sup>[9]</sup>设计了一种将脑电图信号编码为图像的有效方法,对 39 个图像类对应的编码脑电图信号进行分类,在 6 个主题的分层数据集上进行图像分类,获得了 82% 的准确率。Khaleghi 等<sup>[10]</sup>提出基于功能连通性的几何深度神经网络(FC-GDW),利

收稿日期:2023-12-03;修订日期:2024-01-10

基金项目:湖北省重点研发计划项目(2020BAB017);武汉市科技计划项目(2019010701011392);国家语委科研中心项目(ZDI135-135)

作者简介:杨青(1965—),女,湖北武汉人,华中师范大学副教授,博士,主要从事数据挖掘及计算机技术应用方面的研究,E-mail:yangqing@cnu.edu.cn。

引用本文:杨青,王亚群,文斗,等. 基于改进 StackCNN 网络和集成学习的脑电信号视觉分类算法[J]. 郑州大学学报(工学版),2024,45(5):69-76. (YANG Q, WANG Y Q, WEN D, et al. EEG visual classification algorithm based on improved StackCNN network and ensemble learning[J]. Journal of Zhengzhou University (Engineering Science), 2024, 45(5):69-76.)

用图像诱发脑电信号中的时空分布式信息直接提取高分辨率的时间样本隐藏状态,在 EEG-ImageNet 数据集上的平均准确率为 98.4%。以上这些研究表明,利用 EEG 信号对视觉场景进行分类是可行的,结合深度学习的研究结果为进一步探索脑电在视觉分类场景的应用指明了方向,但很少有文献针对所提出模型的泛化能力进行研究,验证模型在图像分类上是否具有一定的应用价值,并且基于 EEG 的现有研究在图像分类上的准确率仍有很大的提高空间。综上,本文针对现有研究存在的问题设计改进了 TextCNN 和 StackCNN 模型,加入  $K$ -max 池化,解决在提取脑电特征时信息丢失问题,并应用在 EEG 特征提取上,显著提升了分类性能。然后分别将 TextCNN 和 StackCNN 模型与 Bagging 算法结合,进一步提高模型的泛化能力,最后将表现更优的 StackCNN-B 方法应用到基于 ResNet 的视觉分类模型上,将图像特征映射到脑电特征,使机器更好地理解视觉场景。

## 1 模型设计

当受试者观看屏幕上显示的图像时,会产生相应的脑电信号。脑电信号可以传递视觉相关信息,可以用于识别不同的图像,从而理解不同的图像内容。EEG 是一种多通道的时间序列类型信号,其中包含一些噪声,因此须利用多通道的序列提取低维和有意义的特征表示视觉相关的信息。当提取到了表示大脑视觉活动的脑电特征后,可以将脑电图像映射到相应的脑电特征上,从而实现视觉分类。

基于上述思想,图像分类过程由两个阶段组成,如图 1 所示。第一阶段是解码人类大脑活动,由于每个脑电序列都对应着相应的图像类别,因此提取的脑电特征可以用于图像分类。第一阶段的目的是

从原始 EEG 信号中提取到低维的 EEG 特征和训练 Softmax 分类器,本文采用基于改进的两种 CNN 模型结合改进的 Bagging 算法实现此目的。第二阶段是实现自动视觉分类,基于上一阶段提取到的 EEG 特征,并在此阶段训练基于残差神经网络(residual network, ResNet)的回归,将提取到的图像特征直接映射到上一阶段提取到的 EEG 特征上,通过上一阶段训练好的 softmax 分类器可以估计新图像的 EEG 特征,并对其进行预测分类。

### 1.1 基于改进的 TextCNN 和 StackCNN 模型设计

本文使用两种不同的 CNN 架构结合受图像刺激的 EEG 数据进行分析。

(1) TextCNN: 利用多个大小不同的卷积核提取 EEG 序列中的关键信息。

(2) StackCNN: 利用多个堆叠的卷积核提取脑电信号特征,并将低维和高维的特征合并。

#### 1.1.1 $K$ -max 池化

上述两种 CNN 结构均使用了最大池化(max-pooling)方法,该池化运算仅取特征向量中的最大值,存在两个明显的缺陷:一是未考虑任何特征的位置信息;二是容易忽视特征的强度,如果得分最高的特征只出现了一次,而得分第二的特征出现多次,得分第二的特征比得分最高的特征更为重要,但却被丢弃,从而造成信息损失。

为解决最大池化带来的问题,本文选择使用  $K$ -max 池化( $K$ -max pooling)<sup>[11]</sup>,此方法选取得分最高的前  $K$  个特征,在一定程度上改善了上述问题。 $K$ -max 池化可以表达同一类特征出现多次的情形,即表达某类特征的强度。另外,因为这些特征值的相对顺序得以保留,所以其保留了部分位置信息,但这种位置信息只是特征间的相对顺序,而非绝对位置信息。

#### 1.1.2 改进的 TextCNN 模型

TextCNN 最初由 Kim<sup>[12]</sup> 提出,是一种非常有用的

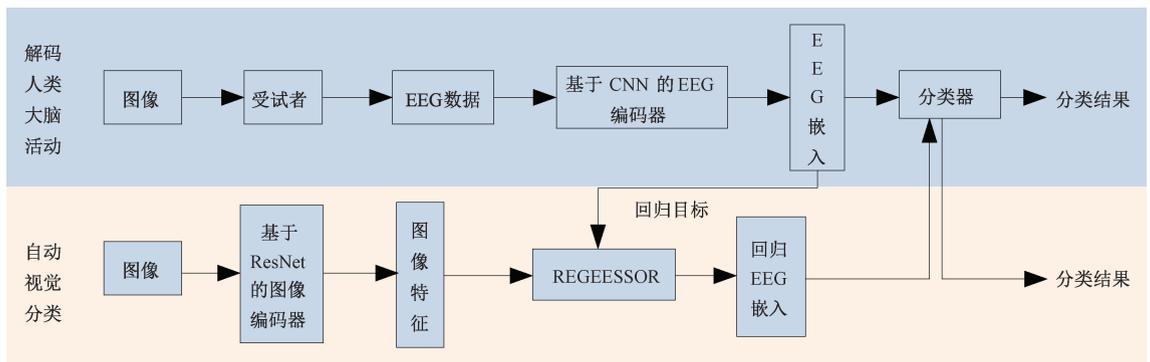


图 1 图像分类体系

Figure 1 Image classification system

和有效的深度学习算法,用于文本分类任务<sup>[13-14]</sup>。

TextCNN 由一维卷积层和时序最大池化层组成,为了保留更多重要的特征,这里由  $K$ -max 池化层替代,其网络结构如图 2 所示。在一维卷积运算中,卷积窗口从输入数组的最左方开始,按从左往右的顺序,依次在输入数组上滑动。当卷积窗口滑动到某一位置时,窗口中的输入子数组与核数组按元素相乘并求和,得到输出数组中相应位置的元素。

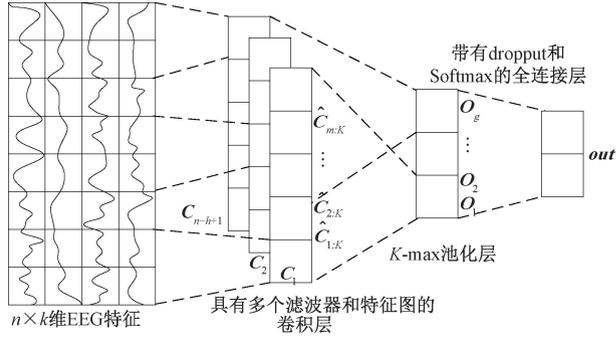


图 2 TextCNN 网络结构

Figure 2 TextCNN network structure

令  $x_i \in \mathbf{R}^k$  是  $k$  通道 EEG 数据中第  $i$  个采样点。这样一个受试者在一张图像上产生的 EEG 数据可以表示为

$$x_{1:n} = x_1 \oplus x_2 \oplus \dots \oplus x_n \quad (1)$$

式中: $\oplus$ 为连接运算符。这样, $x_{i:i+j}$ 就表示 EEG 数据采样点  $x_i, x_{i+1}, \dots, x_{i+j}$  的连接。一维卷积运算使用一个过滤器  $w \in \mathbf{R}^{hk}$ ,它应用于  $h$  个 EEG 样本点的窗口,并产生新的 EEG 特征。例如,一个 EEG 特征  $C_i$  由一个 EEG 样本点窗口  $x_{i:i+h-1}$  生成:

$$C_i = f(w \cdot x_{i:i+h-1} + b) \quad (2)$$

式中: $b$ 为偏置项; $f(\cdot)$ 为一个非线性函数,如 ReLU 函数。这个过滤器被用于 EEG 序列  $\{x_{1:h}, x_{2:h+1}, \dots, x_{n-h+1:n}\}$  每个可能的样本点窗口,并产生一个 EEG 特征向量:

$$C = [C_1, C_2, \dots, C_{n-h+1}] \quad (3)$$

式中: $C \in \mathbf{R}^{n-h+1}$ 。然后对  $C$  应用  $K$ -max 池化运算,选取得分最高的前  $K$  个特征  $\hat{C} = \max(K, C)$ 。这个想法是为每个 EEG 特征向量捕捉最重要的前  $K$  个特征。

一个过滤器有  $m$  个输出通道,每个输出通道产生一个 EEG 特征向量。对每个特征向量应用  $K$ -max 池化,将所有特征向量结果的前  $K$  个值连接起来,得到最后的输出向量:

$$O = \hat{C}_{1:K} \oplus \hat{C}_{2:K} \oplus \dots \oplus \hat{C}_{m:K} \quad (4)$$

TextCNN 使用多个不同大小的过滤器获取不同宽度的视野,从而捕捉到不同个数的相邻 EEG 的相

关性。令有  $g$  个过滤器,把它们池化后的向量拼接起来得到:

$$out = O_1 \oplus O_2 \oplus \dots \oplus O_g \quad (5)$$

最后,把  $out$  馈入全连接层,再使用 Softmax 计算,即可得到分类结果。

### 1.1.3 改进的 StackCNN 模型

StackCNN 和 TextCNN 不同的地方在于 TextCNN 的过滤器为横向排列,彼此之间互不干扰,而 StackCNN 的过滤器为纵向排列,即前一个过滤器的输出作为后一个过滤器的输入,并且 StackCNN 的结构由堆积的小过滤器<sup>[15]</sup>构成,这可以增加网络深度以保证学习更复杂的模式,而且代价更小(参数更少)。StackCNN 将所有过滤器的输出进行池化运算,同样为了保留更多重要的特征,这里由  $K$ -max 池化替代普通池化运算,然后,将输出向量连接起来;最后,馈入完全连接层。

StackCNN 和 VGG<sup>[16]</sup>网络类似,所有过滤器都使用  $3 \times 3$  的小过滤器,填充为 1,步幅为 1,保证每次输出的特征向量维度和输入维度相同。StackCNN 网络结构如图 3 所示,图 3 中输入  $n \times k$  维的 EEG,这里  $n$  为 1,  $k$  为 128。

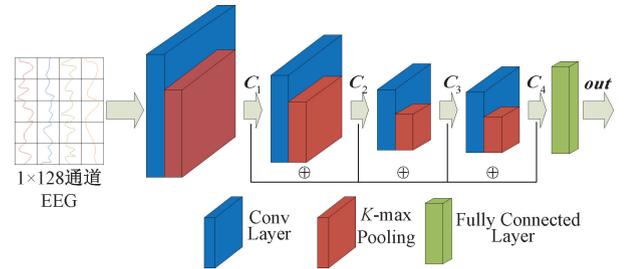


图 3 StackCNN 网络结构

Figure 3 StackCNN network

对于 EEG 序列  $\{x_{1:h}, x_{2:h+1}, \dots, x_{n-h+1:n}\}$ , 假设有  $l$  个过滤器,则有:

$$\begin{cases} C_1 = \text{Conv}_1(w_1, x) + b_1; \\ C_2 = \text{Conv}_2(w_2, C_1) + b_2; \\ \vdots \\ C_l = \text{Conv}_l(w_l, C_{l-1}) + b_l. \end{cases} \quad (6)$$

式中: $\text{Conv}(\cdot)$ 为过滤器; $b$ 为偏置项。对每个 EEG 特征图进行  $K$ -max 池化运算,然后连接起来:

$$out = \max(K, C_1) \oplus \max(K, C_2) \oplus \dots \oplus \max(K, C_l) \quad (7)$$

StackCNN 网络结构中位置靠前的过滤器用于捕捉 EEG 数据之间的局部相关性,可以看作是学习到了 EEG 低维特征,而越到后面的过滤器感受野越大,从而捕捉输入上更大尺寸的 EEG 特征,可以看作是学习到了 EEG 高维特征。因此,StackCNN 可

以同时利用低维和高维的特征,这也是它优于 Text-CNN 的地方。

## 1.2 改进的 Bagging 算法

集成学习属于机器学习,是一种训练思路,并不是指某种具体的方法或者算法。集成学习潜在的思想是即便某一个弱分类器得到了错误的预测,其他的弱分类器也可以将错误纠正回来。

自助法 (bootstrapping) 和引导聚集算法 (bootstrap aggregating, Bagging) 均是集成学习领域的一种基本算法<sup>[17]</sup>,均用于确定训练集和测试集,提高训练有效性。Bagging 算法可与其他分类算法、回归算法结合,提高其准确率、稳定性的同时,通过降低结果的方差,避免过拟合的发生。然而,在传统的 Bagging 算法中,通常使用投票法进行决策,这使得所有的分类器具有相同的决策能力,因而忽略了基分类器的性能差异。本文使用一种改进的 Bagging 算法<sup>[7]</sup>进行集成学习,改进的 Bagging 算法利用基分类器在袋外数据 (out of bag, OOB) 上每个类别的  $F_1$  值确定基分类器在每个类别上的权重,在很细的粒度上考虑了每个基分类器的性能差异,同时避免了对性能好的基分类器的过度依赖,OOB 指的是训练过程中缺失的数据点,由于这些数据没有参与训练集模型的拟合,因此可以用来检测模型的泛化能力。本文对每个类别的权重定义如下:

$$\lambda_{in} = \frac{OOB\_F_{in}}{\sum_{n=1}^N OOB\_F_{in}} \quad (8)$$

式中:  $i$  表示不同的类别 (对于本文使用的数据集,  $i$  的范围从 0 到 39);  $\lambda_{in}$  为第  $n$  个最弱分类器对第  $i$  个类别的权重分配;  $OOB\_F$  为类别  $i$  上第  $N$  个最弱分类器的测试集数据的  $F$  值;  $N$  为基分类器的数量。此外如果分类器在数据集上的性能越好,  $\lambda_{in}$  值则越大。

Zheng 等<sup>[7]</sup>在改进的 Bagging 算法中采用自助法确定训练集和测试集,结果发现,在针对 40 个类别的图像数据集的实验中, Bagging 算法中的分层抽样策略比原先实验取得了更好的效果,如表 1 所示。因此,本文使用了分层抽样法采样进行训练集和测试集的划分。

另外,分层抽样法的采样次数是需要仔细考虑的重要因素。过多的采样次数可能导致过拟合,而太少的采样次数则可能导致欠拟合。因此,本文使用  $k$  折交叉验证方法来确定最佳采样次数,对于  $N$  个基分类器,在  $k$  个类别上进行分类,采样 10 折交叉验证进行分层抽样,如表 2 所示,从表 2 中可以看

出,当采样次数为 20 时,所有基分类器的准确率的平均值达到了最佳水平。因此,最佳采样次数为 20。

表 1 自助法与分层抽样法采样结果对比

采样方法	准确率	召回率	F1 值
自助法	97.13	96.96	96.94
分层抽样法	97.82	97.02	97.54

表 2 不同采样次数在所有基分类器上的结果对比 (取平均值)

采样次数	准确率	召回率	F1 值
5	85.14	86.11	83.22
10	87.12	87.01	85.33
15	90.24	87.74	89.02
20	96.87	95.88	96.01
25	95.35	95.21	93.72

## 1.3 解码人脑活动的 StackCNN-B 模型

基于改进的 Bagging 算法和改进的 StackCNN 模型,本文提出了 StackCNN-B 模型,模型数据流图如图 4 所示。

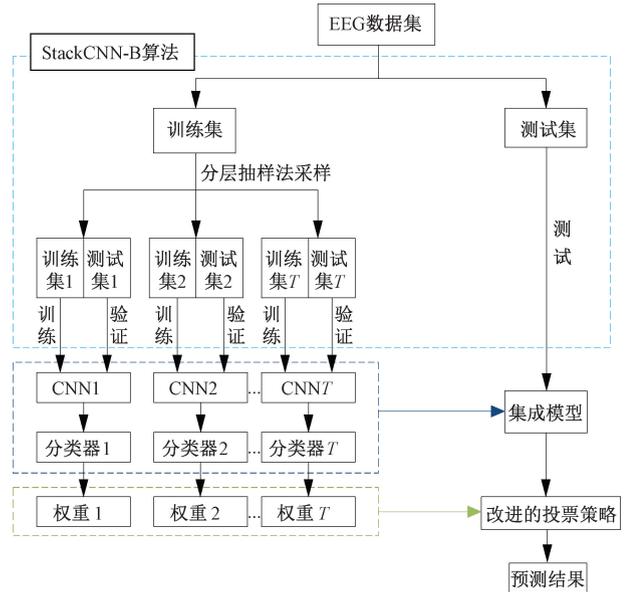


图 4 StackCNN-B 模型流程图

Figure 4 StackCNN-B model flow chart

模型具体实现步骤如下。

步骤 1 将脑电图像数据分为训练集和测试集两部分。

步骤 2 使用 Bagging 算法从训练集中通过分层抽样获得  $T$  个子训练集和测试集。

步骤3 分别训练  $T$  个 CNN 模型,即  $T$  个弱分类器,并将结果输入 Softmax 分类器进行分类。

步骤4 根据训练得到的权重集成模型,使用改进的投票策略在测试集上预测结果。

#### 1.4 EEG 特征回归和视觉分类

本模型首先通过使用结合改进的 Bagging 算法,从 EEG 数据中提取相关的脑电特征表示用于解码人类的大脑活动;其次,使用原始图像数据集和之前提取到的 EEG 特征训练基于 ResNet 的回归,将回归的输出作为特征对图像进行分类,从而实现视觉分类。

完成第一阶段模型的构建和训练后,每个 CNN 的输出(每个 Softmax 分类器的输入)可以被看作是 EEG 的视觉特征向量。为了将集成 CNN 模型学习到的脑电特征扩展到一般的图像中,需要能够跳过脑电记录阶段,直接从图像中提取与视觉相关的脑电特征。为了实现这一假想,使用 ResNet 回归从输入的图像中提取近似的 EEG 特征,ResNet 回归由 ResNet 模型和回归网络组成。ResNet 最初是由 He 等<sup>[18]</sup>提出,因其简单实用,被广泛应用于图像检测、分割和识别<sup>[19]</sup>。它将残差学习应用到每一层,解决了由于网络过深神经网络训练困难的问题。ResNet 通过引入快捷连接,解决了精度退化,即随着网络层数的加深,训练精度快速降低。

本文采用 ResNet-18 模型作为脑电特征提取器,它包含 17 个卷积层和一个完全连接层。基于 ResNet-18 的训练过程包含两个步骤。第一步,通过预训练的 ResNet-18<sup>[20]</sup>模型提取图像的全局和局部特征;第二步,利用回归方法将图像特征映射到 EEG 特征向量。本文对比了常见的几种回归方法<sup>[21-22]</sup>:随机森林回归(random forest regressor)、决策树回归(decision tree regressor)、 $K$  邻近回归( $K$  neighbors regressor)、支持向量机回归(support vector regressor)。由于回归层单元数目与 CNN 模型中学习到的 EEG 特征维数相同,实现了可以从任意图像中提取脑电信号特征。

获得脑电信号后,使用经过训练的 ResNet 对测试集进行回归,提取测试图像集的 EEG 特征,并使用训练过的 Softmax 分类器对特征进行分类。

## 2 实验与分析

实验分为两个部分:一是使用基于改进的 CNN 模型和集成学习相结合的方法学习 EEG 视觉特征;二是使用基于 ResNet 的 EEG 特征回归进行自动视觉分类。

代码使用 Pytorch 深度学习框架进行模型搭建和训练。实验的具体流程如下。首先,在采集到的 EEG 数据上训练模型,学习到 EEG 数据的嵌入表示,再将这个嵌入表示输入到分类器中进行分类;其次,训练一个回归模型将从图像中提取到的特征映射到 EEG 嵌入上;最后,将提取的特征输出到分类器中进行分类。

#### 2.1 EEG 数据集及其预处理

本文使用的数据是由 Spampinato 等<sup>[6]</sup>通过实验收集整理所得到的。该实验共有 6 名参与者(1 名女性和 5 名男性),通过仪器记录参与者的 EEG 数据。当参与者看到视觉刺激图像时,他们的脑电图信号被记录下来。参与者均由专业医生进行检查,排除可能影响实验数据的身体状况。

视觉刺激数据集是 ImageNet<sup>[23]</sup>数据集的一个子集,包含 40 个互不相同且易于识别的类别,共 2 000 个图像,其中每个类别包含 50 个图像。在脑电采集过程中,每幅图像显示 0.5 s,每一类图像连续显示 25 s 作为一批。随后,受试者休息 10 s,并显示一个黑色图像。黑色图像被用于“清除”上一节实验所产生的大脑活动,其中,受试者观看顺序为线性。因此,每次实验只需 1 400 s。实验参数如表 3 所示,该表来自 Spampinato 等<sup>[6]</sup>的视觉分类实验参数。

表 3 实验参数

参数	取值
图像类别数量	40
每种类别的图像数量	50
图像总数量	2 000
每张图像显示时间/s	0.5
每类图像之后的暂停时间/s	10
会话数量	4
单个会话持续时间/s	350
总时间/s	1 400

EEG 数据通过一个 128 通道的有源低噪声电极设备进行采集,且采集过程中使用了脑视觉信号放大器。在运行时设置了陷波滤波器(49~51 Hz)和二阶带通滤波器(截止频率分别为 14、71 Hz),使脑电信号包含视觉认知过程中最有价值的频带( $\beta$  和  $\gamma$ )。采样频率设置为 1 000 Hz,数据分辨率为 16 位。

理论上应该在 1 000 Hz 的采样频率下对每幅图像产生的脑电信号采集 500 个采样点。然而,由于许多因素的影响(如硬件系统、软件过程等),从模拟设备采集的数据呈现出不同的大小。为确保序

列长度固定,对长度小于 500 的样本在后面补 0,并裁剪那些大于 500 的序列。在实验中,每一个 EEG 序列的前 40 个样本(0~40 ms)被丢弃,以最小化上一个刺激图像的干扰。经过以上处理后,总计使用了 460 个样本(40~500 ms)。

## 2.2 使用 StackCNN-B 学习 EEG 视觉特征

为了验证本文改进 CNN 算法的泛化能力,以及优化手段的效果,本文分别验证了各方法在视觉刺激数据集上的准确率、F1 值和召回率,具体实验结果如表 4 所示。在实验过程中,具体参数设置如表 3 所示,利用反向传播(back propagation, BP)算法<sup>[24]</sup>对每个模型进行训练,然后使用小批量梯度下降算法计算梯度。首先,将数据集按 8:2 的比例随机分为训练集和测试集,其中训练集有 1 600 幅图像,测试集有 400 幅图像;其次,确保所有 6 名参与者对同一幅图像记录的脑电图被分割成相同的部分;最后,对训练集进行 10 次随机抽样。各个模型之间的对比情况如表 4 所示。

表 4 最佳参数模型的结果

模型	准确率	召回率	F1 值
TextCNN	91.38	90.92	91.28
TextCNN with 1-max	92.16	91.03	90.87
TextCNN with K-max	94.48	93.96	94.42
TextCNN-B with K-max	98.41	97.84	98.41
StackCNN	97.62	97.14	97.59
StackCNN with 1-max	97.83	96.59	97.64
StackCNN with K-max	98.29	97.98	98.23
StackCNN-B with K-max	99.78	99.02	99.29

由表 4 可知,StackCNN 网络加入 K-max 池化和 Bagging 算法后,在准确率和 F1 值上都优于其他模型。该结果证明 K-max 池化和 Bagging 算法对 CNN 网络有很好的提升作用。并且 StackCNN 在视觉分类问题中明显优于 TextCNN,这说明 StackCNN 不仅能捕捉低维的特征,而且随着过滤器感受野的变大,还能捕捉更高维的特征。

为进一步验证模型的有效性,本文使用视觉刺激数据集再次进行实验,与现有其他 EEG 特征提取方法对比,实验结果如表 5 示。

El-lone 等<sup>[4]</sup>利用经典的支持向量机分类器对脑电信号进行生物和非生物的二分类,分类准确率达到 82.70%。Parekh<sup>[5]</sup>采用基于 CNN 的带离群点去除的 EEG 网络结构对两类 EEG 信号进行分类,得到了 88.00% 的准确率。Zheng 等<sup>[8]</sup>采用了基于注意力的 Bi-LSTM 方法对 40 类 EEG 信号进行分类,获得了 99.50% 的准确率。此外,本研究还使用

表 5 StackCNN-B 方法与其他方法的比较

Table 5 Comparison between StackCNN-B method and other methods

方法	类别	训练时间/min	准确率/%
SVM <sup>[4]</sup>	2	8	82.70
EEG-Net <sup>[5]</sup>	2	17	88.00
BiLSTM+LCA+SVM <sup>[6]</sup>	40	34	84.00
LSTMS-B <sup>[7]</sup>	40	25	97.13
Bi-LSTM-AttGW <sup>[8]</sup>	40	41	99.50
FC-GDN <sup>[10]</sup>	40	40	98.40
StackCNN-B	40	20	99.78

了与文献[6-7,10]相同的脑电图数据集以及类似的流程进行实验,在分类准确率和分类类别上,本文方法 StackCNN-B 取得了更好的效果。

由表 5 可知,本文提出的模型在对于 EEG 信号分类的准确率和分类数目上明显优于其他模型。分析表 5 可知,深度学习的算法明显优于机器学习,因为本实验所用的数据集十分庞大;LSTM 与 Bagging 结合的算法优于 BiLSTM 算法,这说明 Bagging 算法具有较好的泛化能力,这也是本模型使用 Bagging 算法的原因。本实验提出的改进的 StackCNN 模型优于 LSTM 模型,虽然两种模型均适合应用于时序序列,但 StackCNN 使用多个较小的过滤器增加网络深度,不仅可以捕捉到高维的特征,而且实验所需的参数量远小于 LSTM 模型,因此训练时间更短。结果表明,本文所提出的模型在各方面均表现优异。

## 2.3 基于 ResNet 的特征回归与视觉自动分类

基于 ResNet 的回归分析旨在将视觉图像映射到学习的 EEG 特征向量上。本文使用 CNN 提取脑电特征,然后对大脑视觉活动进行分类。对于每个图像,可以从每个 CNN 模型获得一个 EEG 特征向量。将基于 ResNet 的回归作为一个图像的输入,并希望它的输出能够尽可能类似于 CNN 的输出。

为了验证该回归的性能,本实验使用与激发 EEG 信号相同的图像集,并按照上文所提到的分层抽样法将图像数据集分为训练集和测试集,本文采用均方误差 MSE 评估 4 种不同的回归方法的性能,如表 6 所示。根据表 6 的数据可以观察到 K 邻近回归的均方误差最小,于是将 K 邻近作为本文的回归模型,用于作为图像的特征提取器,并映射到 EEG 特征向量。

为了验证本文提出的 StackCNN-B 在图像分类上的应用价值,使用 ResNet 回归对提取到的脑电信号相关的视觉特征表示和原始图像数据集进行训练,再将基于 ResNet 回归的输出作为特征对图像进

行分类,如表7所示。将 ResNet 回归应用到其他模型上,相对比之下,基于 StackCNN-B 模型的图像分类平均准确率更高,表明了集成深度学习方法在自动视觉分类中具有良好的性能和良好的应用价值。再次证明本模型能更好地学习到 EEG 的特征表示,从而使得回归器能够更好地利用 EEG 的数据进行图像分类。

表6 4种回归模型的MSE对比

Table 6 MSE comparison of four regression models

回归模型	MSE
随机森林	1.49
决策树	1.76
K邻近	0.54
支持向量机	1.68

表7 基于 ResNet 回归不同模型关于图像分类性能比较

Table 7 Performance comparison of different models for image classification based on ResNet regression

方法	准确率/%
BiLSTM+LCA+SVM <sup>[6]</sup>	89.70
LSTMS-B <sup>[7]</sup>	90.16
Bi-LSTM-AttGW <sup>[8]</sup>	93.48
FC-GDN <sup>[10]</sup>	92.18
StackCNN-B	96.45

### 3 结论

本文设计了一种改进的 StackCNN 网络和集成学习相结合的方法,用于学习脑电信号相关的视觉特征表示,采用深度残差网络对提取到的脑电信号相关的视觉特征表示和原始图像数据集进行训练,实现视觉图像分类。使用分层抽样法对训练集和测试集划分,将改进的 StackCNN 网络与 Bagging 算法结合提取脑电信号视觉特征,在此基础上训练基于 ResNet-18 的回归实现对原始图像的视觉分类,在 ImageNet 公开数据集上测试的结果表明,本文方法分类准确率均优于现有研究方法。证明本文提出的 StackCNN-B 模型在视觉分类中具有良好的性能和巨大的应用价值。

### 参考文献:

[1] DAS K, GIESBRECHT B, ECKSTEIN M P. Predicting variations of perceptual performance across individuals from neural activity using pattern classifiers[J]. NeuroImage, 2010, 51(4): 1425-1437.

[2] CARLSON T, TOVAR D A, ALINK A, et al. Representational dynamics of object vision: the first 1000 ms[J]. Journal of Vision, 2013, 13(10): 1.

[3] KANESHIRO B, PERREAU GUIMARAES M, KIM H S, et al. A representational similarity analysis of the dynamics of object processing using single-trial EEG classification[J]. PLoS One, 2015, 10(8): e0135697.

[4] EL-LONE R, HASSAN M, KABBARA A, et al. Visual objects categorization using dense EEG: a preliminary study[C]//2015 International Conference on Advances in Biomedical Engineering (ICABME). Piscataway: IEEE, 2015: 115-118.

[5] PAREKH V, SUBRAMANIAN R, ROY D, et al. An EEG-based image annotation system [M]//Communications in Computer and Information Science. Berlin: Springer, 2018: 303-313.

[6] SPAMPINATO C, PALAZZO S, KAVASIDIS I, et al. Deep learning human mind for automated visual classification[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2017: 4503-4511.

[7] ZHENG X, CHEN W Z, YOU Y, et al. Ensemble deep learning for automated visual classification using EEG signals[J]. Pattern Recognition, 2020, 102: 107147.

[8] ZHENG X, CHEN W Z. An attention-based Bi-LSTM method for visual object classification via EEG[J]. Biomedical Signal Processing and Control, 2021, 63: 102174.

[9] MISHRA A, RAJ N, BAJWA G. EEG-based image feature extraction for visual classification using deep learning [C]//2022 International Conference on Intelligent Data Science Technologies and Applications (IDSTA). Piscataway: IEEE, 2022: 181-188.

[10] KHALEGHI N, REZAI T Y, BEHESHTI S, et al. Developing an efficient functional connectivity-based geometric deep network for automatic EEG-based visual decoding [J]. Biomedical Signal Processing and Control, 2023, 80: 104221.

[11] SHU B, REN F J, BAO Y W. Investigating LSTM with k-max pooling for text classification[C]//2018 11th International Conference on Intelligent Computation Technology and Automation (ICICTA). Piscataway: IEEE, 2018: 31-34.

[12] KIM Y. Convolutional neural networks for sentence classification[C]//Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). Stroudsburg: ACL, 2014: 1746-1751.

[13] GUO B, ZHANG C X, LIU J M, et al. Improving text classification with weighted word embeddings via a multi-channel TextCNN model [J]. Neurocomputing, 2019, 363: 366-374.

[14] 周澳回, 翁知远, 周思源, 等. 一种基于主题过滤和语义匹配的服务发现方法[J]. 郑州大学学报(工学

- 版), 2022, 43(6): 36-41, 56.
- ZHOU A H, WENG Z Y, ZHOU S Y, et al. A service discovery method based on topic filtering and semantic matching[J]. Journal of Zhengzhou University (Engineering Science), 2022, 43(6): 36-41, 56.
- [15] MONTANARO A, EBISUZAKI T, BERTAINA M. Stack-CNN algorithm: a new approach for the detection of space objects[J]. Journal of Space Safety Engineering, 2022, 9(1): 72-82.
- [16] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [EB/OL]. (2014-09-04) [2023-07-15]. <http://doi.org/10.48550/arXiv.1409.1556>.
- [17] BAUER E, KOHAVI R. An empirical comparison of voting classification algorithms: Bagging, boosting, and variants[J]. Machine Learning, 1999, 36(1): 105-139.
- [18] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2016: 770-778.
- [19] WU Z F, SHEN C H, VAN DEN HENGEL A. Wider or deeper: revisiting the ResNet model for visual recognition [J]. Pattern Recognition, 2019, 90: 119-133.
- [20] DUBEY V K, SARKAR S, SHUKLA R, et al. Epileptic seizure stage classification from EEG signal using ResNet18 model and data augmentation [C] // 2022 IEEE Delhi Section Conference (DELCON). Piscataway: IEEE, 2022: 1-5.
- [21] 杨剑锋, 乔佩蕊, 李永梅, 等. 机器学习分类问题及算法研究综述[J]. 统计与决策, 2019, 35(6): 36-40.
- YANG J F, QIAO P R, LI Y M, et al. A review of machine-learning classification and algorithms[J]. Statistics & Decision, 2019, 35(6): 36-40.
- [22] SHAH K, PATEL H, SANGHVI D, et al. A comparative analysis of logistic regression, random forest and KNN models for the text classification[J]. Augmented Human Research, 2020, 5(1): 12.
- [23] RUSSAKOVSKY O, DENG J, SU H, et al. ImageNet large scale visual recognition challenge[J]. International Journal of Computer Vision, 2015, 115(3): 211-252.
- [24] 郝旺身, 陈耀, 孙浩, 等. 基于全矢-CNN的轴承故障诊断研究[J]. 郑州大学学报(工学版), 2020, 41(5): 92-96.
- HAO W S, CHEN Y, SUN H, et al. Bearing fault diagnosis based on full vector-CNN[J]. Journal of Zhengzhou University (Engineering Science), 2020, 41(5): 92-96.

## EEG Visual Classification Algorithm Based on Improved StackCNN Network and Ensemble Learning

YANG Qing<sup>1,2,3</sup>, WANG Yaqu<sup>1,2,3</sup>, WEN Dou<sup>1,2,3</sup>, WANG Ying<sup>1,2,3</sup>, WANG Xiangyu<sup>1,2,3</sup>

(1. Hubei Provincial Key Laboratory of Artificial Intelligence and Smart Learning, Central China Normal University, Wuhan 430079, China; 2. School of Computer, Central China Normal University, Wuhan 430079, China; 3. National Language Resources Monitoring & Research Center for Network Media, Central China Normal University, Wuhan 430079, China)

**Abstract:** Aiming at the limited studies researches on visual classification directly using image-induced EEG signals and low average accuracy of visual classification, a method combining convolutional neural networks (CNN) and ensemble learning was designed to learn the visual feature representation related to EEG signals. By adding the  $K$ -max pooling method to the stackCNN network to solve the problem of information loss when extracting EEG features, and combining with Bagging algorithm to enhance the generalization ability of the network, this method was called StackCNN-B. In order to verify the performance of StackCNN-B method in image classification, images were classified using deep residual network regression. The results of ablation experiments and comparative experiments with existing studies showed that the recognition accuracy of this method was high. The average accuracy in learning the visual feature representation of EEG signals was 99.78%, and the average accuracy in image classification was 96.45%. Compared with the most advanced Bi-LSTM-AttGW method, the average accuracy was improved by 0.28 percentage point and 2.97 percentage point. The results verified that EEG signals could effectively decode human brain activities related to visual recognition, proved the advantages of the proposed StackCNN-B model.

**Keywords:** electroencephogram; vision classification; convolutional neural network; Bagging algorithm; ResNet network