

# 基于肌音信号图像化和卷积神经网络的手语识别研究

王新平 夏春明 颜建军

华东理工大学机械与动力工程学院 上海 200237

(showtimewalker@163.com)

**摘要** 时间序列信号被广泛应用于各种模式识别的场合,针对大量目标的时间序列信号模式识别率低的问题,借助多种图像化手段,将时间序列信号转换为图像,采用图像分类算法实现模式识别。实验中采集了前臂上 30 种手语对应的肌音信号(Mechanomyography,MMG),将其转换为不同风格的图像,设计卷积神经网络(Convolution Neural Network,CNN)框架,对图像化的肌音信号训练集建立模式识别的分类模型,并且应用迁移学习(transfer learning)算法对模型进行多次优化,建立的分类模型识别率达 98.7%,高于普通机器学习算法的识别率。实验结果证明了图像化处理时间序列信号可以有效提高多分类肌音信号模式识别的识别率,该研究可以为其他时间序列信号的模式识别研究提供参考。

**关键词**:肌音信号;模式识别;时间序列图像化;卷积神经网络;迁移学习

中图法分类号 TN911.7

## Sign Language Recognition Based on Image-interpreted Mechanomyography and Convolution Neural Network

WANG Xin-ping, XIA Chun-ming and YAN Jian-jun

School of Machinery and Power Engineering, East China University of Science and Technology, Shanghai 200237, China

**Abstract** Time series signals are widely used in various pattern recognition applications. In order to solve the problem of low pattern recognition rate of time series signals for a large number of targets, this article uses a variety of transform methods to convert time series signals into images, and performs pattern recognition using image classification algorithms. In the experiment, the mechanomyography (MMG) corresponding to 30 sign languages on the forearm are collected and converted into different image styles, and a convolution neural network (CNN) framework is designed to establish pattern recognition classification models for the images. The models are optimized twice with the application of transfer learning algorithm, and the recognition rate of the best classification model reaches 98.7%, which is much higher than the recognition rate of traditional machine learning algorithms. The experimental results imply that the image processing of time series signals can effectively improve the recognition rate of multi-target pattern recognition of MMG. This paper can provide references for pattern recognition of other time series signal.

**Keywords** Mechanomyography, Pattern recognition, Image-interpreted time series signal, Convolution neural network, Transfer learning

### 1 引言

手语在人们的生活中应用广泛,是人与人之间交流不可或缺的方式之一,同时也是聋哑人之间以及聋哑人与正常人之间交流的主要途径。由于手语表达形式较为抽象,普通人识别手语较为困难,近年来有大量关于手语识别的研究,通过机器学习建立分类模型,自动识别手语传递的数字信号并作出判断。手语识别需要依赖手语动作发出的信号,这种信号以多种形式存在,如表面肌电信号<sup>[1]</sup>(surface Electromyography, sEMG)、加速度信号<sup>[2]</sup>、姿态信号<sup>[3]</sup>、图像<sup>[4]</sup>、视频<sup>[5]</sup>等,前三者属于一维时间序列信号,第四种和第五种分别为二维、三维图像信号。时间序列信号包含的信息较少,一般用于少量模式的识别,当识别模式种类增多时,很难获得较高的识别效果;图像/视频包含更多的有用信息,在多目标的模式识别中表现更出色,但由于图像数据采集需要借助摄像头等设备,

采集较为困难,并且直接获得的图片数据量庞大,难以在手语识别中普及。

本文结合时间序列信号采集方便以及图像信号信息丰富的特点,设计加速度采集设备采集前臂 3 块肌肉的 6 通道肌音信号,应用时间序列图像化处理算法将其转化为 RGB 图像,并结合深度卷积神经网络建立模式识别分类器,实现了对 30 种动态手语的识别,获得的综合识别率高达 98.7%,超过了普通机器学习算法识别特征向量的效果,为时间序列的模式识别提供了一种新思路。

肌音信号是一种由于肌肉纤维收缩而产生的位于皮肤表面的振动信号,可以反映肌肉的动态特性——当肌肉发生收缩时,肌肉纤维的几何形状发生改变,肌肉之间相互摩擦,产生微小的振动信号,经过细胞膜的传递,穿过肌肉组织到达皮肤表面,形成频率为 2~150 Hz 之间的机械振动信号,通过加速度传感器或者麦克风等硬件可对其进行捕捉<sup>[6]</sup>。

MMG作为一种生物医学信号,可以被应用于动作分类<sup>[7]</sup>、假肢控制<sup>[8]</sup>、肌肉疲劳分析<sup>[9]</sup>等领域。MMG和sEMG有很多相似的地方,分别应用于适合自己的领域中,在某些研究中也同时出现。对比MMG和sEMG,前者具有更高的信噪比,采集成本低,采集方式简单,并且采集设备一般可以间隔衣服、布料等工作;后者是一种肌肉电位差信号,对于采集的环境要求较高,可以获得较为专业的生物医学信号,在医学领域的应用较多,两者在模式识别中均得到了广泛应用。Yang等<sup>[10]</sup>采用卷积神经网络算法直接作用于多通道表面肌电信号,实现了对4种手势动作的分类识别;Ding等<sup>[11]</sup>结合肌音信号和表面肌电信号,采用多种降维算法进行处理,并结合最接近分类等算法实现了11种抓取动作的预测,识别率达95%。Ding等<sup>[11]</sup>根据不同手指的动作,采集了前臂上的肌音信号,并且利用小波包变换等提取数据的特征,通过支持向量机(Support Vector Machine, SVM)算法建立分类模型,实现了多种手指动作的模式识别。Alves等<sup>[12]</sup>采集了前臂多通道的肌音信号,提取了时域、频域上的14种特征,并采用线性判别分类(Linear Discriminate Analysis, LDA)分别对5种和8种手掌动作模式进行了识别,获得的平均识别率为98%和90%。

根据以上调研总结:对于生物医学信号时间序列的模式识别,已有大量相关研究,基本实验流程如下:

(1)采集执行动作时牵动的肌肉群对应皮肤表面的MMG, sEMG。

(2)数据预处理,提取时域、频域,以及小波等特征。

(3)采用SVM, LDA等机器学习算法实现模式识别。

此类研究在少量目标的模式识别中表现较好,但是随着动作识别种类数量增多,识别率随之下降,因此不适合多目标的模式识别。针对这一状况,本文借鉴图像识别中深度学习的模式识别算法,通过设计时间序列信号转图像算法,将采集的多通道MMG转换为图像,并对其建立深度卷积神经网络,利用迁移学习算法对模型进行优化,最终获得的分类器的动作分类效果优于传统的机器学习算法。

## 2 算法设计

设计如图1所示的算法执行流程:原始的MMG经过图像化算法处理后,获得多种不同风格的图像,分别对这些图像建立卷积神经网络分类模型,测试不同分类模型的分类效果。

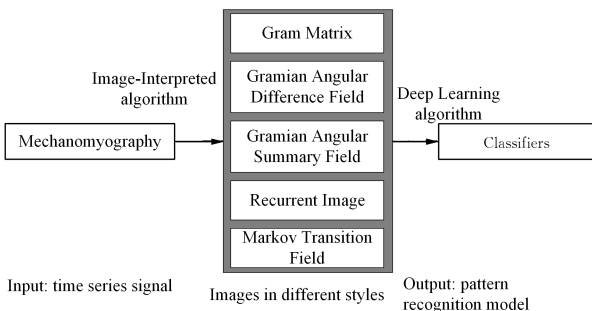


图1 算法运行流程图

Fig. 1 Algorithm execute process

因此算法的设计主要包括时间序列信号的图像化算法以及深度学习算法。

### 2.1 时间序列信号图像化

时间序列信号的图像化可以通过多种方法实现,本文分

别采用格拉姆矩阵(Gram Matrix, GM)、格兰角场差分域(Gramian Angular Difference Field, GADF)<sup>[13]</sup>、格兰角场和域(Gramian Angular Summary Field, GASF)<sup>[13]</sup>、递归图(Recurrent Image, REC)<sup>[14]</sup>和马尔可夫转换域(Markov Transition Field, MTF)<sup>[13]</sup>对肌音信号波段进行图像化转换,获得多种风格的预处理图像数据集。

#### 2.1.1 格拉姆矩阵

格拉姆矩阵是时间序列信号图像化最普遍的处理方法之一,对于一段长度为 $n$ 的时间序列,该算法的处理方式为:

$$X = [x_1, x_2, \dots, x_n]$$

$$\langle x_1 \cdot x_2 \rangle = x_1 \times x_2$$

$$GM = \begin{bmatrix} \langle x_1 \cdot x_1 \rangle & \dots & \langle x_1 \cdot x_n \rangle \\ \vdots & \ddots & \vdots \\ \langle x_n \cdot x_1 \rangle & \dots & \langle x_n \cdot x_n \rangle \end{bmatrix} \quad (1)$$

本文一共采集了30种不同手语对应的6通道肌音信号,运行分割算法后,将每个动作样本划分为长度为300的时间序列波段,经过Gram Matrix算法转换后,获得相应的 $300 \times 300$ 的图像矩阵。由于图像一般只包含RGB三颜色通道,因此,在进行图像化处理前,对6通道数据进行以下预处理:

$$\begin{cases} I_{1i} = \sqrt{CH_{1i}^2 + CH_{2i}^2} \\ I_{2i} = \sqrt{CH_{3i}^2 + CH_{4i}^2}, 1 \leq i \leq 300 \\ I_{3i} = \sqrt{CH_{5i}^2 + CH_{6i}^2} \end{cases} \quad (2)$$

其中, $CH_1 - CH_6$ 为采集的6通道加速度信号,分别由3个加速度传感器进行采集,每个传感器采集两个通道,代表3块前臂肌肉上X, Y方向的振动信号; $I_1 - I_3$ 为Gram Matrix算法转换后的单色灰度图,分别对应R, G, B 3个图像通道。图2为转换风格后其中4种不同动作对应的图像。

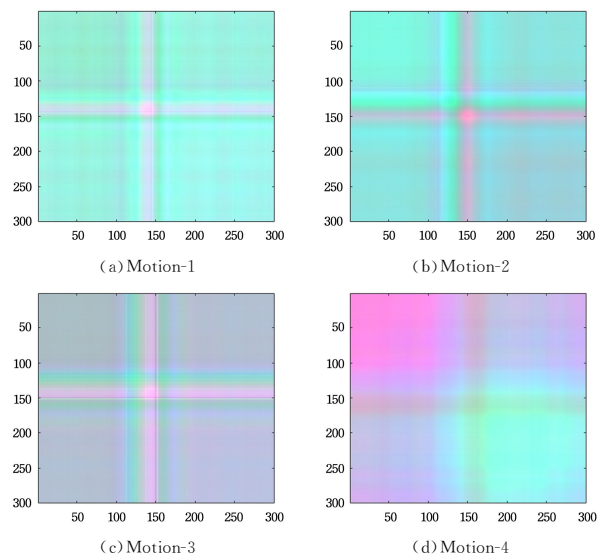


图2 4种手语动作对应的图像化肌音信号

Fig. 2 Four kinds of gesutres' image-interpreted mechanomyography

#### 2.1.2 格兰角转换场

Gram Matrix算法的处理计算位于直角坐标系,因此无法包含时间位置信息。在格兰角场转换算法中,对于一定长度的时间序列信号 $X = [x_1, x_2, \dots, x_n]$ ,先对序列 $X$ 进行归一化预处理,使所有元素等比例映射到 $[-1, 1]$ 区间。

$$\hat{x}_i = \frac{(x_i - \max(X)) + (x_i - \min(X))}{\max(X) - \min(X)} \quad (3)$$

$$\hat{X} = [\hat{x}_1, \hat{x}_2, \dots, \hat{x}_n], 0 < \hat{x}_i < 1$$

之后,将该序列映射到极坐标系下,利用反三角函数获得每个序列点对应的角度:

$$\begin{aligned} \overset{\vee}{x}_i &= \arccos(\overset{\wedge}{x}_i), -1 \leq \overset{\wedge}{x}_i \leq 1 \\ \overset{\vee}{X} &= [\overset{\vee}{x}_1, \overset{\vee}{x}_2, \dots, \overset{\vee}{x}_n], 0 < \overset{\vee}{x}_1 < \pi \end{aligned} \quad (4)$$

图 3 给出了从直角坐标系映射为极坐标系的过程。

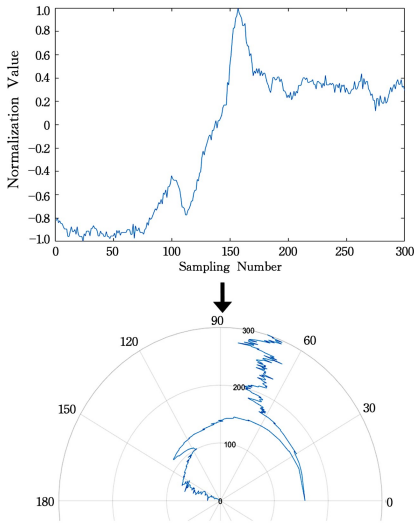


图 3 时间序列信号由直角坐标系过渡至极坐标系

Fig. 3 Time series signal transitions from cartesian coordinate to polar coordinate

在极坐标上,数据的波动以另一种方式展现,相比在直角坐标系中,其可以提供更丰富的动态特性,具体表现为:数据的变化值不仅仅体现于每个数据点的角度值,而且与采样时间和位置有关。在普通的直角坐标系中,  $\Delta x = x_{i+1} - x_i$ , 数据的波动只体现于一维空间;在极坐标系中,  $\Delta s = r_{i+1} \cdot \theta_{i+1} - r_i \cdot \theta_i$ , 数据变化不仅与角度振动有关,而且还与采样点的位置有关。因此可以同时保留时间和空间信息,数据的波动体现于二维空间,类似于水纹的波动。

根据肌音信号在极坐标系下的表现,建立 GASF 和 GADF 如下:

$$\begin{aligned} GASF &= \begin{bmatrix} \cos(\overset{\vee}{x}_1 + \overset{\vee}{x}_1) & \dots & \cos(\overset{\vee}{x}_1 + \overset{\vee}{x}_n) \\ \vdots & \ddots & \vdots \\ \cos(\overset{\vee}{x}_n + \overset{\vee}{x}_1) & \dots & \cos(\overset{\vee}{x}_n + \overset{\vee}{x}_n) \end{bmatrix} \\ &= \overset{\wedge}{X}' \cdot \overset{\wedge}{X} - \sqrt{I - \overset{\wedge}{X}^2}' \cdot \sqrt{I - \overset{\wedge}{X}^2} \\ GADF &= \begin{bmatrix} \cos(\overset{\vee}{x}_1 - \overset{\vee}{x}_1) & \dots & \cos(\overset{\vee}{x}_1 - \overset{\vee}{x}_n) \\ \vdots & \ddots & \vdots \\ \cos(\overset{\vee}{x}_n - \overset{\vee}{x}_1) & \dots & \cos(\overset{\vee}{x}_n - \overset{\vee}{x}_n) \end{bmatrix} \\ &= \overset{\wedge}{X}' \cdot \sqrt{I - \overset{\wedge}{X}^2} - \overset{\wedge}{X} \cdot \sqrt{I - \overset{\wedge}{X}^2}' \end{aligned} \quad (5)$$

将该算法分别应用于 6 通道数据,获得的 GADF/GASF 风格化图像如图 4 所示。

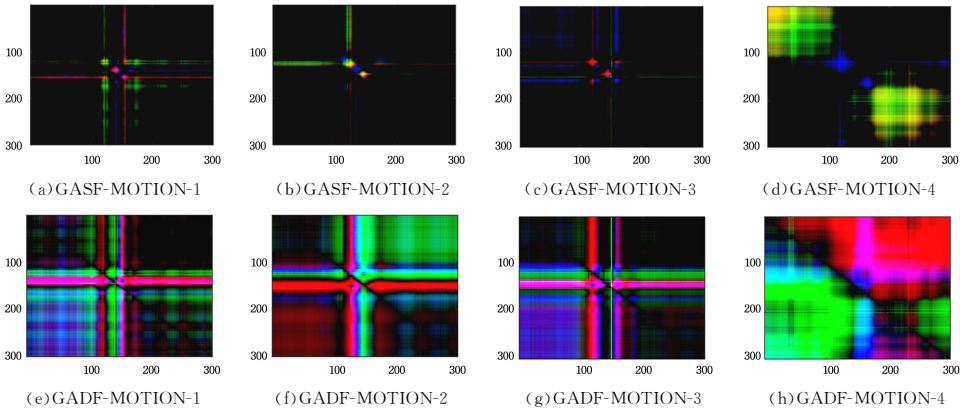


图 4 GASF 和 GADF 转换后的肌音信号图形

Fig. 4 Image-interpreted mechanomyography transformed with GASF and GADF

2.1.3 递归图

递归图可以表现时间序列信号非线性动力学特性,是反映信号内部结构的有效工具<sup>[14]</sup>。肌音信号经递归图算法的处理结果如图 5 所示。该方法的主要步骤是对信号的相空间重构,对于一段长度为  $n$  的序列  $X$ ,选择合适的时间长度  $t$ ,计算不同维度方向上重构向量的差分二阶范数,具体表现为:

$$\begin{aligned} \vec{x}_1 &= [x_1, x_2, \dots, x_t] \\ \vec{x}_2 &= [x_2, x_3, \dots, x_{t+1}] \\ &\vdots \\ \vec{x}_{n-t+1} &= [x_{n-t+1}, x_{n-t+2}, \dots, x_n] \end{aligned} \quad (6)$$

$$REC = \begin{bmatrix} \|\vec{x}_1 - \vec{x}_1\| & \dots & \|\vec{x}_1 - \vec{x}_{n-t+1}\| \\ \vdots & \ddots & \vdots \\ \|\vec{x}_{n-t+1} - \vec{x}_1\| & \dots & \|\vec{x}_{n-t+1} - \vec{x}_{n-t+1}\| \end{bmatrix}$$

$$\|\vec{x}\| = \sqrt{\sum_{i=1}^t |x_i|^2}$$

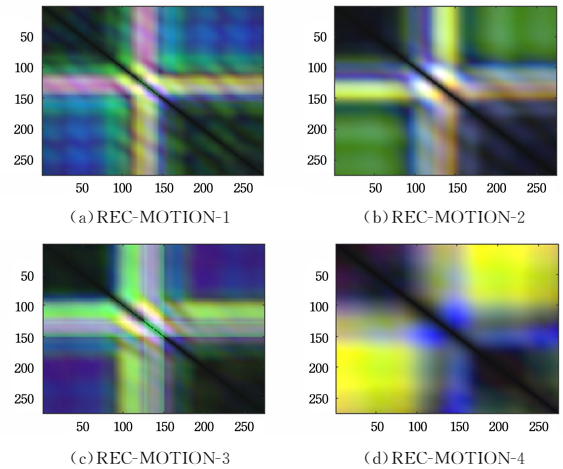


图 5 肌音信号递归图

Fig. 5 Mmechanomyography transferred with recurrent image



如图 8 所示,在该神经网络框架中,保留了 VGG-16 中卷积处理和池化操作的运算,在深度上削减了 7 层卷积运算,在不影响模型表现的情况下压缩模型体积,减少了训练成本,构成了深度为 9 的卷积神经网络 CNN-9,其中包括 5 层卷积层(包含池化运算)、3 层全连接层和 1 层输出层。

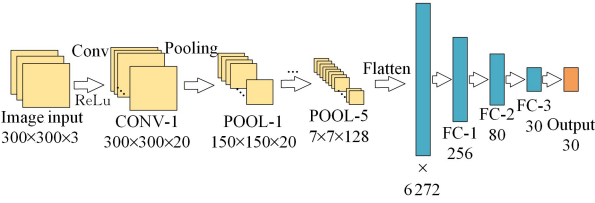


图 8 CNN-9 框架

Fig. 8 Framework of CNN-9

CNN 模型的超参数非常丰富,在实际运行时,设置模型训练的初始学习因子为 0.0001,采用分段式下降法,每 5 个周期减少为原来的 1/2;训练过程中采用规模为 150 的 mini-batch 进行小批量训练,提高了训练速度;模型的代价函数基于交叉熵,结合采用 L2 正则法对模型的权重进行调整,减少过拟合现象;训练时采用 Adam 优化方法,提高了模型的训练效率。

### 2.2.2 迁移学习

迁移学习(transfer learning)在深度学习中应用广泛<sup>[20]</sup>,尤其体现在 CNN 算法中<sup>[21]</sup>,通过将训练后的模型参数迁移至新的待训练模型中,更改输出层及其附近层的结构,训练新的深度神经网络。这种训练算法可以保留上一次训练中可以稳定识别的特征(如车辆识别中的轮子、人脸识别中的五官),因此可以提高第二次训练的效率。迁移学习可以极大提升模型参数的训练效率,缩短模型的训练周期。

在模型完成第一次训练后,将参数导入初始的网络框架中,实现参数的迁移,调整模型的训练超参数,重新训练,完成迁移学习。参数迁移后改变训练策略,重新进行训练,可以在保留上一次训练效果的同时,识别出更多的特征,从而提高模型的训练效果。

卷积神经网络的迁移学习如图 9 所示。

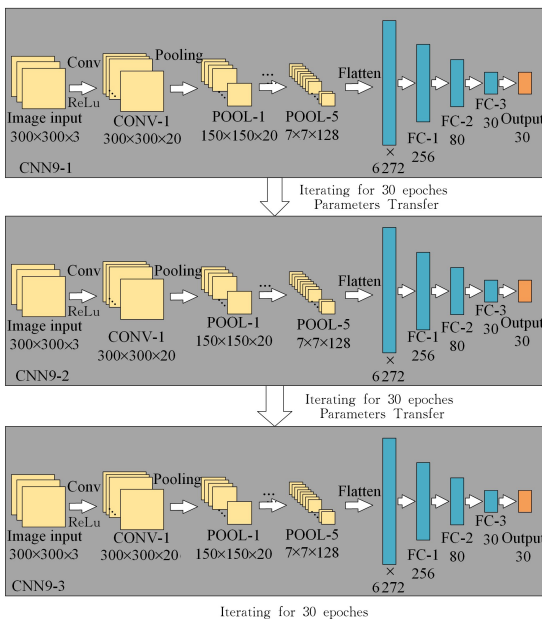


图 9 卷积神经网络的迁移学习

Fig. 9 Transfer learning of convolution neural network

确定模型框架后,输入图像数据,进行 30 次迭代计算,完成第一次模型的训练,此时模型的验证集识别率较低,存在较严重的过拟合现象,增加迭代次数无法提升识别率。用第一次模型训练后的参数代替随机参数,导入原来的模型框架中,调整正则化参数、初始学习率、Adam 参数等,进行第二次训练,此时模型已经具有一定的特征识别能力。在第二次训练过程中可以识别更复杂的特征,提高了验证集,减少了过拟合现象,重复迁移两次,完成最终的模型训练。

## 3 实验

### 3.1 数据采集装置

为了稳定采集多路通道的 MMG,设计并制作如图 10 所示的设备。由于 MMG 可以间隔衣服、表带等进行采集,将采集装置(见图 11)佩戴于前臂。采集工作由 3 个 ADXL213 双轴加速度计完成,将采集的 6 通道 MMG 以 PWM 方式传输至 STM32H743 高性能微处理器,并且在微处理器上同步运行分割算法,将分割后的动作波段通过 WIFI 传输至 PC 平台,由 MATLAB 创建的虚拟机接收,完成数据的收集。

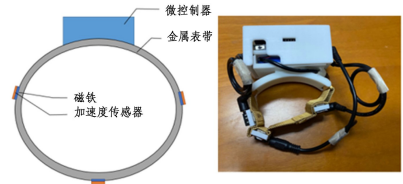


图 10 肌音信号采集装置

Fig. 10 MMG acquisition device



图 11 嵌入式肌音信号采集设备

Fig. 11 Embedded system of mechanomyography collected device

### 3.2 实验对象和内容

有 6 名志愿者参与本课题的数据采集,其中包括 4 名男性和 2 名女性,年龄在 22~26 之间。志愿者提前知晓实验内容,并且签署了安全责任协议。实验过程中,将设备佩戴于前臂肘关节位置下方,传感器通过磁铁吸附于金属表带,调节传感器的位置,使每个传感器分别对应前臂 3 块肌肉的皮肤表面。采集每个志愿者(见图 12)的 30 种手语动作的肌音信号,其分别代表 30 种日常手语交流信息,每个动作重复采集 100 次。



图 12 30 种手语动作

Fig. 12 30 kinds sign language gestures

### 3.3 数据集构造

肌音信号的能量主要集中于 150 Hz 以下,根据奈奎斯特采样定律,在采集过程中调整系统的采样频率为 333 Hz,在分割算法运行时,根据动作的能量波动性,当 6 通道的肌音信号能量值总和超过某一阈值时,记下标记,延长一定时间,取时间序列末尾 300 个采样单位作为该动作的肌音信号波形,完成动作的分割,即每个动作的持续时间为 0.9 s。

每个动作重复 100 次,实验分多次进行,每名志愿者拥有 3000 个数据样本。采用如前所述的多种图像化转换算法,将样本转化成多种图像信号,按照顺序给所有样本设定标签,在组内随机打乱顺序,保证数据的分类不受采样顺序影响。将样本按照 60%,30%,10%的比例划分成训练集、验证集和测试集,采用 hold-out 法在训练中对模型进行交叉验证,模型训练完成后,对测试集样本进行预测分类,统计 6 个志愿者模型的平均识别率的无偏估计。

为了对比图像化风格和特征集的分类效果,设置对照实验,对每个样本提取时域、频域特征,并进行归一化处理,组成特征矩阵,作为训练对象,结合迁移学习算法,分别运行 BPNN 和 CNN,统计模型的识别率。

特征集包含 12 种时域特征和 4 种频域特征。时域特征:1-绝对平均值;2-标准差;3-方差;4-能量值;5-均方根;6-四分位差;7-偏度;8-峰度;9-中心距;10-绝对平均离差;11-中位绝对差;12-混乱熵;频域特征:13-第一频率位置;14-最大频谱能量值;15-中值频率;16-平均能量频率。

### 4 结果与分析

本文算法运行于 MATLAB 2020a,CPU 主频为 2.2 GHz, GPU 采用 NVIDIA GeForce GTX1070,内存为 8 GB,用于加速图像处理。根据基于 MMG 构造的多种图像信号以及特征图谱,运行 CNN 和迁移学习算法,将获得的综合识别率、混

淆矩阵、代价函数等作为模型评估标准。

#### 4.1 神经网络模型的表现

实验中将每个样本分别转化成 GM,GADF,GASF,MTF 和 REC 风格的图像,其中前四者转换后的图像尺寸均为  $300 \times 300 \times 3$ ,可以直接用作卷积神经网络的模型输入,递归图的图像大小与延时参数有关,本文获得的递归图尺寸为  $285 \times 285 \times 3$ ,图像训练之前进行尺寸等比例调整。训练后获得的模型表现如图 13 所示。

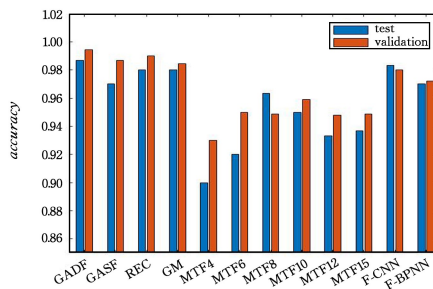


图 13 肌音信号转换成不同风格图片后的模型识别率

Fig. 13 Accuracy of the classification model with different image styles

MTF 后面的数字代表不同的 Q 值,F-CNN 和 F-BPNN 分别代表提取特征后采用卷积神经网络和反向传播神经网络算法训练的结果。由结果可知,将 MMG 转换为不同风格的图像后,获得的识别结果也不同,其中获得的 GADF 风格的图像表现最佳,综合识别率高达 98.67%。马尔可夫概率转移域图表现较差,不同数量的子值域下模型的表现不同,当 Q 的值为 10 左右时,模型表现最优。对于特征提取的方式,应用 CNN 建立的模型表现比 BPNN 更好,其测试集识别率可达 98.33%。

图 14 给出了 GADF 风格图像的分类混淆矩阵。

	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12	M13	M14	M15	M16	M17	M18	M19	M20	M21	M22	M23	M24	M25	M26	M27	M28	M29	M30	accuracy
M1	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	
M2	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	
M3	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	
M4	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	
M5	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	
M6	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	
M7	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	
M8	0	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	
M9	0	0	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	
M10	0	0	0	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	
M11	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	
M12	0	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	
M13	0	0	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	
M14	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	
M15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	
M16	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	8	0	0	0	0	0	0	0	0	0	0	0	0	80%	
M17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0	100%	
M18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	100%	
M19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	100%	
M20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0	100%	
M21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0	100%	
M22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	9	0	0	0	0	0	0	0	90%	
M23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0	0	0	0	100%	
M24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0	0	0	100%	
M25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0	0	100%	
M26	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0	100%	
M27	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	9	0	0	0	90%	
M28	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0	0	100%	
M29	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0	100%	
M30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	100%	

图 14 GADF 图像的识别率混淆矩阵

Fig. 14 Confusion matrix of GADF style mechanomyography

采用迁移学习可以大大提高模型的学习效率,在实验中,通过将前一识别的参数导入下次训练的模型框架中,再改变训练参数,可以使模型突破一些之前难以识别的特征障碍,从而达到提升识别率和降低过拟合的目的。由图 15 可知,在模

型的第一次训练中,随着迭代次数的增加,识别率和损失函数分别提高和减小,但逐渐趋近一个阈值,难以突破。迁移学习后调试训练参数,再次进行 CNN 训练时可以识别一些在第一次训练中难以识别的特征,使模型得到优化。

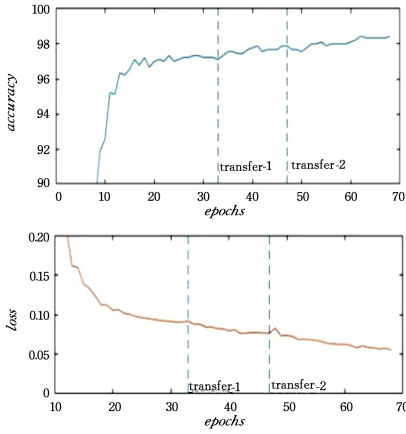


图 15 迁移学习后模型表现的变化曲线

Fig. 15 Curve after transfer learning of model performance

图 16 中, *accuracy-1*, *accuracy-2*, *accuracy-3* 分别为原始神经网络模型的识别率、一次迁移学习和二次迁移学习的识别率。实验结果证明, 迁移学习可以有效提高卷积神

经网络算法的识别率。

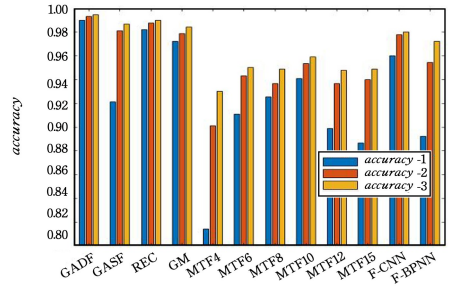


图 16 迁移学习提升模型的识别率

Fig. 16 Transfer learning can improve accuracy of the model

4.2 过拟合分析

本实验中, 卷积神经网络模型交叉验证采用的是 hold-out 法, 因此可能存在模型的过拟合现象, 在算法中使用正则化矫正、迁移学习以及采用优化学习法等可以在一定程度上降低模型的过拟合现象。表 1 列出了迁移学习前后三次模型训练后的手语识别精度。

表 1 过拟合分析

Table 1 Overfitting analysis

(单位: %)

	<i>test-1</i>	<i>validation-1</i>	<i>test-2</i>	<i>validation-2</i>	<i>test-3</i>	<i>validation-3</i>
GADF	98.3	99.0	98.7	99.3	98.7	99.4
GASF	96.3	92.1	97.0	98.1	97.0	98.7
REC	97.3	98.2	98.3	98.8	98.0	99.0
GM	95.3	97.2	96.3	97.9	98.0	98.4
MTF4	81.0	81.4	88.0	90.1	90.0	93.0
MTF6	86.3	91.1	91.7	94.3	92.0	95.0
MTF8	93.3	92.6	95.7	93.7	96.3	94.9
MTF10	93.0	94.1	94.0	95.3	95.0	95.9
MTF12	87.7	89.9	91.3	93.7	93.3	94.8
MTF15	86.7	88.7	92.7	94.0	93.7	94.9
F-CNN	96.0	96.0	98.3	97.8	98.3	98.0
F-BPNN	90.0	89.2	95.0	95.4	97.0	97.2

*test* 结果为模型对测试集的识别率, 为无偏估计, *validation* 结果为模型对验证集的自测表现, 当 *validation* 值大于 *test* 值时, 表示模型可能存在一定的过拟合, 差距越大, 模型的过拟合现象越严重。

由表 1 可知, 大部分模型在完成第一次训练后, 测试集的识别率和验证集相差较大, 经过迁移学习之后, 两者的差值明显缩小, 证明了通过迁移学习可以消除部分过拟合情况, 同时也大大提高了模型的表现效果。

4.3 卷积神经网络模型与机器学习模型表现的对比

为了验证卷积神经网络对图像识别效果的优越性, 建立对照组, 将所有样本以 60%、30%、10% 的比例分成训练集、验证集和测试集, 分别采用决策树 (Decision Tree, DT)、最近邻分类器 (*k*-Nearest Neighbors, KNN)、线性判别分析 (Linear Discriminant Analysis, LDA)、朴素贝叶斯 (Naive Bayes, NB)、支持向量机 (Vector Support Machine, SVM) 以及随机子空间判别 (Random Subspace, RS) 对特征数据集进行建模, 同时对 6 个人的测试集进行模式识别, 获得的综合平均识别率如表 2 所列。

表 2 不同分类器表现的对比

Table 2 Model Performance of different classifiers

Classifier	dataset		Accuracy of test set / %
	Sample type	Sample number	
CNN	Image	3 000	98.67
DT	Feature set	3 000	84.33
KNN	Feature set	3 000	96.33
LDA	Feature set	3 000	93.78
NB	Feature set	3 000	95.33
SVM	Feature set	3 000	94.67
RS	Feature set	3 000	96.00

由表 2 可得, 采用图像化处理 MMG 结合卷积神经网络算法获得的综合识别率远高于其他机器学习算法对特征向量模式识别的结果, 证明了该方法的科学性和先进性。

**结束语** 本文提出了将肌音信号时间序列信号转换成 5 种不同风格的图片信号, 并构建卷积神经网络框架, 实现了 30 种不同手语对应肌音信号波形的分类。实验结果表明, 将时间序列转换为图像不仅可以避免手工提取时频域特征的步骤, 而且可以获得更优的识别效果。

研究中尚存在训练模型过拟合的现象, 导致训练测试结

果与实际应用效果存在一定偏差,另外,该分类模型仅针对单个个体而建立,不同个体之间的模型不具有通用性。因此,后续的研究工作主要是建立更科学的深度学习框架和研发其他图像处理、图像分类算法,以消除模型过拟合现象和提高模型的普适性。

本文提供了一种模式识别应用中处理时间序列的新思路,并证明了该方法的科学性和先进性,可为其他领域的时间序列信号模式识别提供参考和借鉴。

## 参 考 文 献

- [1] ZHANG C Z, GU X T, ZHANG Y M. Gesture Recognition Based on Deep Convolutional Neural Network[J]. *Radio Engineering*, 2019, 49(7): 587-591.
- [2] KAKOTY N M, SHARMA M D. Recognition of Sign Language Alphabets and Numbers based on Hand Kinematics using a Data Glove[J]. *Procedia Computer Science*, 2018, 133: 55-62.
- [3] SURI K, GUPTA R. Continuous sign language recognition from wearable IMUs using deep capsule networks and game theory [J]. *Computers & Electrical Engineering*, 2019, 78: 493-503.
- [4] LI L X, WANG Y, WU J J, et al. Micro-motion hand gesture recognition method based on improved multiple dimensional convolution neural network[J]. *Computer Engineering*, 2018, 44(9): 243-249.
- [5] AZAR S G, SEYEDARABI H. Trajectory-based recognition of dynamic Persian sign language using hidden Markov model[J]. *Computer Speech & Language*, 2020, 61: 101053.
- [6] WOLCZOWSKI A, ZDUNEK R. Electromyography and mechanomyography signal recognition: Experimental analysis using multi-way array decomposition methods[J]. *Biocybernetics and Biomedical Engineering*, 2017, 37(1): 103-113.
- [7] ZHANG Y, XIA C. A preliminary study of classification of upper limb motions and forces based on mechanomyography[J/OL]. *Medical Engineering & Physics*. [https://www.researchgate.net/publication/341471100\\_A\\_preliminary\\_study\\_of\\_classification\\_of\\_upper\\_limb\\_motions\\_and\\_forces\\_based\\_on\\_mechanomyography](https://www.researchgate.net/publication/341471100_A_preliminary_study_of_classification_of_upper_limb_motions_and_forces_based_on_mechanomyography).
- [8] XIE H B, ZHENG Y P, GUO J Y. Classification of the mechanomyogram signal using a wavelet packet transform and singular value decomposition for multifunction prosthesis control[J]. *Physiological Measurement*, 2009, 30(5): 441-457.
- [9] JIANG W D, XIA C M, ZHANG Y, et al. Static Fatigue Analysis of Upper Trapezius Muscle Based on Fractal Dimension of Mechanomyography[J]. *Space Medicine & Medicine Engineering*, 2019(3): 259-264.
- [10] YANG Y H, XIE H. Research on surface EMG signal gesture recognition based on convolution neural network [J/OL]. *Microcomputer and Applications*. [http://en.cnki.com.cn/Article\\_en/CJFDTotat-WXJY201715017.htm](http://en.cnki.com.cn/Article_en/CJFDTotat-WXJY201715017.htm).
- [11] DING H, HE Q, ZENG L, et al. Motion intent recognition of individual fingers based on mechanomyogram[J]. *Pattern Recognition Letters*, 2017, 88(1): 41-48.
- [12] ALVES N, CHAU T. Uncovering patterns of forearm muscle activity using multi-channel mechanomyography[J]. *Journal of Electromyogr Kinesiol*, 2010, 20(5): 777-786.
- [13] WANG Z, OATES T. Encoding time series as images for visual inspection and classification using tiled convolutional neural networks[C]// *Workshops at the twenty-ninth AAAI conference on artificial intelligence*. 2015: 1.
- [14] YAN J J, CHEN S Y, YAN H X, et al. Wrist pulse analysis and recognition based on recurrent plot and convolution neural network[J]. *Computer Engineering and Applications*, 2020(7): 26.
- [15] ZANGENEH E, RAHMATI M, MOHSENZADEH Y. Low resolution face recognition using a two-branch deep convolutional neural network architecture[J]. *Expert Systems with Applications*, 2020, 139: 112854.
- [16] WANG Z, YAN W, OATES T. Time series classification from scratch with deep neural networks: A strong baseline[C]// *2017 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2017: 1578-1585.
- [17] FAHIM S R, SARKER Y, SARKER S K, et al. Self attention convolutional neural network with time series imaging based feature extraction for transmission line fault detection and classification[J]. *Electric Power Systems Research*, 2020, 187: 106437.
- [18] SANAYHA M, VATEEKUL P. Remaining Useful Life Prediction Using Enhanced Convolutional Neural Network on Multivariate Time Series Sensor Data[J]. *Walailak Journal of Science and Technology (WJST)*, 2019, 16(9): 669-679.
- [19] LIU S, DENG W. Very deep convolutional neural network based image classification using small training sample size[C]// *2015 3rd IAPR Asian conference on pattern recognition (ACPR)*. IEEE, 2015: 730-734.
- [20] WANG S, ZHANG L, FU J. Adversarial transfer learning for cross-domain visual recognition[J]. *Knowledge-Based Systems*, 2020, 204: 106258.
- [21] ZHAO B, ZHANG X, ZHAN Z, et al. Deep multi-scale convolutional transfer learning network: A novel method for intelligent fault diagnosis of rolling bearings under variable working conditions and domains[J]. *Neurocomputing*, 2020, 407: 24-38.



**WANG Xin-ping**, born in 1995, post-graduate. His main research interests include pattern recognition of mechanomyography, embedded system development.



**XIA Chun-ming**, born in 1969, Ph. D., professor. His main research interests include mechanomyography, traditional Chinese medicine signal analysis.