



智能系统学报

CAAI Transactions on Intelligent Systems

ISSN 1673-4785, CN 23-1538/TP

《智能系统学报》网络首发论文

题目: 基于增强 AlexNet 的音乐流派识别研究
作者: 刘万军, 孟仁杰, 曲海成, 刘腊梅
收稿日期: 2019-09-16
网络首发日期: 2020-08-28
引用格式: 刘万军, 孟仁杰, 曲海成, 刘腊梅. 基于增强 AlexNet 的音乐流派识别研究 [J/OL]. 智能系统学报.
<https://kns.cnki.net/kcms/detail/23.1538.TP.20200827.1317.010.html>



网络首发: 在编辑部工作流程中, 稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定, 且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式 (包括网络呈现版式) 排版后的稿件, 可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定; 学术研究成果具有创新性、科学性和先进性, 符合编辑部对刊文的录用要求, 不存在学术不端行为及其他侵权行为; 稿件内容应基本符合国家有关书刊编辑、出版的技术标准, 正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性, 录用定稿一经发布, 不得修改论文题目、作者、机构名称和学术内容, 只可基于编辑规范进行少量文字的修改。

出版确认: 纸质期刊编辑部通过与《中国学术期刊 (光盘版)》电子杂志社有限公司签约, 在《中国学术期刊 (网络版)》出版传播平台上创办与纸质期刊内容一致的网络版, 以单篇或整期出版形式, 在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊 (网络版)》是国家新闻出版广电总局批准的网络连续型出版物 (ISSN 2096-4188, CN 11-6037/Z), 所以签约期刊的网络版上网络首发论文视为正式出版。

DOI: [10.11992/tis.201909032](https://doi.org/10.11992/tis.201909032)

基于增强 AlexNet 的音乐流派识别研究

刘万军, 孟仁杰, 曲海成, 刘腊梅

(辽宁工程技术大学软件学院, 辽宁 葫芦岛 125105)

摘要: 针对机器学习模型对音乐流派特征识别能力较弱的问题, 提出了一种基于深度卷积神经网络的音乐流派识别 (DCNN-MGR) 模型。该模型首先通过快速傅里叶变换提取音频信息, 生成可以输入 DCNN 的频谱并切割生成频谱切片。然后通过融合带泄露整流 (Leaky ReLU) 函数、双曲正切 (Tanh) 函数和 Softplus 分类器对 AlexNet 进行增强。其次将生成的频谱切片输入增强的 AlexNet 进行多批次的训练与验证, 提取并学习音乐特征, 得到可以有效分辨音乐特征的网络模型。最后使用输出模型进行音乐流派识别测试。实验结果表明, 增强的 AlexNet 在音乐特征识别准确率和网络收敛效果上明显优于 AlexNet 及其他常用的 DCNN、DCNN-MGR 模型在音乐流派识别准确率上比其他机器学习模型提升了 4%~20%。

关键词: 音乐流派识别; 深度卷积神经网络; 机器学习; 深度学习; AlexNet; 音频特征提取; 音乐特征识别

中图分类号: TP181 **文献标志码:** A **文章编号:** 1673-4785(2020)04-0001-08

中文引用格式: 刘万军, 孟仁杰, 曲海成, 等. 基于增强 AlexNet 的音乐流派识别研究 [J]. 智能系统学报, 2020, 15(4): 1-8.

英文引用格式: LIU Wanjun, MENG Renjie, QU Haicheng, et al. Music genre recognition research based on enhanced AlexNet[J].

CAAI transactions on intelligent systems, 2020, 15(4): 1-8.

Music genre recognition research based on enhanced AlexNet

LIU Wanjun, MENG Renjie, QU Haicheng, LIU Lamei

(College of Software, Liaoning Technical University, Huludao 125105, China)

Abstract: To solve the problem that machine learning model has weak ability to identify music genre features, a music genre recognition model based on deep convolutional neural network (DCNN-MGR) is proposed in this paper. At first, the model extracts audio information through Fast Fourier Transformation, generating spectrums that can be input to the DCNN and slicing the generated spectrums. Then AlexNet is enhanced by fusion of Leaky ReLU function, Tanh function and Softplus classifier. The generated spectrum slices are input into the enhanced AlexNet for multi-batch training and verification. Music features are extracted and learned, and a network model that can effectively distinguish music features is obtained. At last, the output model is applied to music genre recognition and test. The experimental results show that the enhanced AlexNet is superior to AlexNet and other commonly used DCNN in terms of accuracy of music feature recognition and network convergence effect. The DCNN-MGR model is 4%~20% higher than other machine learning models in music genre recognition accuracy.

Keywords: music genres recognition; deep convolutional neural network; machine learning; deep learning; AlexNet; audio feature extraction; audio feature extraction

音乐流派是被提及最多的音乐标签之一。随着互联网曲库容量的增加, 按流派检索音乐成为音乐信息检索的主流方法, 同时也是音乐服务平台为用户推荐音乐的重要基础。自动且精准地进

行音乐流派识别可以有效减少人力成本。常用的音乐流派识别模型一般包括训练和测试两个阶段。在训练阶段, 首先通过建立数学模型描绘具有区分度的音乐流派数字特征; 然后采用预加重、梅尔滤波、倒谱提升等方式提取音乐文件的数字特征; 最后基于不同流派的数字特征和分布特性训练分类器。在测试阶段, 使用与训练阶段

收稿日期: 2019-09-16.

基金项目: 国家自然科学基金青年基金项目 (41701479).

通信作者: 孟仁杰. E-mail: mengrenjie95@163.com.

相同的方式进行数字特征提取,利用在训练阶段得到的分类器对提取到的数字特征进行计算并评估流派。

常用的机器学习音乐流派识别模型主要有支持向量机 (support vector machine, SVM) 模型、K-近邻 (K-nearest neighbor, KNN) 模型、梯度提升 (gradient boosting) 模型和极端随机树 (extra trees) 模型。邵曦等^[1]利用音乐的梅尔频率倒谱系数 (Mel frequency cepstrum coefficient, MFCC) 特征构造了 SVM 模型,对 5 个流派的音乐进行识别,取得了 81.20% 的识别准确率。Ali 等^[2]将 MFCC 特征输入 KNN 模型,在 GTZAN 数据集^[3]上进行实验,验证了 KNN 模型对音乐流派识别的有效性。Murauer 等^[4]对比了 Gradient Boosting 模型和 Extra Trees 模型对音乐流派多维数字特征的提取能力,证明了两个模型能够有效识别音乐流派特征信息。但常用的机器学习音乐流派识别模型存在容易过拟合的问题,在完成一定数据量的训练之后,输入更多的数据进行训练,无法得到更好的扩展。因此常用的机器学习音乐流派识别模型在处理大量的训练样本时表现较差^[5]。

近年来,学者们尝试使用深度学习模型来提高音乐流派自动识别的准确率,采用循环神经网络 (recurrent neural network, RNN) 构建深度学习音乐流派识别模型逐渐成为当前的研究热点。Dai 等^[6]提出了融合段特征的概念,验证了音乐特征分段提取的有效性,成功将长短时记忆网络 (long short-term memory, LSTM) 引入音乐流派识别领域。Jakubik^[7]采用音乐长片段和短片段两种数据格式进行模型训练,建立了有效的 LSTM 和门控循环单元 (gated recurrent unit, GRU) 音乐流派识别模型,并将 LSTM 模型和 GRU 模型进行了对比,在 GTZAN 数据集上,GRU 模型取得了更高的音乐流派识别准确率。但是 LSTM 和 GRU 难以表达具有良好区分度的音乐特征^[8],因此 LSTM 和 GRU 在音乐流派识别领域表现不够理想。而图像识别领域中的深度卷积神经网络 (deep convolutional neural network, DCNN) 模型具有权值共享和良好的局部感知特性^[9],能够学习音乐的频谱图像中局部细腻的音乐流派特征和频率节奏等变化。因此本文在对比了 AlexNet^[10]、VGGNet16^[11]、VGGNet19^[11]、GoogLeNet^[12] 和 ResNet-34B^[13] 后,选择了参数较少,参数空间较小,不易导致过拟合和梯度消失的 AlexNet^[14] 做出增强,基于增强的 AlexNet 构建了深度卷积神经网络音乐流派识别 (deep convolutional neural network music genre recognition, DCNN-MGR) 模型,旨在进一步提高机器学习模型对音乐流派自动识别的

准确率。

1 AlexNet 网络

Krizhevsky 等^[10]提出了由 5 个卷积层、3 个最大池化层、3 个局部响应归一化 (local response normalization, LRN) 层、3 个全连接层堆叠成的 AlexNet。AlexNet 使用线性整流 (rectified linear unit, ReLU) 函数作为卷积层和全连接层的激活函数,有效解决了网络较深时的梯度消失问题。AlexNet 利用 GPU 的并行计算来处理神经网络训练时大量的矩阵运算,使卷积神经网络的训练变得更加高效,从而增强了 DCNN 在工业生产中的实用性。AlexNet 在医疗诊断^[15-17]、目标检测^[18-19]、人脸识别^[20] 等领域有着广泛的应用, AlexNet 的网络结构如图 1 所示。

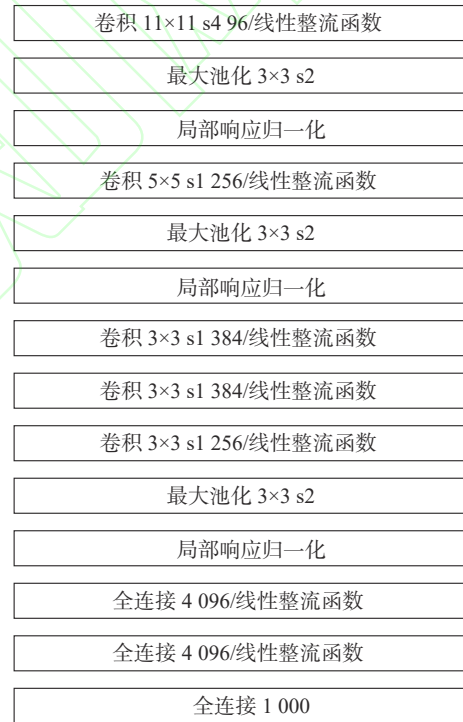


图 1 AlexNet 网络结构

Fig. 1 AlexNet network structure

为了避免网络在训练过程中出现过拟合现象, AlexNet 在全连接层间应用 Dropout。在进行前向传播时以概率 $p = 0.5$ 随机删除网络中隐藏层的神经元。然后进行反向传播,更新参数,恢复被删除的神经元,再随机删除神经元,重复以上过程,保持输入输出的神经元数量不变,减少神经元间的相互作用。Dropout 的计算过程如式 (1)~(4) 所示。

$$\mathbf{r}_j^{(l)} \sim \text{Bernoulli}(p) \quad (1)$$

$$\tilde{\mathbf{y}}^{(l)} = \mathbf{r}^{(l)} \times \mathbf{y}^{(l)} \quad (2)$$

$$z_i^{(l+1)} = w_i^{(l+1)} \tilde{y}^l + b_i^{(l+1)} \quad (3)$$

$$y_i^{(l+1)} = f(z_i^{(l+1)}) \quad (4)$$

式中: l 表示隐藏层数; $\text{Bernoulli}(p)$ 生成概率向量 r ; \tilde{y} 表示以概率 r 使得部分神经元停止工作后的中间结果; z 表示输入到 l 层中的向量; w 和 b 表示 l 层的权重和偏差。

2 DCNN-MGR 模型

2.1 AlexNet-enhanced 网络

2.1.1 网络结构及模型

本文在 AlexNet 的基础上, 针对 DCNN 的音乐特征提取能力和频谱识别性能, 对网络性能进行增强并改进其部分结构, 提出了 AlexNet-enhanced。AlexNet-enhanced 由 5 个卷积层、3 个最大池化层、3 个全连接层及 Softplus 分类器构成, AlexNet-enhanced 的网络结构如图 2 所示, 网络模型如图 3 所示。

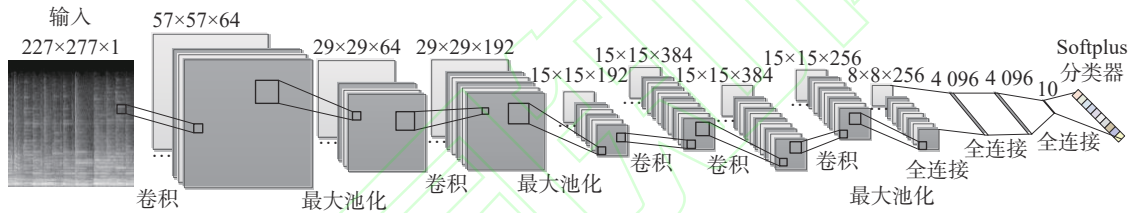


图 3 AlexNet-enhanced 网络模型

Fig. 3 AlexNet-enhanced network model

2.1.2 网络性能增强

增强 1 AlexNet-enhanced 将卷积层的激活函数——ReLU 函数替换为带泄露整流 (leaky rectified linear unit, Leaky ReLU) 函数。Leaky ReLU 激活函数计算公式^[21]为

$$f(x) = \begin{cases} 0.01x, & x < 0 \\ x, & x \geq 0 \end{cases} = \max(0.01x, x) \quad (5)$$

Leaky ReLU 函数可以有效解决 ReLU 函数带来的神经元坏死问题。Leaky ReLU 函数的导数不为零, 可以减少静默神经元的出现。同时其自变量在负区间取值时, 输出值不为 0, 避免了 ReLU 函数进入负区间后神经元不学习的问题。因此, 在特征提取过程中, Leaky ReLU 函数能够有效提高特征信息的完整性和利用率。

增强 2 AlexNet-enhanced 将全连接层的激活函数 ReLU 函数替换为双曲正切 (hyperbolic tangent, Tanh) 函数。Tanh 激活函数计算公式为

$$f(x) = \frac{\sinh(x)}{\cosh(x)} = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (6)$$

Tanh 函数在自然语言处理领域有着广泛的应用, 其函数的零均值化处理, 可以在迭代过程

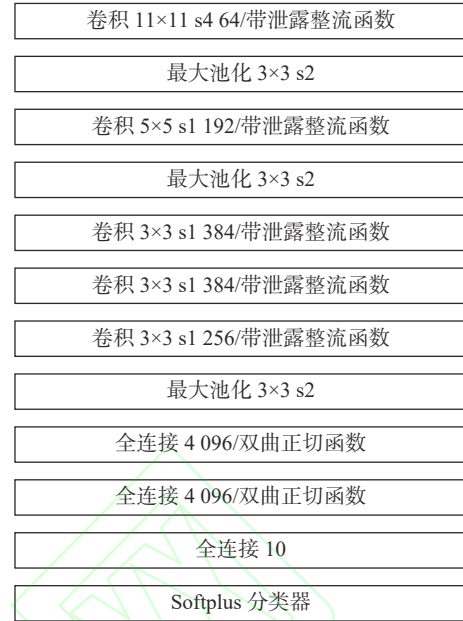


图 2 AlexNet-enhanced 网络结构

Fig. 2 AlexNet-enhanced network structure

中不断增强网络对音乐特征的提取效果。

2.1.3 网络结构改进

改进 1 为了使网络更易并行化, AlexNet-enhanced 去掉了对音乐特征识别泛化能力增益较弱的 LRN 层。

改进 2 AlexNet-enhanced 将第 1 个卷积层的输出节点减少至 64 个, 第 2 个卷积层的输出节点减少至 192 个, 最后一个全连接层的输出节点减少至 10 个, 通过减少冗余参数来加速收敛网络, 减少网络的训练时间。

改进 3 AlexNet-enhanced 在提取特征的全连接层后加入 Softplus^[22] 分类器进行特征分类。Softplus 计算公式为

$$\zeta(x) = \log(1 + e^x) \quad (7)$$

Softplus 具有单侧抑制、相对宽阔的兴奋边界等优点, 其在定义域内连续可导, 使得梯度可在整个定义域内传播, 更接近生物特性^[23]。

2.2 模型构建

2.2.1 模型流程图

DCNN-MGR 模型通过快速傅里叶变换 (fast

fourier transform, FFT)^[24] 提取音频信息,生成可以输入 DCNN 的频谱,并将频谱进行切割后输入 DCNN。以模型完成训练集中全部音乐文件的一次训练,验证集中全部音乐文件的一次验证,作为一个批次,多次迭代训练模型和验证模型,在达到指定批次时输出音乐流派识别模型。模型流程如图 4 所示。

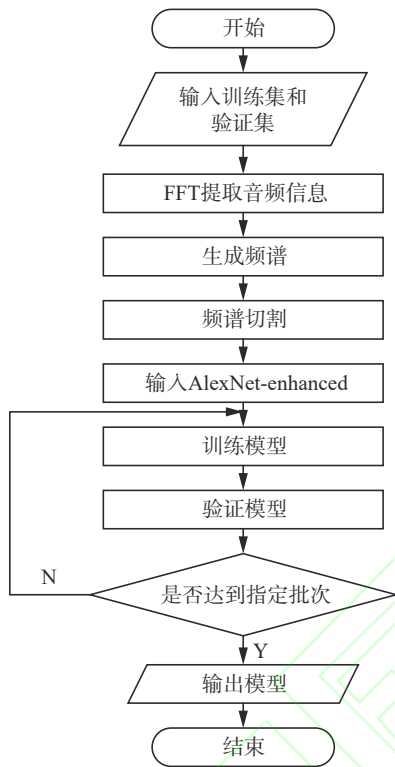


图 4 DCNN-MGR 模型流程图

Fig. 4 DCNN-MGR model flow chart

2.2.2 音频处理

DCNN-MGR 模型提取音频信息的方法为: 通过将音频数据的声道由立体声转换为单声道,降低信息冗余;通过 FFT 将时域数据转换到频域,使数据更加简单和紧凑,生成频谱。模型将频谱切分为 227 像素×227 像素的切片后输入 DCNN,以此更加拟合人类的学习和认知规律。音频处理得到的频谱切片如图 5 所示。



图 5 音频处理得到的频谱切片图

Fig. 5 Spectrum sliced image obtained by audio processing

频谱切片以 x 轴表示时间, y 轴表示频率,其中顶部为最高频率,底部为最低频率;以灰度显示频率的缩放幅度,其中白色为最大值,黑色为最小值。

3 实验及结果分析

3.1 实验环境

本文实验主要在 Intel Core i7 2.9 GHz CPU + Nvidia GeForce GTX 2080 Ti GPU 配置的服务器上进行,另外采用基于 Python 语言的 TensorFlow 深度学习框架构建 DCNN 进行对比实验。

3.2 数据集

GTZAN 数据集是音乐流派识别领域常用的公开数据集,其音乐数据分为 10 个流派,分别为流行、古典、金属、爵士、雷鬼、蓝调、迪斯科、嘻哈、乡村、摇滚。

GTZAN 数据集共有 1 000 条音乐数据,本文实验以 800 条音乐数据作为训练集,100 条音乐数据作为验证集进行监督学习,100 条音乐数据作为测试集进行音乐流派识别准确率测试。

3.3 DCNN 对比实验

3.3.1 网络评价指标

本文采用频谱识别准确率和频谱识别损失函数值作为 DCNN 的性能评价指标。其中频谱识别准确率指网络对频谱切片的识别准确率,频谱识别损失函数为适用于多分类问题的交叉熵函数。交叉熵函数计算公式为

$$\text{loss} = - \sum_{i=1}^n p(x_i) \log q(x_i) \quad (8)$$

式中: n 表示样本数; i 表示类别数; $p(x)$ 是分布目标; $q(x)$ 是预测的匹配分布,可以将交叉熵函数看作通过 $q(x)$ 表示 $p(x)$ 的困难程度,交叉熵函数的函数值越小,网络收敛效果越好。

3.3.2 实验结果分析

AlexNet-enhanced 与常用 DCNN 以学习率 0.001 迭代 31 000 次的训练集频谱识别准确率变化曲线如图 6 所示。在迭代 31 000 次后的训练集频谱识别准确率见表 1。

DCNN 通过梯度下降算法训练网络,迭代更新权重参数,网络从欠拟合趋向于最佳拟合^[25],应用更新后的权重参数计算频谱识别准确率。随着迭代次数的增加,频谱识别准确率逐渐提升,并趋于稳定。根据图 6 看出,各网络迭代 31 000 次后,训练集频谱识别准确率均趋于稳定, AlexNet-enhanced 的训练集频谱识别准确率高出常用

DCNN。表1数据表明, AlexNet-enhanced 的训练集频谱识别准确率比常用 DCNN 高出 0.50%~4.05%。

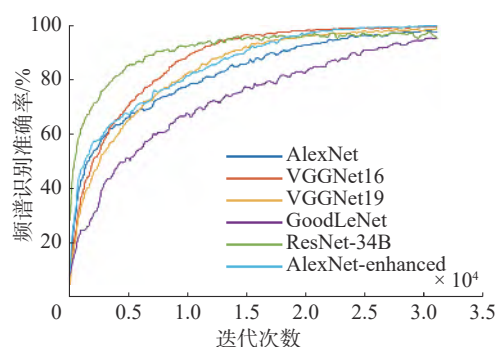


图6 各网络训练集频谱识别准确率变化曲线

Fig. 6 Spectrum recognition accuracy curve of each network in the training set

表1 各网络训练集频谱识别准确率

Table 1 Spectrum recognition accuracy of each network in the training set

网络	频谱识别准确率/%
AlexNet	97.40
VGGNet16	98.93
VGGNet19	97.93
GoogLeNet	95.38
ResNet-34B	95.66
AlexNet-enhanced	99.43

AlexNet-enhanced 与常用 DCNN 以学习率 0.001 迭代 31 000 次的训练集频谱识别损失函数值变化曲线如图7所示。在迭代 31 000 次后的训练集频谱识别损失函数值见表2。

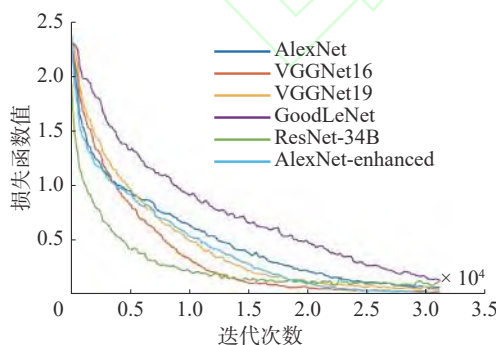


图7 各网络训练集频谱识别损失函数值变化曲线

Fig. 7 Loss curve of each network in the training set

DCNN 在最小化频谱识别损失函数值时,通过梯度下降算法逐步迭代求解。随着迭代次数的增加,频谱识别损失函数值逐渐下降,并趋于稳定。根据图7可以看出,各网络在迭代 31 000 次后,训练集频谱识别损失函数值均趋于稳定,

AlexNet-enhanced 的训练集频谱识别损失函数值低于常用 DCNN。表2数据表明, AlexNet-enhanced 的训练集频谱识别损失函数值比常用 DCNN 低 0.028 1~0.132 4。

表2 各网络训练集频谱识别损失函数值

Table 2 Loss of each network in the training set

网络	损失函数值
AlexNet	0.122 4
VGGNet16	0.101 5
VGGNet19	0.205 8
GoogLeNet	0.136 4
ResNet-34B	0.147 1
AlexNet-enhanced	0.073 4

AlexNet-enhanced 与常用 DCNN 以学习率 0.001 迭代 31 000 次的验证集频谱识别准确率变化曲线如图8所示。在迭代 31 000 次后的验证集频谱识别准确率见表3。

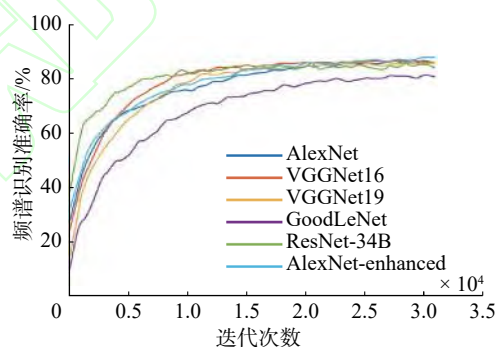


图8 各网络验证集频谱识别准确率变化曲线

Fig. 8 Spectrum recognition accuracy curve of each network in the validation set

表3 各网络验证集频谱识别准确率

Table 3 Spectrum recognition accuracy of each network in the validation set

网络	频谱识别准确率/%
AlexNet	86.26
VGGNet16	86.31
VGGNet19	86.11
GoogLeNet	81.18
ResNet-34B	84.67
AlexNet-enhanced	88.16

根据图8可以看出,各网络在迭代 31 000 次后,对验证集音乐的频谱识别准确率趋于稳定,且 AlexNet-enhanced 的准确率达到最高,表3数据表明, AlexNet-enhanced 的验证集频谱识别准确

率比常用 DCNN 高出 1.85%~6.98%。

AlexNet-enhanced 与常用 DCNN 以学习率 0.001 迭代 31 000 次的验证集频谱识别损失函数值变化曲线如图 9 所示。在迭代 31 000 次后的验证集频谱识别损失函数值见表 4。

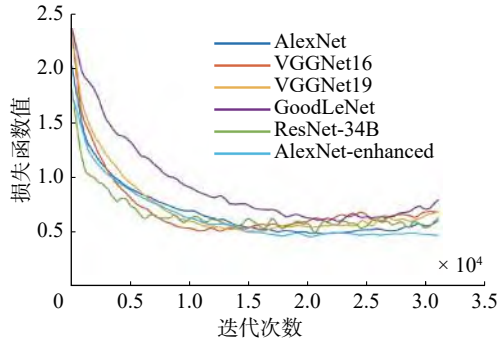


图 9 各网络验证集频谱识别损失函数值变化曲线
Fig. 9 Loss curve of each network in the validation set

表 4 各网络验证集频谱识别损失函数值
Table 4 Loss of each network in the validation set

网络	损失函数值
AlexNet	0.587 4
VGGNet16	0.700 8
VGGNet19	0.669 3
GoogLeNet	0.776 2
ResNet-34B	0.599 5
AlexNet-enhanced	0.465 3

根据图 9 可以看出,各网络在迭代 31 000 次后,对验证集音乐的频谱识别损失函数值趋于稳定,且 AlexNet-enhanced 的损失函数值降到了最低,表 4 数据表明, AlexNet-enhanced 的验证集频谱识别损失函数值比常用 DCNN 低 0.122 1~0.310 9。

对比实验结果表明, DCNN 在经过 31 000 次迭代之后,频谱识别准确率和频谱识别损失函数值均趋于稳定。 AlexNet-enhanced 的训练集频谱识别准确率和验证集频谱识别准确率分别达到了 99.43% 和 88.16%, 训练集频谱识别损失函数值和验证集频谱识别损失函数值分别降低到 0.073 4 和 0.465 3。由此可以看出, AlexNet-enhanced 具有更好的频谱识别性能。

3.4 机器学习模型对比实验

将基于 AlexNet-enhanced 的 DCNN-MGR 模型和基于 AlexNet、VGGNet16、VGGNet19、GoogLeNet、ResNet-34B 的 DCNN-MGR 模型与其他机器学习模型对 GTZAN 数据集的测试集音乐进行流派识别,识别准确率见表 5。

表 5 各模型测试集音乐识别准确率

Table 5 Music recognition accuracy of each model in the test set

模型	准确率/%
SVM ^[1]	73.00
KNN ^[2]	59.00
Gradient Boosting ^[4]	69.00
Extra Trees ^[4]	72.00
LSTM ^[7]	74.00
GRU ^[7]	75.00
DCNN-MGR(AlexNet)	76.00
DCNN-MGR(VGGNet16)	77.00
DCNN-MGR(VGGNet19)	76.00
DCNN-MGR(GoogLeNet)	70.00
DCNN-MGR(ResNet-34B)	72.00
DCNN-MGR(AlexNet-enhanced)	79.00

实验结果表明,基于 AlexNet-enhanced 的 DCNN-MGR 模型在 GTZAN 数据集中测试集上的音乐流派识别准确率达到 79.00%,比基于其他 DCNN 的 DCNN-MGR 模型高出 2.00%~9.00%;比其他机器学习模型高出 4.00%~20.00%。由此可以看出,相较于基于其他 DCNN 的 DCNN-MGR 模型,基于 AlexNet-enhanced 的 DCNN-MGR 模型具有更高的音乐流派识别准确率;相较于其他机器学习模型,基于 AlexNet-enhanced 的 DCNN-MGR 模型具有更高的音乐流派识别准确率。

4 结束语

本文针对机器学习模型对音乐流派特征识别能力较弱的问题,提出了基于增强 AlexNet 的音乐流派识别 DCNN-MGR 模型。一方面通过融合 Leaky ReLU 函数、Tanh 函数和 Softplus 分类器对 AlexNet 做出增强,并对 AlexNet 做出了部分网络结构调整,得到了频谱识别准确率更高,网络收敛效果更好的 DCNN;另一方面通过 FFT 提取音频信息,生成可以输入 DCNN 的频谱,并将频谱进行切割后输入 DCNN 进行深度学习,以此构建音乐流派识别模型。将本文模型与其他机器学习模型在 GTZAN 数据集上进行对比实验,实验结果表明,对于音乐流派的识别,基于 AlexNet-enhanced 的 DCNN-MGR 模型具有较高的准确率,证明了本文模型的有效性。本文模型在 GTZAN 数据集上取得了 79% 的音乐流派识别准确率,仍有较大的提高空间,下一步将充分发挥 DCNN 的特性以进一步提高音乐流派识别的准确率。

参考文献:

- [1] 邵曦,姚磊. 基于 SVM 主动学习的音乐分类[J]. 计算机工程与应用, 2016, 52(6): 127–133.
SHAO Xi, YAO Lei. Music classification based on SVM active learning[J]. Computer engineering and applications, 2016, 52(6): 127–133.
- [2] ALI M A, SIDDIQUI Z A. Automatic music genres classification using machine learning[J]. International journal of advanced computer science and applications, 2017, 8(8): 337–344.
- [3] TZANETAKIS G, COOK P. Musical genre classification of audio signals[J]. IEEE transactions on speech and audio processing, 2002, 10(5): 293–302.
- [4] MURAUER B, SPECHT G. Detecting music genre using extreme gradient boosting[C]//Companion of the The Web Conference 2018. Lyon, France, 2018: 1923–1927.
- [5] 焦李成, 杨淑媛, 刘芳, 等. 神经网络七十年:回顾与展望[J]. 计算机学报, 2016, 39(8): 1697–1716.
JIAO Licheng, YANG Shuyuan, LIU fang, et al. Seventy years beyond neural networks: retrospect and prospect[J]. Chinese journal of computers, 2016, 39(8): 1697–1716.
- [6] DAI Jia, LIANG Shan, XUE Wei, et al. Long short-term memory recurrent neural network based segment features for music genre classification[C]//2016 10th International Symposium on Chinese Spoken Language Processing. Tianjin, China, 2016: 1–5.
- [7] JAKUBIK J. Evaluation of gated recurrent neural networks in music classification tasks[C]//International Conference on Information Systems Architecture and Technology. Szklarska Poreba, Poland, 2017: 27–37.
- [8] 马世龙, 乌尼日其其格, 李小平. 大数据与深度学习综述[J]. 智能系统学报, 2016, 11(6): 728–742.
MA Shilong, WUNIRI Qiqige, LI Xiaoping. Deep learning with big data: state of the art and development[J]. CAAI transactions on intelligent systems, 2016, 11(6): 728–742.
- [9] 苗北辰, 郭为安, 汪镭, 等. 隐式特征和循环神经网络的多声部音乐生成系统[J]. 智能系统学报, 2019, 14(1): 158–164.
MIAO Beichen, GUO Weian, WANG Lei. A polyphony music generation system based on latent features and a recurrent neural network[J]. CAAI transactions on intelligent systems, 2019, 14(1): 158–164.
- [10] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[C]//Advances in neural Information Processing Systems. Lake Tahoe, USA, 2012: 1097–1105.
- [11] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[J]. Computer science, 2014: 1409–1556.
- [12] SZEGEDY C, LIU W, JIA Y, et al. Going deeper with convolutions[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Boston, USA, 2015: 1–9.
- [13] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA, 2016: 770–778.
- [14] 吴进, 钱雪忠. 紧凑型深度卷积神经网络在图像识别中的应用[J]. 计算机科学与探索, 2019, 13(2): 275–284.
WU Jin, QIAN Xuezhong. Compact deep convolutional neural network in image recognition[J]. Journal of frontiers of computer science and technology, 2019, 13(2): 275–284.
- [15] 吕鸿蒙, 赵地, 迟学斌. 基于增强 AlexNet 的深度学习阿尔茨海默病的早期诊断[J]. 计算机科学, 2017, 44(6): 50–60.
LYU Hongmeng, ZHAO Di, CHI Xuebin. Deep learning for early diagnosis of Alzheimer's disease based on intensive AlexNet[J]. Computer science, 2017, 44(6): 50–60.
- [16] 甘岚, 郭子涵, 王瑶. 基于径向变换和改进 AlexNet 的胃肿瘤细胞图像识别方法[J]. 计算机应用, 2019, 39(10): 2923–2929.
GAN Lan, GUO Zihan, Wang Yao. Gastric tumor cell image recognition method based on radial transformation and improved AlexNet[J]. Journal of computer applications, 2019, 39(10): 2923–2929.
- [17] 陈思文, 刘玉江, 刘冬, 等. 基于 AlexNet 模型和自适应对比度增强的乳腺结节超声图像分类[J]. 计算机科学, 2019, 46(6): 146–152.
CHEN Siwen, LIU Yujiang, LIU Dong, et al. AlexNet model and adaptive contrast enhancement based ultrasound imaging classification[J]. Computer science, 2019, 46(6): 146–152.
- [18] 王文秀, 傅雨田, 董峰, 等. 基于深度卷积神经网络的红外船只目标检测方法[J]. 光学学报, 2018, 38(7): 160–166.
WANG Wenxiu, FU Yutian, DONG Feng, et al. Infrared ship target detection method based on deep convolution neural network[J]. Acta optica sinica, 2018, 38(7): 160–166.
- [19] 李祥鹏, 闵卫东, 韩清, 等. 基于深度学习的车牌定位和识别方法[J]. 计算机辅助设计与图形学学报, 2019, 31(6): 979–987.
LI Xiangpeng, MIN Weidong, HAN Qing, et al. License plate location and recognition based on deep learning[J]. Journal of computer-aided design & computer graphics,

- 2019, 31(6): 979–987.
- [20] 赵远东, 刘振宇, 柯丽, 等. 人脸识别中 AlexNet 网络设计和改进方法研究 [J]. 通信技术, 2019, 52(3): 592–598.
ZHAO Yuandong, LIU Zhenyu, Ke li, et al. Alexnet network design and improvement methods in face recognition[J]. Communications technology, 2019, 52(3): 592–598.
- [21] 盖杉, 鲍中运. 基于改进深度卷积神经网络的纸币识别研究 [J]. 电子与信息学报, 2019, 41(8): 1993–2000.
GAI Shan, BAO Zhongyun. Banknote recognition research based on improved deep convolutional neural network[J]. Journal of electronics and information technology, 2019, 41(8): 1993–2000.
- [22] ZHENG Hao, YANG Zhanlei, LIU Wenju, et al. Improving deep neural networks using softplus units[C]//2015 International Joint Conference on Neural Networks. Killarney, Ireland, 2015: 1–4.
- [23] 赵慧珍, 刘付显, 李龙跃. 一种新的深度卷积神经网络的 SLU 函数 [J]. 哈尔滨工业大学学报, 2018, 50(4): 117–123.
ZHAO Huizhen, LIU Fuxian, LI Longyue. A novel soft-plus linear unit for deep CNN[J]. Journal of Harbin institute of technology, 2018, 50(4): 117–123.
- [24] ELBIR A, İLHAN H O, SERBES G, et al. Short time Fourier transform based music genre classification[C]//2018 Electric Electronics, Computer Science, Biomedical Engineerings' Meeting. Istanbul, Turkey, 2018: 1–4.
- [25] 毛勇华, 桂小林, 李前, 等. 深度学习技术应用研究 [J]. 计算机应用研究, 2016, 33(11): 3201–3205.
MAO Yonghua, GUI Xiaolin, LI Qian, et al. Study on application technology of deep learning[J]. Application research of computers, 2016, 33(11): 3201–3205.

作者简介:



刘万军, 教授, 主要研究方向为数字图像处理、运动目标检测与跟踪。主持国家级和省部级科研项目 20 余项。发表学术论文 120 余篇。



孟仁杰, 硕士研究生, 主要研究方向为深度学习、自然语言处理。



曲海成, 副教授, 主要研究方向为高光谱遥感图像处理、GPU 并行计算。主持辽宁省科技厅和教育厅一般项目各 1 项, 参与国家自然科学基金项目 2 项。发表学术论文 30 余篇。