

本文引用格式: 黄伟坤,谢伟.基于跨模态信息迁移的发音想象脑电信号分类方法[J].自动化与信息工程,2023,44(3):23-28.

HUANG Weikun, XIE Wei. Classification method of EEG signals of pronunciation in imagined based on cross modal information transfer[J]. Automation & Information Engineering, 2023,44(3):23-28.

基于跨模态信息迁移的发音想象脑电信号分类方法*

黄伟坤 谢伟

(广东工业大学, 广东 广州 510006)

摘要: 针对基于发音想象的脑机接口样本数据数量小、数据噪声大, 导致模型泛化能力差的问题, 提出一种基于跨模态信息迁移的发音想象脑电信号分类方法。该方法通过知识蒸馏, 将音频模态信息迁移到脑电模态, 从而提高模型的泛化能力; 通过多尺度学习来提高模型性能。在数据集 Kara One 中, 两个二分类任务的 AUC 分别为 68.28% 和 69.53%。实验结果表明, 该方法有效地提高了模型的性能。

关键词: 发音想象; 脑机接口; 跨模态; 知识蒸馏; 信息迁移

中图分类号: TP391

文献标志码: A

文章编号: 1674-2605(2023)03-0004-06

DOI: 10.3969/j.issn.1674-2605.2023.03.004

Classification Method of EEG Signals of Pronunciation in Imagined Based on Cross Modal Information Transfer

HUANG Weikun XIE Wei

(Guangdong University of Technology, Guangzhou 510006, China)

Abstract: Aiming at the problem that the datasets of brain-computer interface based on pronunciation in imagined is small and the data noise is loud, which leads to the poor generalization ability of the model, a classification method of EEG signals of pronunciation in imagined based on cross modal information transfer is proposed. In this method, the audio modal information is transferred to the EEG modality by knowledge distillation, so as to improve the generalization ability of the model. The method also improves the performance of the model through multi-scale learning. In the dataset Kara One, the AUC of two binary classification tasks is 68.28% and 69.53%, respectively. Experimental results demonstrate that this method effectively enhances the performance of the model.

Keywords: pronunciation in imagined; brain-computer interface; cross modal; knowledge distillation; information transfer

0 引言

基于发音想象的脑机接口能够帮助语言障碍患者与外界方便地沟通, 受到人们广泛关注, 具有广阔的应用前景^[1-2]。目前, 可用于脑机接口的脑信号主要有脑电图(electroencephalogram, EEG)、近红外光谱、脑磁图、磁共振成像等。相比于其他脑信号, EEG 具有采集安全方便、时间分辨率高且成本低等特点, 但其中含有多种伪迹和噪声, 且目前基于发音想象的脑机接口数据样本较少。在噪声干扰大且训练样本少的

情况下, 现有的机器学习模型无法取得较好的泛化性能, 容易出现过拟合现象。

为此, 本文提出一种基于跨模态信息迁移的发音想象脑电信号分类方法, 通过多模态协同学习的方式, 将一种模态数据(音频信号)中学习到的知识迁移到另一种模态数据(脑电信号)中, 即通过知识蒸馏的方法将音频模态数据与脑电模态数据的关联信息从教师网络(双流架构, 包括音频数据、脑电数据)迁移到学生网络(单流架构, 只有脑电数据)。

* 基金项目: 多模态舌部运动解码方法及其应用研究 (2018A030313306)

1 跨模态信息迁移方法

2015 年加拿大多伦多大学的 ZHAO 等^[3]提出基于深度置信网络 (deep belief network, DBN) 的发音想象脑电信号分类方法, 先提取 EEG 的经验特征, 再将这些经验特征作为 DBN 的输入对网络进行训练; 虽然采用了多模态数据融合的方法, 但没有采用跨模态信息迁移的方法, 导致模型在预测时需使用多个模态数据, 无法仅使用 EEG 单模态数据进行预测。2019 年加拿大不列颠哥伦比亚大学的 SAHA 等^[4]提出一种新的分层深度神经网络, 该网络由空间和时间卷积神经网络 (convolutional neural network, CNN) 分层组合后再与深度自编码器级联而成, 利用 6 个语音类别 (如鼻音、双唇音) 的预测发音信息作为音素和单词分类的中间步骤, 找到负责自然语音合成的辨别信号; 但仅使用了脑电信号这一单模态信息进行实验。2020 年马什哈德医科大学的 BAKHSHALI 等^[5]提出基于相关熵谱密度黎曼距离的发音想象脑电信号分类方法, 对来自不同通道脑电信号的相关熵谱密度 (correntropy spectral density, CSD) 矩阵进行估计, 并将这些矩阵之间的距离作为语音识别的度量, 但仅使用脑电信号这一单模态信息进行实验。

综上所述, 现有的发音想象脑电信号分类方法,

没有利用其他模态 (如音频信号) 与脑电模态的关联信息帮助单一模态 (脑电信号) 模型进行训练, 在小样本的情况下容易出现过拟合。为此, 本文提出一种基于跨模态信息迁移的发音想象脑电信号分类方法。该方法自适应地提高更具判别力的特征在局部近邻关系计算中的权重, 并集成不同 batch size 的知识蒸馏模型, 能够更加充分地利用多模态教师网络迁移潜在的特征分布信息来监督单模态学生网络的训练, 进一步缓解小样本问题的影响, 提高分类模型的泛化能力。

本文改进了 CHEN 等^[6]提出的局部性保留损失 (locality preserving loss, LP) 的知识蒸馏方法。CHEN 等通过保持学生、教师网络特征空间样本的拓扑关系一致性, 增强学生网络学习。但该方法在计算教师网络特征空间样本的近邻关系时, 没有考虑信噪比的高低特征对近邻关系计算的重要程度差异。另外, 由于网络训练采用小批量梯度下降法, 样本近邻关系并不是在所有的训练样本中计算, 而是在随机批次中计算, 因此不同大小的 batch size 中, 样本近邻关系的尺度不一样。如, 基于 t-SNE 的教师网络隐含层样本近邻分布可视化图如图 1 所示, 图中三角形和圆圈散点分别对应不同发音想象任务的脑电样本。

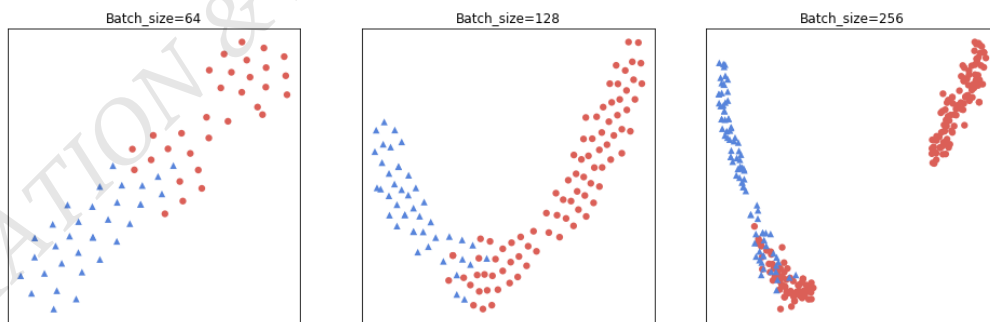


图 1 基于 t-SNE 的教师网络隐含层样本近邻分布可视化图

由图 1 可以看出, 在不同 batch size 训练的教师网络特征空间中, 样本间的近邻拓扑关系不一样。当 batch size 较大时, 更多地考虑样本间短程的拓扑关系; 当 batch size 较小时, 更多地考虑样本间长程的拓扑

关系。

鉴于以上分析, 本文方法首先计算教师网络特征与标签之间的相关系数, 并利用相关系数对样本间的近邻关系进行加权求和, 增加具有高信噪比 (或判别

力)的特征在样本间近邻关系计算中的重要程度;然后,集成不同 batch size 的近邻关系知识蒸馏模型,更加全面地利用教师网络特征空间中不同尺度的近邻关系对应的样本分布信息。

集成学习通过构建多个学习器并将其结合(模型融合),能更好地完成预测任务,提高模型性能。常用的集成学习方法有 Bagging^[7-9]、Boosting^[10-12]、Stacking^[13-15]等。由于神经网络训练是通过小批量样本计算近邻关系,通过设置不同的 batch size 来训练

模型学习不同的知识,如不同尺度的样本近邻关系。本文通过集成不同 batch size (64, 128, 256) 训练得到的 3 个模型来提高学生网络的预测能力,从而实现多尺度学习。

2 师生架构模型

本文的网络模型训练分为 2 个阶段:第一阶段是 1 个双流架构的教师网络;第二阶段是 1 个单流架构的学生网络,如图 2 所示。

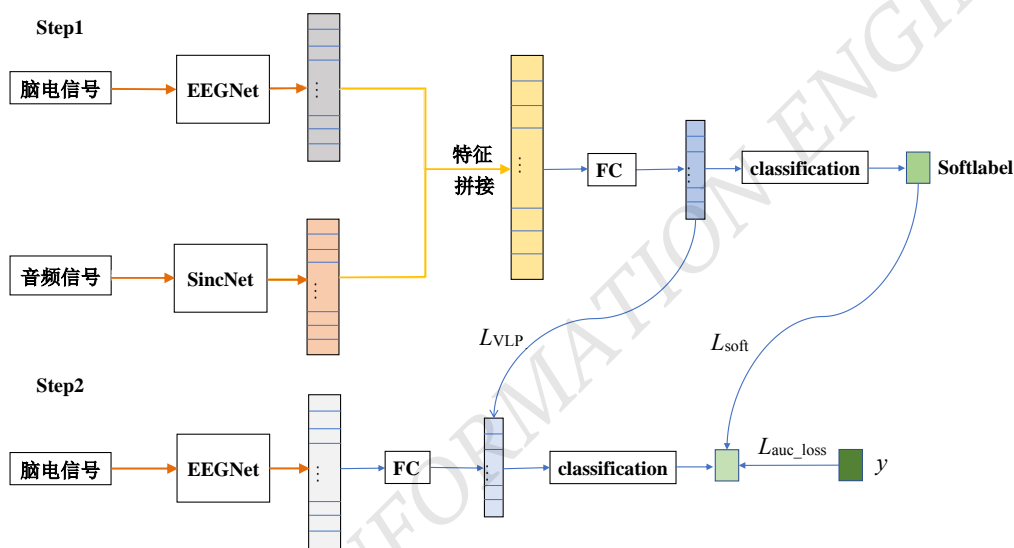


图 2 跨模态信息迁移的网络架构

第一阶段网络训练的过程:首先,使用 EEGNet^[16]和 SincNet^[17]特征提取网络分别提取脑电信号特征和音频信号特征;然后,将脑电信号特征和音频信号特征拼接;最后,经全连接层 FC 和分类层 classification 得到正样本的概率值。

教师网络的损失函数为

$$L_T = L_{\text{auc_loss}} \quad (1)$$

式中: L_T 为教师网络的损失函数, $L_{\text{auc_loss}}$ 为优化 ROC 曲线下的面积 (area under curve, AUC) 指标的损失函数,教师网络的损失函数可以采用文献[18]的设计:

$$R(x_i, y_j) = \begin{cases} -(x_i - y_j - \gamma)^p, & x_i - y_j < \gamma \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

$$L_{\text{auc_loss}} = \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} R(x_i, y_j) \quad (3)$$

式中: $\{x_0, \dots, x_{m-1}\}$ 为 m 个正样本的分类输出; $\{y_0, \dots, y_{n-1}\}$ 为 n 个负样本的分类输出; γ 为正负样本对分类输出差的阈值, $0 < \gamma \leq 1$; p 为调节因子, $p > 1$ 。

第一阶段的教师网络训练结束后,将教师网络的分类输出作为软标签约束学生网络的输出,达到信息迁移的目的。这种知识蒸馏方法是由 HINTON 等^[19]提出的。

学生网络的损失函数为

$$L_{\text{soft}} = -\frac{1}{N} \sum_{j=0}^{N-1} (p_j^T \log(q_j^T) + (1-p_j^T) \log(1-q_j^T)) \quad (4)$$

$$p_j^T = \frac{1}{1 + \exp(-v_j/T)}, q_j^T = \frac{1}{1 + \exp(-z_j/T)} \quad (5)$$

式中：\$N\$ 为样本的总数，\$j\$ 为样本序号，\$T\$ 为温度超参数，\$v_j\$ 为第一阶段教师网络分类输出的 logits 值，\$z_j\$ 为第二阶段学生网络分类输出的 logits 值。

由于教师网络和学生网络的输入和结构差异较大，因此，仅约束最终输出层是不够的。考虑到网络隐含层中的特征也包含有用信息，本文利用变权的局部保留损失来保证学生、教师网络特征空间拓扑关系的一致性。

假设 \$n\$ 个带标签的训练集表达式为 \$\{(x^1, y^1), (x^2, y^2), \dots, (x^n, y^n)\}\$，\$f_T^i\$ 和 \$f_S^i\$ 分别表示样本 \$x^i\$ 经过教师网络和学生网络提取后的特征，变权的局部保留损失函数表达式为

$$L_{\text{VLP}} = \frac{1}{k} \sum_{i,j} \alpha_{i,j} \|f_S^i - f_S^j\|_2^2 \quad (6)$$

$$\alpha_{i,j} = \begin{cases} \exp(-s \sum_d u_d^\beta (f_T^{i,d} - f_T^{j,d})^2), & \text{if } j \in N(i) \\ 0, & \text{otherwise} \end{cases} \quad (7)$$

$$w_d = \frac{\text{Cov}(f_T^d, y)}{\sqrt{\text{Var}(f_T^d) \text{Var}(y)}} \quad (8)$$

$$u_d^\beta = |w_d^\beta| / \sum_d |w_d^\beta| \quad (9)$$

式中：\$\alpha_{i,j}\$ 为教师网络隐含层特征空间样本间的局部关系，\$N(i)\$ 为特征 \$f_T^i\$ 的 \$k\$ 近邻，\$d\$ 为教师网络隐含层特征向量的维度序号，\$s\$ 为变权的尺度因子，\$\beta\$ 为指数因子，\$w_d\$ 为教师网络隐含层特征与标签之间的相关系数，\$u_d\$ 为变权的权重因子。

学生网络的损失函数表达式为

$$L_S = L_{\text{auc_loss}} + \eta_1 L_{\text{soft}} + \eta_2 L_{\text{VLP}} \quad (10)$$

式中：\$\eta\$ 为损失函数的折中系数。

由于教师网络隐含层特征空间的样本近邻关系

是在随机批次中计算的，不同 batch size 的训练模型，迁移教师网络隐含层样本近邻分布信息的尺度也不同。因此，本文采用线性回归法集成 3 个 batch size (64、128、256) 训练的学生网络，对脑电信号类别进行预测输出。

3 实验

本文实验使用 Kara One 公开数据集^[3]，用于脑机接口发音想象的研究。该数据集包含了音节和单词形式的想象语音，包括对应 7 个音位/音节 (/iy/, /piy/, /tiy/, /diy/, /uw/, /m/, /n/) 和 4 个单词 (pat, pot, knew, gnaw) 的多模态数据 (脑电信号、音频信号、面部跟踪信号)，共 14 个参与者的数据。本文采用脑电信号和音频信号 2 种模态数据。将数据集中的 10 个参与者的数据作为训练集 (含样本 1 353 例，其中正样本 492 例，负样本 861 例)；4 个参与者的数据作为测试集 (含样本 460 例，其中正样本 203 例，负样本 257 例)。

实验环境：计算机的处理器为 3.4 GHz Intel Core i7-6800K，内存为 NVIDIA GeForce RTX 2070 24 GB，深度学习框架为 Keras。

本文考虑 2 种二分类任务，即是否存在高前元音 (presence of high-front vowel, \$\pm\$ /iy/) 和是否存在高后元音 (presence of high-back vowel, \$\pm\$ /uw/)。AUC 是衡量二分类模型优劣的一种评价指标，本质是从样本集中随机选择一个正样本和负样本，模型预估正样本得分大于负样本得分的概率，计算公式为

$$AUC = \frac{\sum_{i \in \text{positiveClass}} \text{rank}_i - \frac{M(1+M)}{2}}{M \times N} \quad (11)$$

式中：\$\text{positiveClass}\$ 为正样本，\$\text{rank}_i\$ 为第 \$i\$ 条样本的序号 (将所有的样本的预测输出进行排序，排在第 \$\text{rank}\$ 个位置)，\$M\$ 为正样本的数量，\$N\$ 为负样本的数量。

本文比较了单模态方法 EEGNet (方法 1)、决策层知识蒸馏 (方法 2)、局部近邻关系保留知识蒸馏

(方法 3)、变权局部近邻关系保留知识蒸馏(方法 4)、及多尺度变权局部近邻关系保留知识蒸馏等方法(本文方法)的性能优劣,评价指标为 AUC,实验结果如表 1、表 2 所示。

表 1 \pm /iy/分类任务的实验结果

方法	使用模态	AUC (%)
EEGNet ^[16] (方法 1)	训练: 脑电 测试: 脑电	52.38
EEGNet ^[16] +SincNet ^[17] + 决策层知识蒸馏 ^[19] (方法 2)	训练: 脑电+音频 测试: 脑电	54.80
EEGNet ^[16] +SincNet ^[17] +决策层知识蒸馏 ^[19] + 局部近邻关系保留知识蒸馏 ^[6] (方法 3)	训练: 脑电+音频 测试: 脑电	61.81
EEGNet ^[16] +SincNet ^[17] +决策层知识蒸馏 ^[19] + +变权局部近邻关系保留知识蒸馏 (方法 4)	训练: 脑电+音频 测试: 脑电	64.40
EEGNet ^[16] +SincNet ^[17] +决策层知识蒸馏 ^[19] + 多尺度变权局部近邻关系保留知识蒸馏 (本文方法)	训练: 脑电+音频 测试: 脑电	68.28

由表 1 可知: 方法 1 中仅使用脑电信号模态训练模型, AUC 最低, 仅为 52.38%; 方法 2 在方法 1 的基础上增加了音频信号帮助模型进行训练, 并使用决策层知识蒸馏方法将教师网络中的知识迁移到学生网络, AUC 提高了约 2.42%, 验证了跨模态信息迁移方法的有效性; 方法 3 在方法 2 的基础上增加了局部近邻关系保留知识蒸馏方法, AUC 提高了约 7%, 表明教师网络和学生网络的输入和结构差异较大时, 利用局部近邻关系保留知识蒸馏方法可以保证学生、教

师网络特征空间拓扑关系的一致性, 提升学生网络性能; 方法 4 在局部近邻关系保留知识蒸馏方法中增加了变权因子, AUC 提高了约 2.59%, 表明变权因子可以增加具有高信噪比(或判别力)的特征在样本间近邻关系计算中的重要程度, 提高模型的性能; 本文方法在方法 4 的基础上集成 3 个 batch size (64、128、256)的近邻关系知识蒸馏模型, 更加全面地利用教师网络特征空间中不同尺度的近邻关系对应的样本分布信息, 性能表现最好。

表 2 \pm /uw/分类任务的实验结果

方法	使用模态	AUC (%)
EEGNet ^[16] (方法 1)	训练: 脑电 测试: 脑电	53.53
EEGNet ^[16] +SincNet ^[17] + 决策层知识蒸馏 ^[19] (方法 2)	训练: 脑电+音频 测试: 脑电	56.10
EEGNet ^[16] +SincNet ^[17] +决策层知识蒸馏 ^[19] + 局部近邻关系保留知识蒸馏 ^[6] (方法 3)	训练: 脑电+音频 测试: 脑电	58.61
EEGNet ^[16] +SincNet ^[17] +决策层知识蒸馏 ^[19] + +变权局部近邻关系保留知识蒸馏 (方法 4)	训练: 脑电+音频 测试: 脑电	65.52
EEGNet ^[16] +SincNet ^[17] +决策层知识蒸馏 ^[19] + 多尺度变权局部近邻关系保留知识蒸馏 (本文, 方法 5)	训练: 脑电+音频 测试: 脑电	69.53

由表 1、表 2 可知, 方法 5 相比于方法 1 性能有较大提升, AUC 提高了约 16%, 可见本文方法能够提升脑电信号分类的性能。

4 结论

针对在噪声干扰大且训练样本少的情况下, 现有

的机器学习模型无法取得较好的泛化性能,容易出现过拟合现象,本文提出一种基于跨模态信息迁移的发音想象脑电信号的分类方法,并在公开数据集 Kara One 上获得较高的 AUC 得分(±/iy/: 68.28%, ±/uw/: 69.53%)。该方法通过变权局部邻关系保留知识蒸馏自适应地提高了更具有判别力的特征在局部邻关系计算中的权重,并利用多尺度学习集成了不同 batch size (64、128、256) 的知识蒸馏模型,能够更加充分地多模态教师网络迁移潜在的特征分布信息监督单模态学生网络的训练,从而进一步缓解小样本问题的影响,提高分类模型的泛化能力。

参考文献

- [1] 陈霏,潘昌杰.基于发音想象的脑机接口的研究综述[J].信号处理, 2020,36(6):86-830.
- [2] 韩震坤,陶庆关,向恒.脑机接口技术的仿人控制综述[J].机电工程技术,2021,50(4):7-13;34.
- [3] ZHAO S, RUDZICZ F. Classifying phonological categories in imagined and articulated speech[C]. 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2015: 992-996.
- [4] SAHA P, ABDUL-MAGEED M, FELLS S. Speak your mind! Towards imagined speech recognition with hierarchical deep learning[J]. CoRR, arXiv preprint arXiv:1904.05746, 2019.
- [5] BAKHSHALI MA, KHADEMI M, EBRAHIMMOGH-ADAM A, et al. EEG signal classification of imagined speech based on riemannian distance of correntropy spectral density[J]. Biomedical Signal Processing and Control, 2020,59(C):101899.
- [6] CHEN H, WANG Y, XU C, et al. Learning student networks via feature embedding[J]. IEEE Transactions on Neural Networks and Learning Systems, 2021,32(1):25-35.
- [7] LEE Tae-Hwy, ULLAH Aman, WANG Ran. Bootstrap aggregating and random forest[J]. Macroeconomic Forecasting in the Era of Big Data. Springer Cham, 2020:389-429.
- [8] WANG RU, PENG J. Learning directed acyclic graphs via bootstrap aggregating[J]. arXiv preprint arXiv: 1406.2098, 2014.
- [9] HASSAN Ahnaf Rashik, BHUIYAN Mohammed Imamul Hassan. Computer-aided sleep staging using complete ensemble empirical mode decomposition with adaptive noise and bootstrap aggregating[J]. Biomedical Signal Processing and Control, 2016,24:1-10.
- [10] SCHAPIRE, ROBERT E. The boosting approach to machine learning: an overview[J]. Nonlinear Estimation and Classification, Springer, 2003:149-171.
- [11] LIU Shixia, XIAO Jiannan, LIU Junlin, et al. Visual diagnosis of tree boosting methods[J]. IEEE Transactions on Visualization and Computer Graphics, 2018,24(1):163-173.
- [12] JIANG J, WANG R, WANG M, et al. Boosting tree-assisted multitask deep learning for small scientific datasets[J]. Journal of Chemical Information and Modeling, 2020,60(3):1235-1244.
- [13] DIVINA Federico, GILSON Aude, GOMÉZ Vela Francisco, et al. Stacking ensemble learning for short-term electricity consumption forecasting[J]. Energies, 2018,11(4):949.
- [14] RIYAZ Sikora. A modified stacking ensemble machine learning algorithm using genetic algorithms[J]. Handbook of Research on Organizational Transformations Through Big Data Analytics. IGI Global, 2015:43-53.
- [15] CUI Shaoze, YIN Yunqiang, WANG Dujuan, et al. A stacking-based ensemble learning method for earthquake casualty prediction[J]. Applied Soft Computing, 2021,101: 107038.
- [16] LAWHERN VJ, SOLON AJ, WAYTOWICH NR, et al. EEGNet: a compact convolutional network for EEG-based brain-computer interfaces[J]. Journal of Neural Engineering, 2016,15(5):056013.1-056013.17.
- [17] RAVANELLI M, BENGIO Y. Interpretable convolutional filters with SincNet[J]. arXiv preprint arXiv:1811. 09725, 2018.
- [18] YAN L, DODIER RH, MOZER M, et al. Optimizing classifier performance via an approximation to the Wilcoxon-Mann-Whitney statistic[C]. Proceedings of the Twentieth International Conference on Machine Learning (ICML-2003), Washington DC, 2003:848-855.
- [19] HINTON G, VINYALS O, DEAN J. Distilling the knowledge in a neural network[J]. Computer Science, 2015,14(7):38-39.

作者简介:

黄伟坤,男,1996年生,硕士研究生,主要研究方向:深度学习。E-mail: 532190822@qq.com

谢伟,男,1995年生,硕士研究生,主要研究方向:深度学习。E-mail: 1341601296@qq.com