ISSN 1004-9037,CODEN SCYCE4
Journal of Data Acquisition and Processing Vol. 41,No. 1,Jan. 2026,pp. 244－258
DOI:10. 16337/j. 1004-9037. 2026. 01. 017

http://sjcj. nuaa. edu. cn
E-mail:sjcj@nuaa. edu. cn
Tel/Fax:＋86-025-84892742

基于双残差LSTM和DCGAN的脑电信号驱动视觉图像重建模型

倪哲文，全海燕

(昆明理工大学信息工程与自动化学院,昆明 650500)

摘　要: 近年来,计算机视觉的进步使基于脑电信息重建图像成为可能,这在医学图像重建和脑机接口等领域具有重要意义。然而,由于脑电信号的复杂性和时序特性,现有模型在特征提取和图像生成任务上面临诸多挑战。为此,本文提出了一种基于双残差长短期记忆网络(Long short-term memory,LSTM)和深度卷积生成对抗网络(Deep convolution generative adversarial network,DCGAN)的脑电信号驱动视觉图像重建模型。该模型引入基于注意力残差网络和三元组损失函数的长短期记忆网络(LSTM network based on attention residual network and triplet loss,ARTLNet),以提升脑电信号特征提取的质量。ARTLNet融合了残差网络、长短期记忆网络和注意力机制,通过残差连接改善深层网络训练,长短期记忆网络捕捉时间序列特征,注意力机制增强对关键特征的关注;同时结合批量归一化和全局平均池化,确保信号稳定传递。在图像生成阶段,模型引入自行设计的DCGAN与特征融合策略,有效提升了生成图像的质量和多样性。实验结果表明,改进后的ARTLNet在Characters和Objects数据集上,结合不同的分类和聚类算法均获得了更高的准确率;所提模型在图像生成质量方面也表现优越,尤其在图像清晰度和类别区分度方面展现出显著优势。

关键词: 深度学习;生成对抗网络;脑电信号解码;自注意力;残差连接

中图分类号: TP391.41;TP181　　**文献标志码:** A

引用格式: 倪哲文,全海燕.基于双残差LSTM和DCGAN的脑电信号驱动视觉图像重建模型[J].数据采集与处理,2026,41(1):244-258. NI Zhewen,QUAN Haiyan. EEG signal-driven visual image reconstruction model based on double residual LSTM and DCGAN[J].Journal of Data Acquisition and Processing,2026,41(1):244-258.

引　　言

通过大脑思维信号与外界交互始终是人类的前沿目标,这个目标推动了脑机接口(Brain-computer interface, BCI)等技术的发展。随着科技的进步,BCI技术在医疗康复和神经科学等领域得到了广泛应用,为神经康复提供了重要的工具[1]。BCI技术通过解码和分析大脑信号,使人们能够使用脑电控制外部设备,同时解码大脑对想象或视觉刺激的反应,极大提升了残疾人的日常生活质量。脑电图(Electro-encephalography, EEG)信号因其高时间分辨率、便携性和低成本等优点,在BCI领域得到了广泛应用。

如何从脑电信号中学习语义特征至关重要。近年来,研究者们在特征提取任务上进行了广泛探索。传统机器学习方法,如随机森林(Random forest,RF)和极端梯度提升(Extreme gradient boosting,XGBoost)[2],已成功应用于EEG解码任务。然而,这些方法在特征提取上依赖手工设计,导致模型的泛化能力和解码精度受到限制。

基金项目: 国家自然科学基金(61861023)。
收稿日期: 2024-12-19;**修订日期:** 2025-05-08

随着深度学习的发展,研究者们提出了多种脑电解码的新方法。例如,Lawhern等[3]提出的EEG-Net为运动想象EEG(Motor imagery EEG, MI-EEG)特征解码提供了一个有效的框架;Ingolfsson等[4]在此基础上引入时序卷积网络(Temporal convolutional network, TCN)[5],进一步提升了特征解码的准确性;徐晴等[6]将癫痫脑电的一维时间序列数据转为二维图像,利用EfficientNetV2模型实现癫痫检测。此外,Kavasidis等[7]首次将长短期记忆网络(Long short-term memory, LSTM)[8]与生成技术相结合,用于生成ImageNet中的已知图像。然而,这些模型在利用时序数据的局部和全局信息的充分性上仍存在不足,它们往往面临梯度消失与特征分离需求之间的平衡问题,导致特征提取不充分,进而影响模型的训练稳定性。此外,这些模型过于依赖监督学习,并不适合转移到图像生成任务,导致图像生成质量较低。

基于这种背景,神经-图像编解码技术近年来取得了显著进展,并为脑电信号的应用开辟了新的思路。Ren等[9]提出了一种多源域自适应成对迁移学习模型,通过对脑电信号进行空间-频率编码,并使用成对学习对比相同类别样本完成脑电情绪识别任务。此外,Song等[10]提出了EEG Conformer模型,结合卷积神经网络和Transformer结构,高效地融合时空特征,提升了脑电信号处理的精度和鲁棒性。Qian等[11]提出了EEG-Diffusion框架,利用扩散模型(Diffusion model)从脑电信号重建视觉图像,显著提升了脑电解码的视觉保真度。

另外,图像生成质量一直是脑电重建领域的一个关键问题。Kavasidis等[7]使用生成对抗网络(Generative adversarial network, GAN)和变分自编码器(Variational auto-encoder, VAE)[12]从脑电信号中合成图像,验证了GAN的方法在合成图像的真实性方面优于VAE。然而,传统GAN生成器仅以噪声作为输入,无法控制生成特定类别的图像,因此在生成图像时可能出现类别混淆和模糊不清等问题。为了解决这一问题,Khare等[13]提出使用条件渐进生长GAN(cProGAN)学习提取的特征向量和相应图像之间的映射;Mishra等[14]提出使用感知损失和注意力模块来生成高质量的图像;Zheng等[15]提出频谱归一化GAN(SNGAN),通过引入频谱归一化来提升训练稳定性和生成图像的质量。

在上述研究基础上,本文提出了一种基于注意力残差网络和三元组损失函数的长短期记忆网络模型(Attention residual network and triplet loss function based LSTM network,ARTLNet),结合双层残差型LSTM网络,使用度量学习的方法进行脑电特征提取,同时结合深度卷积生成对抗网络(Deep convolution GAN, DCGAN),根据编码后的特征生成清晰图像。该方法不仅有效提高了特征提取的准确性,还能提取出更具判别性和泛化性的特征,更符合图像生成任务。此外,通过将特征与噪声融合并输入到生成器,利用条件嵌入控制生成图像的类别,能够生成指定类别的清晰图像。结合谱归一化和数据增强策略,可进一步增强模型的生成能力,有效应对生成图像的模糊性和不一致性问题,从而提升生成图像的质量与多样性。

1 本文方法

本节首先概括所提算法模型的整体架构,然后分别对所提架构的两个阶段的模型(特征提取ARTLNet模型和图像生成DCGAN模型)进行详细的描述。

1.1 整体架构

本文所提模型的整体结构如图1所示,包含两个阶段:第1阶段采用特征提取网络ARTLNet,通过双层残差型LSTM对脑电信号进行时序特征建模,并结合三元组损失函数优化特征表达,所得特征以检查点形式保存,供后续图像生成使用;第2阶段引入DCGAN模型,将第1阶段提取的EEG特征与随机噪声拼接后作为输入,进行图像生成。考虑到特征提取与图像生成任务在优化目标上的差异,为避免生成器过度依赖既有特征分布,DCGAN模型不采用预训练方式,而是从头训练,以更有效地适应

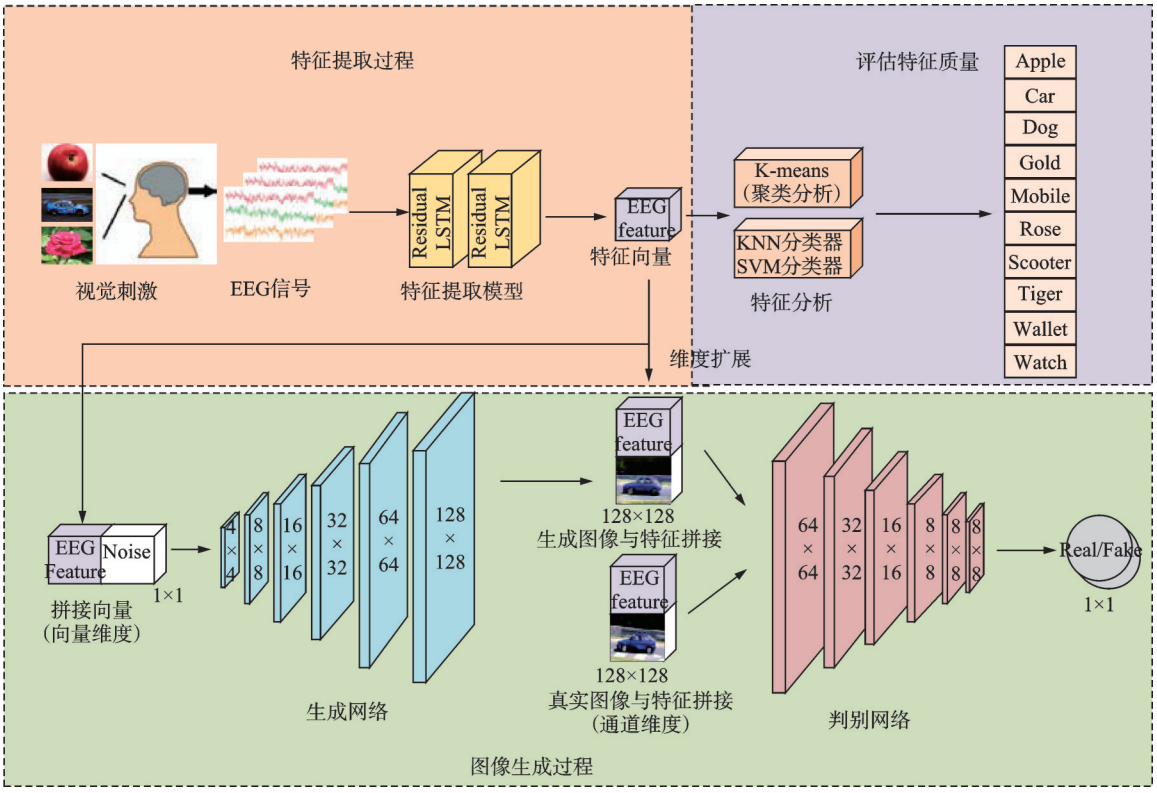


图1 本文算法模型框架

Fig.1 Framework of the proposed algorithm model

EEG图像生成任务的特性。

生成器将EEG特征与随机噪声在向量维度上拼接后输入网络,生成对应的EEG图像;判别器则通过将扩展后的EEG特征与输入图像在通道维度拼接,判断图像的真实性。在两个阶段中,特征提取模型和图像生成模型分别在不同的训练循环中独立优化,确保训练过程的有效性和合理性。

为了验证EEG特征的有效性,特征提取阶段的输出被用于分类和聚类实验,评估提取特征在已知标签数据上的分类能力,以及在无监督条件下的分布一致性。这些验证方法独立于GAN训练过程,旨在全面评估特征质量。

1.2 特征提取模型

本文的特征提取模型采用自主设计的ARTLNet网络模型,其框架如图2所示。ARTLNet的核心功能是对输入数据进行编码,并结合TripletSemiHardLoss[16]进行度量学习,优化特征之间的距离,增强不同类别特征的区分能力。

为提升脑电信号这种复杂时序数据的表征能力,本文所提的ARTLNet模型在设计上融合了残差结构、注意力机制以及长短时记忆网络(LSTM),其动机源于3个方面的互补特性。首先,残差结构能够有效缓解深层网络中的梯度消失问题,同时促进特征的多层次抽象表达,从而提升模型的表达能力;其次,注意力机制可动态分配特征维度的重要性权重,使模型聚焦于更具判别性的特征区域,有助于增强类间区分性;最后,LSTM作为处理时序数据的经典结构,能够捕捉输入信号中的长距离依赖关系,提升对时间动态特征的建模能力。在此基础上,结合TripletSemiHardLoss进行度量学习,以应对脑电特征类别边界模糊、类间重叠的问题,通过构建判别性嵌入空间提升类间可分性与类内一致性。整体

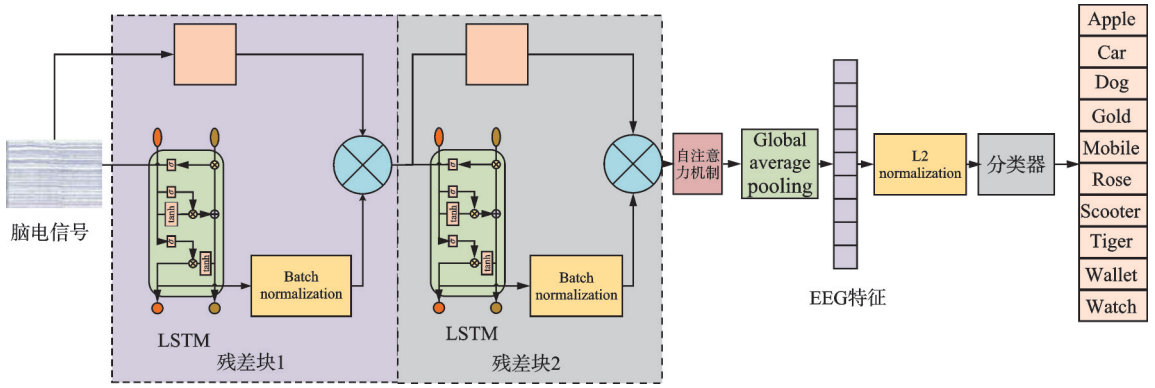


图2 ARTLNet框架

Fig.2 ARTLNet framework

架构设计立足于脑电信号的本质特性,旨在实现更鲁棒的特征表示与性能提升。

首先,模型利用两个残差块对EEG信号进行处理,每个残差块包括1个LSTM层、批量归一化层和1个Dense全连接层。LSTM层能够捕捉EEG信号中的时间序列特征,而Dense层则确保LSTM输出的维度与输入维度保持一致。残差连接通过将输入直接加到输出上,帮助模型在训练中减轻梯度消失问题。然后通过自注意力机制提高模型对关键特征的敏感度和提取能力,后续的平均池化用于降低特征维度,减少模型的复杂性,同时保留全局信息,从而降低过拟合风险。最后L2归一化层进一步优化了特征的分布,使其在欧氏空间中更具辨识度。

1.2.1 深度残差型LSTM

脑电信号具有丰富的时间、空间和频率特征。研究表明,增加神经网络层数可以提升脑电特征提取能力,但这往往带来识别率下降、计算开销上升及梯度不稳定等问题。为此,本文采用双层LSTM结构,在保证性能的同时有效控制计算资源消耗。该结构能够捕捉复杂的时间序列特征,同时避免深层网络带来的训练难题。

此外,本文使用TripletSemiHardLoss进行嵌入学习,优化特征间距离关系。该方法对网络深度要求较低,双层LSTM已能提供充分表达能力,进一步加深模型并不能显著改善性能,反而可能增加训练复杂度。

为提升特征提取效果与稳定性,模型引入残差连接,通过跳跃结构保留关键信息并缓解梯度消失问题。结合LSTM的动态建模能力与残差结构的信息保留优势,模型在捕捉EEG中的时序依赖关系方面表现更优。图3(a)展示了基本残差结构,图3(b)展示了包含批归一化(Batch normalization, BN)的残差型LSTM架构。

假设残差架构输入为$\boldsymbol{i}_{\mathrm{R}}$,输出为$O(\boldsymbol{i}_{\mathrm{R}})$,残差映射模型为$R(\boldsymbol{i}_{\mathrm{R}})$,输出可得$O(\boldsymbol{i}_{\mathrm{R}})=R(\boldsymbol{i}_{\mathrm{R}})+\boldsymbol{i}_{\mathrm{R}}$,则对应训练目标为$R(\boldsymbol{i}_{\mathrm{R}})=O(\boldsymbol{i}_{\mathrm{R}})-\boldsymbol{i}_{\mathrm{R}}$,$\sigma$和tanh分别表示Sigmoid激活函数和tanh激活函数。

LSTM层间的计算公式为

$$\boldsymbol{i}_t=\sigma W_{hi}\boldsymbol{h}_{t-1}+W_{xi}\boldsymbol{x}_t+\boldsymbol{b}_{\mathrm{i}} \tag{1}$$

$$\boldsymbol{f}_t=\sigma(W_{hf}\boldsymbol{h}_{t-1}+W_{xf}\boldsymbol{x}_t+\boldsymbol{b}_{\mathrm{f}}) \tag{2}$$

$$\boldsymbol{o}_t=\sigma(W_{ho}\boldsymbol{h}_{t-1}+W_{xo}\boldsymbol{x}_t+\boldsymbol{b}_{\mathrm{o}}) \tag{3}$$

$$\tilde{\boldsymbol{c}}_t=\tanh(W_{hc}\boldsymbol{h}_{t-1}+W_{xc}\boldsymbol{x}_t+\boldsymbol{b}_{\mathrm{c}}) \tag{4}$$

$$\boldsymbol{c}_t=\boldsymbol{f}_t\otimes\boldsymbol{c}_{t-1}+\boldsymbol{i}_t\otimes\tilde{\boldsymbol{c}}_t \tag{5}$$

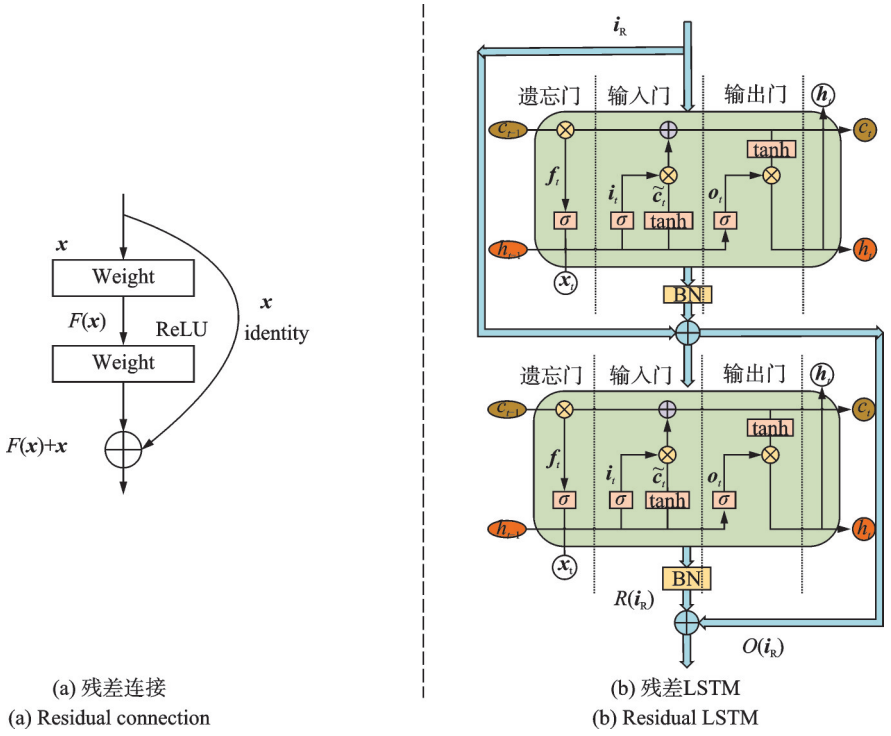


图3 残差连接和残差LSTM示意图

Fig. 3 Schematics of residual connection residual LSTM

$$\boldsymbol{h}_t = \boldsymbol{o}_t \otimes \tanh(\boldsymbol{c}_t) \tag{6}$$

式中：下标i、o和f分别表示输入门、输出层和遗忘门；$\boldsymbol{c}$、$\tilde{\boldsymbol{c}}$和$\boldsymbol{h}$依次对应单元的记忆状态、候选值和隐藏状态，$\boldsymbol{W}$为各状态对应的权重矩阵，$\boldsymbol{b}$是偏置项，而t和$t-1$则指代当前时刻和前一时刻的状态信息。

1.2.2 自注意力机制

在所有残差块处理后，模型引入自注意力机制以进一步提升特征提取效果。残差块已经成功提取了丰富的时间序列特征，并增强了特征表达的紧凑性。此时加入注意力机制，能够帮助模型聚焦关键时间步或特征，从而提升区分能力和全局依赖的建模精度，有效减少噪声和冗余信息的干扰，提高关键特征的提取效率。同时，残差块的输出维度保持一致，使得注意力机制的计算更加简洁，从而降低了计算成本。

自注意力机制基于“查询-键-值”模型，能够捕捉序列中任意位置的依赖关系，特别适合处理时序数据。每个输入生成查询、键和值向量，通过计算查询与键的相似度来加权值向量，最终生成输出。其计算公式为

$$y = \mathrm{Attention}(\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V}) = \mathrm{softmax}\left(\frac{\boldsymbol{Q}\boldsymbol{K}^{\mathrm{T}}}{\sqrt{d_k}}\right)\boldsymbol{V} \tag{7}$$

式中d_k为键的维度。首先通过查询矩阵$\boldsymbol{Q}$和键矩阵$\boldsymbol{K}$的点积计算得出相似度，点积的结果会被除以$\sqrt{d_k}$，以防止随着维度增加，点积值过大而导致梯度消失。然后通过softmax函数将上述相似度转换为注意力权重，这一步的目的是将相似度归一化，确保权重之和为1。最后使用注意力权重$A=$

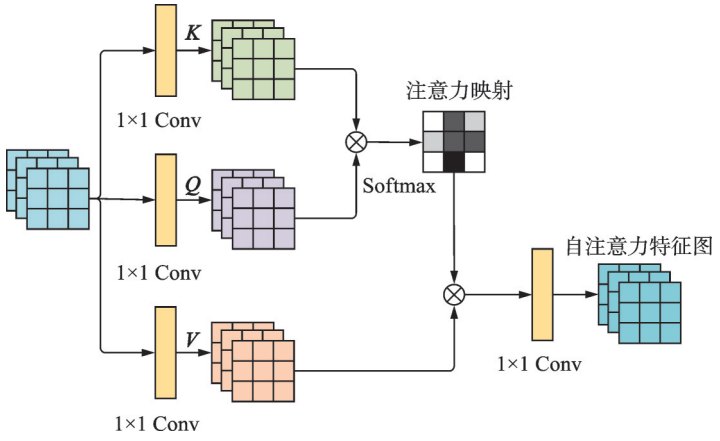


图4 自注意力机制结构图

Fig.4 Structural diagram of the self-attention mechanism

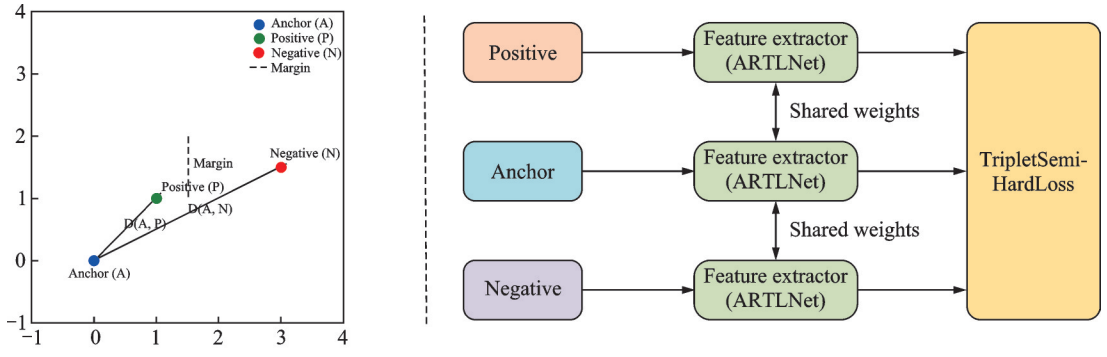
$\mathrm{softmax}\left(\frac{\boldsymbol{QK}^{\mathrm{T}}}{\sqrt{d_k}}\right)$对值矩阵$\boldsymbol{V}$进行加权求和,生成最终的输出表示。

图4展示了该机制的结构,包括输入特征、查询矩阵$\boldsymbol{Q}$、键矩阵$\boldsymbol{K}$、值矩阵$\boldsymbol{V}$和输出特征图。

1.2.3 半难样本三元组损失函数(TripletSemiHardLoss)

在ARTLNet模型中,本文采用了TripletSemiHardLoss,这是一种改进的Triplet Loss。传统的Triplet Loss由锚点样本(Anchor)、正样本(Positive)和负样本(Negative)构成,目标是使锚点与正样本距离更近,而与负样本距离更远。TripletSemiHardLoss则选择那些距离锚点比正样本远但距离还不够大的"半难"负样本,以提高训练的稳定性和效率,避免对极端困难样本的过度关注。如图5(a)所示,蓝、绿和红色点分别表示锚点(A)、正样本(P)和负样本(N),线段代表它们之间的距离,虚线表示正负样本之间的分离边界(Margin)。

如图5(b)所示,特征提取过程使用度量学习模型,以适应小样本学习环境。通过Anchor、Positive和Negative三种样本,模型学习特征空间中样本的距离关系,目标是使同类样本(Anchor与Positive)距离更近,不同类样本(Anchor与Negative)距离更远,从而提升相似性特征的学习效果。ARTLNet模型



(a) TripletSemiHardLoss损失
(a) TripletSemiHardLoss loss

(b) 度量学习网络结构
(b) Metric learning network structure

图5 TripletSemiHardLoss损失和度量学习网络结构示意图

Fig.5 Schematics of TripletSemiHardLoss loss and metric learning network structure

的3个部分同时训练，权重共享，以增强模型的判别能力，在小样本场景中能提高特征提取的稳定性和对相似性度量的敏感性。这种方法有效提高了不同类别特征的区分度，使生成器能基于明确的类别特征生成高质量、多样性的图像。其计算公式为

$$L(\boldsymbol{x}_a,\boldsymbol{x}_p,\boldsymbol{x}_n)=\max\left(0,\left\|f(\boldsymbol{x}_a)-f(\boldsymbol{x}_p)\right\|_2^2-\left\|f(\boldsymbol{x}_a)-f(\boldsymbol{x}_n)\right\|_2^2+\alpha\right) \tag{8}$$

式中：$f(\boldsymbol{x})$为特征提取网络的输出，即样本$\boldsymbol{x}$的特征向量；α表示边际，希望锚点样本与正样本之间的距离$\left\|f(\boldsymbol{x}_a)-f(\boldsymbol{x}_p)\right\|_2^2$尽可能小，锚点样本与负样本之间的距离$\left\|f(\boldsymbol{x}_a)-f(\boldsymbol{x}_n)\right\|_2^2$尽可能大。

1.3 图像生成模型

1.3.1 生成器判别器网络结构

图像生成模型是一个基于DCGAN架构的深度卷积神经网络，任务是从随机噪声和EEG特征中生成128×128的高质量图像。如图6所示，生成器包含多层反卷积层，用于逐步上采样输入特征，最终生成目标尺寸的图像。判别器接收图像输入，通过多层卷积逐步下采样，并最终输出一个标量作为真假判断的标准。

生成器首先接收随机噪声向量和EEG特征拼接后的1×1×256组合向量。通过6层反卷积，每一层在增大空间维度的同时减少通道数，最后通过卷积层生成128×128×3的RGB图像。反卷积层后接

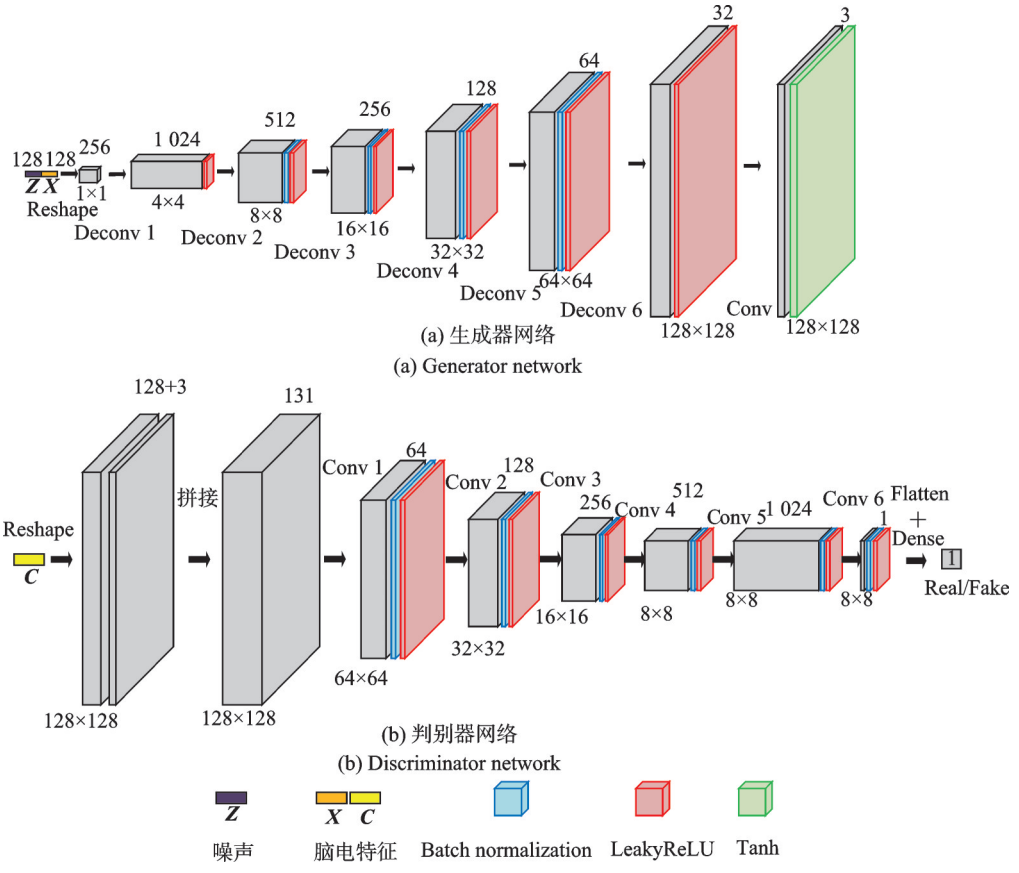


图6 图像生成模型的网络结构

Fig.6 Network structure of the image generation model

LeakyReLU激活函数和批归一化层,输出图像经过tanh激活,使像素值在[−1, 1]之间。

判别器则通过卷积网络对生成图像和真实图像进行区分。它接收拼接后的EEG特征与图像(128×128×128和128×128×3),通过6层卷积将通道数翻倍,最后通过平展层和全连接层进行二分类输出,判定图像的真伪。每层卷积后都有LeakyReLU激活函数和批归一化层。

1.3.2 损失函数

本文使用的DCGAN模型中,生成器和判别器所使用的损失函数为hinge loss,并在生成器损失中加入了模式损失,通过衡量生成器在给定不同噪声输入时生成的图像之间的差异,确保生成器输出具有多样性,能减少模式崩溃(Mode collapse)的风险。在这个模型中,生成器试图使判别器对假图像的评分趋向于正值。生成器和判别器的的损失函数可以分别表示为

$$L_G=-E_{\boldsymbol{x}\sim p_{z(z)},\psi\sim p_{\text{data}}(\boldsymbol{x})}\left[D\left(T(G(\boldsymbol{z},\psi)),\psi\right)\right]+\lambda\times L_{\text{M}} \tag{9}$$

$$L_{\text{M}}=\frac{E\left[\left\|G(\boldsymbol{z}_1)-G(\boldsymbol{z}_2)\right\|_1\right]}{E\left[\left\|\boldsymbol{z}_1-\boldsymbol{z}_2\right\|_1\right]} \tag{10}$$

$$L_D=E_{(\boldsymbol{x},\psi)\sim p_{\text{data}}(\boldsymbol{x})}\left[\max(0,1-D(T(\boldsymbol{x}),\psi))\right]+E_{\boldsymbol{x}\sim p_{z(z)},\psi\sim p_{\text{data}}(x)}\left[\max\left(0,1+D\left(T(G(\boldsymbol{z},\psi)),\psi\right)\right)\right] \tag{11}$$

式中:L_{G}为生成器损失;L_{D}为判别器损失;L_{M}为模式损失;$G(\boldsymbol{z}_1)$和$G(\boldsymbol{z}_2)$分别为不同的噪声$\boldsymbol{z}_1$和$\boldsymbol{z}_2$生成的假图像;λ为模式损失的权重系数;T表示数据增强策略。

2 实验分析

2.1 数据集

本研究使用的EEG数据由Kumar等[17]收集,数据来源于23名年龄在15至40岁的参与者。参与者观看字符(Characters)、数字(Digits)和物体(Objects)三类刺激,刺激图像取自Chars74K[18]、MNIST和ImageNet数据集[19]。每类包含10个例子,每个例子记录了23名参与者的10 sEEG信号。数据使用Emotiv EPOC+传感器采集,共14个通道,采样率为128 Hz。

为提高实验效率并突出模型在不同认知复杂度下的表现,本研究仅选用字符与物体两个子数据集。字符与数字在视觉形式和脑电反应上高度相似,且整体复杂度较低,为提升实验效率,本研究选择字符数据集使用。相比之下,物体数据集具有更高的复杂度,字符与物体在认知复杂度上的差异,有助于从不同层面评估模型的表征能力与泛化性能。

2.2 实验设置

模型训练分两个阶段:第一阶段在字符和物体数据集上训练,提取EEG特征,参数设置包括Batch size=256,学习率=0.000 3,Adam优化器,300次迭代;第二阶段基于提取的EEG特征生成图像,主要使用物体数据集,Batch size=128,学习率和优化器与第一阶段相同,迭代300次。实验在Windows 10系统、Python平台和TensorFlow框架下使用RTX 4090显卡进行。

2.3 评价指标

第一阶段实验,为了避免单一算法可能带来的结果不稳定性[20],确保评估结果的可靠性,使用3种评价指标来评价提取特征的质量。提取的特征被用于分类和聚类实验。分类实验采用K近邻法(K-nearest neighbors, KNN)[21]和支持向量机法(Support vector machine, SVM)[22],评估提取特征在已知标签数据上的分类能力;聚类实验则采用K均值聚类法(K-means clustering, K-means),衡量特征在

无监督条件下的分布一致性。这3种方法得到的准确率越高,表明特征提取效果越好。其中,KNN和SVM作为传统算法,计算速度更快、训练过程简单,尤其适合小规模数据集。

第二阶段实验,评估生成图像的质量,同样采用3种指标:IS(Inception score)[23]、FID(Frechet inception distance)[24]和CDS(Class diversity score)[25]。IS通过比较生成图像的类别分布与整体类别分布的差异来评估图像的质量与多样性。分数越高,表示质量与多样性越高,其计算公式为

$$\mathrm{IS}(G)=\exp\left(E_{G(z)\sim p_g}\left[D_{\mathrm{KL}}\left(p(y|G(z))\|\,p(y)\right)\right]\right) \tag{12}$$

式中:$G(z)$表示生成模型生成的图像,$p(y|G(z))$为条件类别分布,表示分类模型对生成图像的类别预测概率;$p(y)$为所有生成图像的边际类别分布;D_{KL}表示KL(Kullback-Leibler)散度,衡量两个分布之间的不一致性。

CDS通过衡量生成图像类别的熵来反映单个类别的图像的多样性。分数越低,表明指定类别中其他类别图片更少,生成效果更好,其计算公式为

$$\mathrm{CDS}=\frac{1}{\log N}H\left(\frac{1}{|X|}\sum_{G(z)\in X}C(G(z))\right) \tag{13}$$

式中:N为类别总数;|X|为从特定类别的EEG信号生成的样本总数;C为分类器。CDS的范围在0到1之间。

FID通过计算两个多维高斯分布之间的Frechet距离来评估生成样本与真实样本在统计层面的接近程度。分数越低,表明生成图像与真实图像之间的分布越接近,生成质量越高,其计算公式为

$$\mathrm{FID}=\|\boldsymbol{\mu}_1-\boldsymbol{\mu}_2\|_2^2+\mathrm{Tr}\left(\boldsymbol{\Sigma}_1+\boldsymbol{\Sigma}_2-2\sqrt{\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2}\right) \tag{14}$$

式中:$\boldsymbol{\mu}_1$和$\boldsymbol{\mu}_2$分别表示真实图像和生成图像在特征空间中的均值向量,$\boldsymbol{\Sigma}_1$和$\boldsymbol{\Sigma}_2$则分别对应其协方差矩阵。

2.4 实验结果与分析

第一阶段的实验选择EEGNet[3]、TSception[26]、ShallowConvNet[27]、EEGConformer[10]、LCM[28]和WaveNet[29]网络模型,以及EEG2Image[30]论文中的方法作为对比模型,全面评估不同网络结构在脑电特征提取任务中的表现。其中EEGNet提出采用轻量级卷积结构来高效提取EEG信号中的时空特征;TSception通过引入时序-空间感受野模块,提取多尺度的动态脑电特征;ShallowConvNet使用浅层卷积架构聚焦于低频脑电波段的判别性信息提取;EEGConformer提出结合卷积与Transformer编码器,捕捉局部和全局的时空依赖;LCM通过局部校准机制增强关键通道特征的表达能力;WaveNet采用因果卷积结构建模EEG信号的时序依赖;EEG2Image提出利用LSTM编码器提取脑电特征并生成相应的图像。为了确保实验的公平性,所有实验均使用相同的训练集、测试集、训练迭代次数、学习率、批次大小和优化算法等参数。

对于上述各类模型,实验使用t-SNE工具[31]对Objects数据集的测试数据进行聚类分析,结果如图7所示,Characters数据集的结果类似。结果显示,ARTLNet提取的特征在不同类别之间具有明显的分隔,表明其能够有效地提取和区分不同类别的EEG信号特征,从而提升下游任务的分类准确率。相比于其他方法,ARTLNet在特征空间中表现出了较好的聚类效果,验证了自注意力机制和残差结构在捕捉有效特征方面的显著贡献。

如表1所示,ARTLNet在K-means、SVM和KNN分类器上的准确率均高于目前一些SOTA模型,且相较于先进的EEG2Image方法,K-means、SVM和KNN准确率在Characters数据集上分别提高了

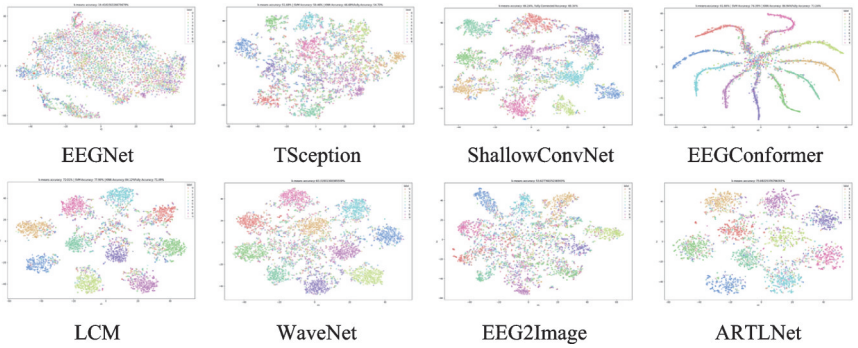


图7 各模型方法的t-SNE图

Fig.7 t-SNE plots for each modeling approach

表1 Characters和Objects数据集上不同模型的K-means、SVM和KNN准确率分数对比

Table 1 Comparison of K-means, SVM and KNN accuracy scores for different models on the Characters and Objects datasets

Method	Characters数据集			Objects数据集		
	K-means	SVM	KNN	K-means	SVM	KNN
EEGNet[3]	0.18	0.36	0.56	0.16	0.28	0.53
TSception[26]	0.48	0.56	0.68	0.52	0.58	0.67
ShallowConvNet[27]	0.66	0.73	0.78	0.71	0.76	0.81
EEGConformer[10]	0.62	0.74	**0.87**	0.69	0.76	0.86
LCM[28]	0.72	0.77	0.84	0.72	0.78	0.84
WaveNet[29]	0.71	0.75	0.84	0.64	0.71	0.78
EEG2Image[30]	0.54	0.55	0.66	0.53	0.64	0.74
ARTLNet	**0.75**	**0.79**	0.85	**0.76**	**0.79**	**0.86**

0.21、0.24和0.19,在Objects数据集上分别提高了0.23、0.15和0.12。这表明,ARTLNet在特征提取方面表现出色,可为生成对抗网络提供更有效的特征,从而生成更高质量的EEG图像。这得益于注意力机制在特征提取过程中有效捕捉了EEG信号中的关键时间序列信息,同时残差连接确保了特征在深度网络中的有效传递和保留。

第二阶段的实验选择AC-GAN[32]、C-Former[33]、ThoughtViz[34]、SPIE work[35]、NeuroGAN[14]和EEG2Image[30]作为对比模型方法,在Objects数据集上对生成图像进行评估。AC-GAN提出采用标签条件化的GAN变体;C-Former通过专门的Transformer-encoder模块对EEG信号进行编码,并将得到的高维特征作为条件输入生成网络来生成目标图像;ThoughtViz使用条件生成对抗网络(cGAN)将编码的EEG信号作为输入并生成相应的图像;SPIE work使用像素级损失以提升图像结构与语义还原效果;NeuroGAN提出使用感知损失和注意力模块来生成高质量的图像;EEG2Image提出使用LSTM提取脑电特征来生成刺激图像。尽管这些方法均在不同程度上提升了生成图像的质量,但它们在特征的细粒度提取与生成图像的精细度上存在一定的局限性。

如表2和表3所示,与ThoughtViz、EEG2Image等主流方法相比,本文方法在Objects数据集上,无论是整体还是单个类别的生成图像,其质量均优于其他模型。更高的IS分数、更低的CDS和FID分数

表2 不同方法在Objects数据集上的生成性能对比

Table 2 Comparison of generation performance of different methods on the Objects dataset

Method	IS	FID	CDS
AC-GAN(one-hot)[32]	3.10	/	/
AC-GAN[32]	4.90	/	/
C-Former[33]	5.10	/	0.650 1
ThoughtViz[34]	5.43	/	0.789 7
SPIE work[35]	5.70	/	/
NeuroGAN[14]	6.02	/	**0.405 1**
EEG2Image[30]	6.78	7.28	0.602 2
Ours	**6.98**	**6.25**	0.533 0

表3 各类别在Objects数据集上不同方法的CDS分数和FID分数对比

Table 3 Comparison of CDS and FID scores for different methods on the Objects dataset for each category

Class (ImageNet)	CDS				FID	
	ThoughtViz	C-Former	EEG2Image	Ours	EEG2Image	Ours
Apple	0.849 7	0.721 3	0.605 9	**0.556 0**	2.63	**2.17**
Car	0.839 1	0.562 1	0.599 1	**0.519 1**	4.15	**3.61**
Dog	0.696 5	0.632 7	0.614 2	**0.610 4**	2.41	**2.26**
Gold	0.656 1	**0.539 1**	0.626 9	0.624 9	8.40	**8.23**
Mobile	0.854 1	0.725 2	0.618 1	**0.567 4**	13.31	**11.76**
Rose	0.830 9	**0.584 3**	0.604 4	0.590 0	4.84	**4.21**
Scooter	0.630 9	0.591 5	0.542 6	**0.352 5**	10.35	**8.45**
Tiger	0.906 8	0.759 4	0.593 9	**0.398 7**	5.16	**4.28**
Wallet	0.815 3	0.755 3	0.601 1	**0.554 9**	8.32	**7.13**
Watch	0.821 5	0.629 7	0.616 0	**0.556 4**	13.21	**10.36**
Mean	0.789 7	0.650 1	0.602 2	**0.533 0**	7.28	**6.25**

表明,本文模型在生成图像的质量、类别区分以及生成图像与真实图像的相似性方面展现了更强的能力。通过精细的特征提取与条件控制,本文方法在时空特征的捕捉和图像生成的精度上具备明显的优势,从而可以实现更好的生成效果。

图8展示了主流的实验方法对比本文方法在Objects数据集上重建的图像,显然在图像清晰度方面,本文方法相比其他一些主流模型会更好,重建过程中的伪影也更少一些。

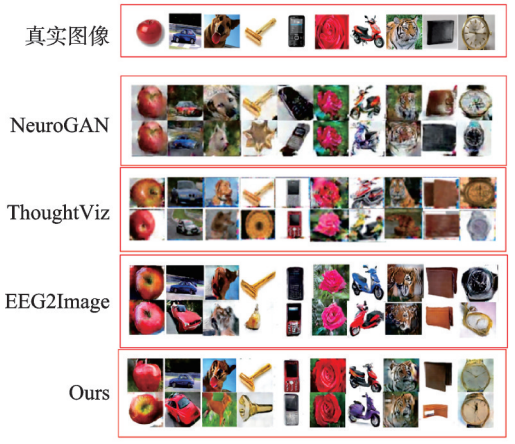


图8 本文方法与4种主流方法生成图像的结果对比

Fig.8 Comparison of image generation results between the proposed method and four mainstream methods

2.5 消融实验

为了验证本文所提ARTLNet模型的有效性和必要性,分别在Objects和Characters数据集上进行了系统的消融实验。首先,将本文提出的基于三元组损失的半监督模型与采用对比学习方法(SimCLR)在无标签数据中学习相似样本特征表示的无监督模型进行对比,然后再对ARTLNet模型的自注

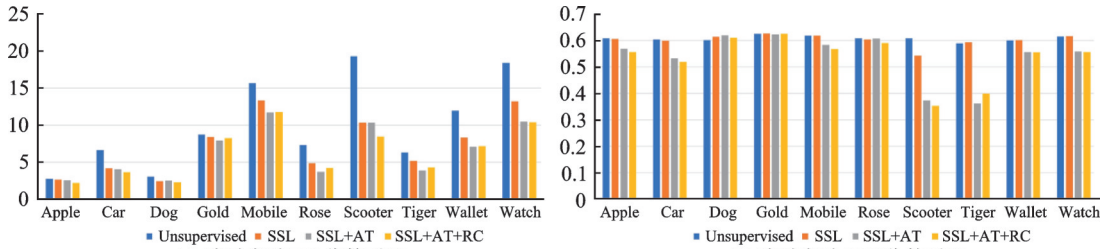
意力机制和残差模块进行逐一消融,通过逐步引入这些模块并比较模型性能的变化,分析各组件对整体模型性能的具体贡献。

表4显示了不同模型配置在Characters和Objects数据集上的分类准确率,其中Unsupervised表示无监督模型,SSL表示使用三元组损失的半监督模型,AT表示注意力模块,RC表示残差模块。从表4可以看出,半监督模型的分类准确率显著高于无监督模型,且加入自注意力机制后的模型的分类准确率明显上升,这表明自注意力机制在提升EEG信号中关键特征的提取效率方面起到了重要作用。再加入残差连接后的模型的准确率又得到了进一步提升,有效地缓解了梯度消失的问题,使得模型在深层网络中能够稳定地学习到有效特征。此外,图9展示了不同模型配置在Objects数据集上针对10种类别生成图像的质量评分对比,证明了ARTLNet模型更能适合图像生成任务,提高了图像生成质量。

表4 不同模型配置在Characters和Objects数据集上的准确率对比

Table 4 Accuracy comparison of different model configurations on Characters and Objects datasets

Method	Characters数据集			Objects数据集		
	K-means	SVM	KNN	K-means	SVM	KNN
Unsupervised	0.16	0.23	0.44	0.15	0.20	0.38
SSL	0.51	0.62	0.71	0.53	0.64	0.74
SSL+AT	0.73	0.75	0.81	0.70	0.70	0.73
SSL+AT+RC	**0.75**	**0.77**	**0.85**	**0.76**	**0.79**	**0.86**



(a) 消融实验FID分数对比

(a) Comparison of FID scores in ablation study

(b) 消融实验CDS分数对比

(b) Comparison of CDS scores in ablation study

图9 不同模型配置在Objects数据集上生成图像的质量分数对比

Fig.9 Comparison of mass fraction of images generated by different model configurations

3 结束语

为了解决EEG信号解码研究中遇到的特征提取不充分,图像生成质量差的问题,本文提出了一种新的用于解码脑电信号的ARTLNet网络模型,结合LSTM、残差块与注意力机制,实现了EEG信号的有效特征提取,并进一步利用深度卷积生成对抗网络DCGAN生成了高质量的图像。该模型利用自注意力机制聚焦于EEG信号中的关键时序特征,增强了模型的全局语义理解能力;残差连接则有效缓解了梯度消失问题,提升了特征的传递性和保留性。结合批量归一化和全局平均池化的生成过程优化,生成的图像在清晰度和一致性上有显著提升。实验结果显示,本文提出的ARTLNet在脑电信号解码与图像生成任务中表现出色,显著提升了生成图像的质量和类别区分度,证明了其在脑机接口等应用领域的巨大潜力。未来将研究如何在无监督学习框架下进一步提升EEG信号解码能力,特别是探索如何利用更大规模的EEG数据集与更加复杂的生成模型,以实现更高质量的图像生成和更加智能化的脑机接口应用。

参考文献：

[1] ORBAN M, ELSAMANTY M, GUO K, et al. A review of brain activity and EEG-based brain-computer interfaces for rehabilitation application[J]. Bioengineering, 2022, 9(12): 768.

[2] SHUQFA Z, BELKACEM A N, LAKAS A. Decoding multi-class motor imagery and motor execution tasks using Riemannian geometry algorithms on large EEG datasets[J]. Sensors, 2023, 23(11): 5051.

[3] 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): 056013.

[4] INGOLFSSON T M, HERSCHE M, WANG X, et al. EEG-TCNet: An accurate temporal convolutional network for embedded motor-imagery brain-machine interfaces[C]//Proceedings of 2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC). [S.l.]: IEEE, 2020: 2958-2965.

[5] BAI S, KOLTER J Z, KOLTUN V. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling[EB/OL]. (2018-03-04)[2018-04-19]. https://doi.org/10.48550/arXiv.1803.01271.

[6] 徐晴, 葛成, 蔡标, 等. 基于深度学习的癫痫脑电信号分类[J]. 数据采集与处理, 2022, 37(4): 787-797.
XU Qing, GE Cheng, CAI Biao, et al. Epileptic EEG signal classific-ation based on deep learning[J]. Journal of Data Acquisition and Processing, 2022, 37(4): 787-797.

[7] KAVASIDIS I, PALAZZO S, SPAMPINATO C, et al. Brain2image: Converting brain signals into images[C]//Proceedings of the 25th ACM International Conference on Multimedia. [S.l.]: ACM, 2017: 1809-1817.

[8] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. Neural Computation, 1997, 9(8): 1735-1780.

[9] REN C, CHEN J, LI R, et al. Semi-supervised pairwise transfer learning based on multi-source domain adaptation: A case study on EEG-based emotion recognition[J]. Knowledge-Based Systems, 2024, 305112669-112669.

[10] 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, 2023, 31: 710-719.

[11] QIAN D, ZENG H, CHENG W, et al. NeuroDM: Decoding and visualizing human brain activity with EEG-guided diffusion model[J]. Computer Methods and Programs in Biomedicine, 2024, 251: 108213.

[12] KINGMA D P, WELLING M. Auto-encoding variational bayes[EB/OL]. (2014-05-01). https://doi. org/10.48550/arXiv.1312.6114.

[13] KHARE S, CHOUBEY R N, AMAR L, et al. NeuroVision: Perceived image regeneration using cProGAN[J]. Neural Computing and Applications, 2022, 34(8): 5979-5991.

[14] MISHRA R, SHARMA K, JHA R R, et al. NeuroGAN: Image reconstruction from EEG signals via an attention-based GAN [J]. Neural Computing and Applications, 2023, 35(12): 9181-9192.

[15] ZHENG X, CHEN W, LI M, et al. Decoding human brain activity with deep learning[J]. Biomedical Signal Processing and Control, 2020, 56: 101730.

[16] SCHROFF F, KALENICHENKO D, PHILBIN J. FaceNet: A unified embedding for face recognition and clustering[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2015: 815-823.

[17] KUMAR P, SAINI R, ROY P P, et al. Envisioned speech recognition using EEG sensors[J]. Personal and Ubiquitous Computing, 2018, 22: 185-199.

[18] DE CAMPOS T E, BABU B R, VARMA M. Character recognition in natural images[C]//Proceedings of International Conference on Computer Vision Theory and Applications. 2009, 1: 273-280.

[19] DENG J, DONG W, SOCHER R, et al. ImageNet: A large-scale hierarchical image database[C]//Proceedings of 2009 IEEE Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2009: 248-255.

[20] GONI M R, RAHMAN T. Predictive modeling on MEG signal to classify hand and wrist movement using UNEQ and KNN [C]//Proceedings of 2020 IEEE Region 10 Symposium (TENSYMP). [S.l.]: IEEE, 2020: 815-818.

[21] MUCHERINO A, PAPAJORGJI P J, PARDALOS P M. K-nearest neighbor classification[M]//in Data Mining in Agriculture. New York, USA: Springer New York, 2009: 83-106.

[22] HEARST M A, DUMAIS S T, OSUNA E, et al. Support vector machines[J]. IEEE Intelligent Systems and Their Applications, 1998, 13(4): 18-28.
[23] SALIMANS T, GOODFELLOW I, ZAREMBA W, et al. Improved techniques for training GANs[J]. Advances in Neural Information Processing Systems, 2016, 29: 2234-2242.
[24] HEUSEL M, RAMSAUER H, UNTERTHINER T, et al. GANs trained by a two time-scale update rule converge to a local nash equilibrium[EB/OL]. (2018-01-12). https://doi.org/10.48550/arXiv.1706.08500.
[25] BEN-YOSEF M, WEINSHALL D. Multi-modal generative adversarial networks for diverse datasets[J].
[26] DING Y, ROBINSON N, GUAN Z C. TSception: Capturing temporal dynamics and spatial asymmetry from EEG for emotion recognition[J]. IEEE Transactions on Affective Computing, 2023, 14(3):2238-2250.
[27] TIBOR SCHIRRMEISTER R, GEMEIN L, EGGENSPERGER K, et al. Deep learning with convolutional neural networks for decoding and visualization of EEG pathology[J]. Human Brain Mapping, 2017, 38(11): 5391-5420.
[28] CHEN C S, CHEN Y J, TSAI H W. Large cognition model: Towards pretrained EEG foundation model[EB/OL]. (2025-02-11). https://doi.org/10.48550/arXiv.2502.17464.
[29] VAN DEN OORD A, DIELEMAN S, ZEN H, et al. WaveNet: A generative model for raw audio[EB/OL]. (2016-09-19). https://doi.org/10.48550/arXiv.1609.03499.
[30] SINGH P, PANDEY P, MIYAPURAM K, et al. EEG2IMAGE: Image reconstruction from EEG brain signals[C]//Proceedings of 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). [S.l.]: IEEE, 2023: 1-5.
[31] VAN DER MAATEN L, HINTON G. Visualizing data using t-SNE[J]. Journal of Machine Learning Research, 2008, 9(11).
[32] ODENA A, OLAH C, SHLENS J. Conditional image synthesis with auxiliary classifier GANs[C]//Proceedings of International Conference on Machine Learning. [S.l.]: PMLR, 2017: 2642-2651.
[33] MISHRA R, BHAVSAR A. Generating visual stimuli from EEG recordings using Transformer-encoder based EEG encoder and GAN[EB/OL]. (2024-11-20). https://doi.org/10.48550/arXiv.2402.10115.
[34] TIRUPATTUR P, RAWAT Y S, SPAMPINATO C, et al. Thoughtviz: Visualizing human thoughts using generative adversarial network[C]//Proceedings of the 26th ACM International Conference on Multimedia. [S.l.]: ACM, 2018: 950-958.
[35] MISHRA R, SHARMA K, BHAVSAR A. Reconstruction of visual stimulus from the EEG recordings via generative adversarial network[C]//Proceedings of Medical Imaging 2022: Image Processing. [S.l.]: SPIE, 2022, 12032: 512-520.

作者简介:

倪哲文(2000-),男,硕士研究生,研究方向:脑电信号处理,E-mail:2391598953@qq.com。

全海燕(1970-),**通信作者**,男,博士,副教授,硕士生导师,研究方向:信号处理、智能优化等,E-mail:quanhaiyan@163.com。

(编辑:王静)

EEG Signal-Driven Visual Image Reconstruction Model Based on Double Residual LSTM and DCGAN

NI Zhewen, QUAN Haiyan*

(School of Information Engineering and Automation, Kunming University of Science and Technology, Kunming 650500, China)

Abstract: Reconstructing visual images from electroencephalogram (EEG) signals has become an emerging frontier in brain-computer interface (BCI) research, offering substantial potential in medical image reconstruction, neural decoding, and cognitive state analysis. However, the inherently noisy, low-amplitude, and highly temporal characteristics of EEG signals pose considerable challenges to robust feature extraction and high-fidelity image synthesis. To address these limitations, this study aims to establish an effective EEG-driven visual reconstruction framework capable of capturing fine-grained temporal dynamics while ensuring semantic consistency in the generated images. The proposed model integrates a double residual long short-term memory (LSTM) architecture with a self-designed deep convolutional generative adversarial network (DCGAN). Specifically, an LSTM network based on attention residual network and Triplet loss (ARTLNet) is constructed to improve EEG feature extraction by combining residual learning, temporal modeling, and self-attention mechanisms. Batch normalization and global average pooling are further employed to enhance signal stability and suppress feature redundancy. In the reconstruction stage, a customized DCGAN incorporating feature fusion is adopted to enrich semantic representation and improve image clarity and diversity. Experimental evaluations on both Characters and Objects EEG datasets demonstrate that ARTLNet achieves consistently higher classification and clustering accuracy across multiple algorithms compared with baseline LSTM and non-residual architectures. The generated images exhibit clearer structural details and more distinguishable category attributes, verifying the effectiveness of the proposed generative strategy. The results demonstrate that the combination of residual enhanced temporal modeling and feature-fusion-based adversarial generation can significantly improve EEG-driven visual reconstruction performance. This study confirms the viability of exploiting advanced deep learning mechanisms to decode and visualize EEG information with improved interpretability, providing methodological support for future BCI-based image reconstruction and neural representation studies.

Highlights:

1. This paper introduces ARTLNet, an attention-enhanced double-residual LSTM architecture that significantly strengthens EEG temporal feature extraction for downstream reconstruction tasks.

2. This paper develops a customized DCGAN with a feature fusion mechanism to improve semantic consistency, image clarity, and category separability during EEG-driven visual generation.

Key words: deep learning; generative adversarial network; EEG signal decoding; self-attention; residual connectivity

---

Foundation item: National Natural Science Foundation of China (No.61861023).
Received: 2024-12-19; **Revised:** 2025-05-08
***Corresponding author**, **E-mail:** quanhaiyan@163.com.