**Software Engineering and Applications** 软件工程与应用, 2023, 12(2), 185-197
Published Online April 2023 in Hans. https://www.hanspub.org/journal/sea
https://doi.org/10.12677/sea.2023.122019

Hans汉斯

# 基于卷积神经网络的人脸表情识别算法

**韩　博[1*]，关珍博[1]，陈路路[1]，温　博[1]，周　云[2]**

[1]中国电子科技集团公司第五十四研究所，河北 石家庄
[2]陆装驻石家庄地区第一军代室，河北 石家庄

收稿日期：2023年2月20日；录用日期：2023年3月24日；发布日期：2023年3月31日

## 摘　要

**人脸表情识别是图像识别的一个重要领域。传统人脸表情识别主要基于人工提取特征，由于人脸表情丰富、背景复杂、差异范围大等问题，其存在算法鲁棒性较差、易受人脸身份信息干扰等问题，以及传统卷积神经网络易发生过拟合、梯度弥散、梯度爆炸等问题的现状，因此本文提出一种多层特征融合非反向传播稠密卷积神经网络的人脸表情识别算法。该算法应用了改进的HSIC (Hilbert-Schmidt Independence Criterion，希尔伯特－施密特独立性)-bottleneck来代替传统反向传播(Back Propagation, BP)，具有诸多独特的优点。在特征提取过程中，为了充分利用得到的特征图像，将卷积层稠密连接并引入attention机制，最终通过softmax分类器分类，得到分类结果。在FER2013数据集上进行了多次实验，与传统BP算法的卷积神经网络算法相比，不仅有效减轻过拟合现象，并且在模型收敛速度上更快、计算量更小、内存占用更小，证明了在人脸表情识别问题中非反向传播稠密卷积神经网络模型结构有效、提出的分类优化方法可行。**

## 关键词

**非反向传播稠密卷积神经网络，人脸表情识别，特征融合，HSIC-Bottleneck**

# Facial Expression Recognition Algorithm Based on Convolutional Neural Network

**Bo Han[1*], Zhenbo Guan[1], Lulu Chen[1], Bo Wen[1], Yun Zhou[2]**

[1]China Electronics Technology Group Corporation 54th Research Institute, Shijiazhuang Hebei
[2]In Shijiazhuang 1st Military Representative Office, Shijiazhuang Heibei

Received: Feb. 20th, 2023; accepted: Mar. 24th, 2023; published: Mar. 31st, 2023

## Abstract

**Facial expression recognition is an important field of image recognition. Traditional facial expres-**

*通讯作者。

文章引用: 韩博, 关珍博, 陈路路, 温博, 周云. 基于卷积神经网络的人脸表情识别算法[J]. 软件工程与应用, 2023, 12(2): 185-197. DOI: 10.12677/sea.2023.122019

**sion recognition is mainly based on manual extraction of features, which has the problems of poor algorithm robustness and susceptibility to interference by face identity information due to rich facial expressions, complex background and large range of differences, as well as the current situation that traditional convolutional neural networks are prone to overfitting, gradient dispersion and gradient explosion, etc. Therefore, this paper proposes a multilayer feature fusion using dense convolutional neural network without Back Propagation (BP) for face expression recognition algorithm. The algorithm applies a modified HSIC (Hilbert-Schmidt independence criterion)-bottleneck instead of the traditional BP, which has many unique advantages. In the feature extraction process, in order to make full use of the obtained feature images, the convolutional layers are densely connected and attention mechanism is introduced, and finally the classification results are obtained by the softmax classifier. Compared with the traditional BP algorithm of convolutional neural network algorithm, it not only effectively reduces the overfitting phenomenon, but also has faster model convergence, smaller computation and smaller memory consumption, which proves that the structure of non-back propagation dense convolutional neural network model is effective and the proposed classification optimization method is feasible in the face expression recognition problem.**

## Keywords

**Dense Convolutional Network without Back Propagation, Facial Expression Recognition, Features Fusion, HSIC-Bottleneck**



## 1. 引言

在计算机视觉领域中，图像识别技术日益成熟，人脸表情识别成为了热点研究课题。人脸识别问题相较其他问题而言，不仅需要考虑拍摄光照、遮挡和角度问题，同时还需要考虑人脸身份特征及人脸表情非刚性变化的特征。美国著名心理学家 Mehrabian 提出[1]，在人类的日常交流中，通过语言、声音传递的信息分别占全部的信息总量的 7%和 38%，而通过人脸表情传递的信息量则占到了 55%。美国心理学家 Ekman 和 Friesen 通过大量实验[2]，定义了人类六种基本表情：高兴、生气、惊讶、害怕、厌恶和悲伤。2013 年，Shaohua Wan 等人使用 Gabor 变换来提取特征[3]，降低了人脸姿态变化对识别准确率的影响。2015 年，Ali 等人利用经验模态分解(Empirical Mode Decomposition, EMD)技术[4]进行人脸表情识别，将二维图像经过连续投影得到面部特征图，并使用 EMD 技术对面部特征图进行分解。Liu 等人将 Riesz 基量局部二元模式和直推式传输线性判别分析相结合[5]，针对野外环境进行人脸表情识别，实现了较高的识别准确率。另外，Shin 等人使用独立分量分析和主成分分析在情感维度上进行表情识别[6]。Lajevardi 等人将三维彩色图像展开为二维矩阵[7]，使用 Lob-Gabor 滤波器提取特征，并且使用线性判别分析分类器对特征进行分类。传统人脸表情识别过程中的特征提取很大程度上依靠人为干预，算法鲁棒性差且精度不高。

现阶段，深度学习作为机器学习研究的一个新的领域，受到人们的广泛关注。深度学习在时效性和准确性上有了显著的提高。深度学习为越来越广泛的图像分类识别任务带来了新的性能水平，但目前深度学习算法的误差反向传播过程通常在生物学上被认为是不合理的[8] [9] [10]，不符合人脑神经网络的信号传播过程，同时随机梯度下降算法(Stochastic Gradient Descent, SGD)等相关梯度下降算法既耗费时间、

占用大量内存，且伴随需要不断搜索合适的诸多超参数和梯度弥散、梯度爆炸的问题。2019 年，Wan-Duo Kurt Ma 等人采用统计特征与输入和标签之间的相关独立性的方法[11]，代替传统的反向传播训练，但该方法在复杂的卷积神经网络以及特征重用网络中的结果与反向传播(Backward Propagation, BP)算法的结果还有一定差距。因此，本文提出了改进的 HSIC-Bottleneck 方法在理论和实际上替代神经网络的梯度反向传播算法，同时为了更好地利用提取到的图像特征，借鉴 FPN (Feature Pyramid Networks)，采用多个稠密卷积模块，在每个模块中将不同尺度的特征图在 channel 维度上进行拼接，并赋予每层特征在拼接时的权重，在前向传播过程中采用稠密连接的卷积层，且每层的输入由前面所有层的输出拼接而成，最大程度的利用提取到的特征。

在本文中，通过最大化隐藏层的输出与标签之间的互信息(Mutual Information)，且最小化隐藏层与输入之间的相互依赖，来确定损失函数的表示，这样就可以用最少的输入特征来预测输出，去掉冗余的特征，使得隐层特征高度效率化，增加了模型抑制过拟合的能力，提高了模型的泛化性，从原理和实际上改进了神经网络的训练方法，与 2019 年 Wan-Duo Kurt Ma 等人提出的 HSIC-Bottleneck 算法[12]相比，本文提出的算法使用 trace 法代替行列式法得到目标函数，且在特征之间加入特征重用，并引入注意力机制，这样可以将较大的感受野与较小的感受野相结合，使得特征的提取更加高效，对较大目标和小物体的检测效果都更加精准。本文试验使用 FER 2013 数据集进行测试，与传统反向梯度传播算法相比，在准确率明显提升的同时收敛速度明显提高，且泛化能力加强，计算量和内存占用大幅减少。

## 2. 非反向传播的卷积神经网络

目前国内外的卷积神经网络主要优化方法是 SGD 等相关的梯度下降算法[13]，使用从全局各个分类的误差获得的负梯度最大的方向，将全局寻优任务分解为一个小的子问题的集合，来逐层更新权重和偏置参数，这样使其具有了在较大的模型结构中训练参数多、所需算力较高、计算量庞大等缺点。反向梯度传播算法在生物学中的合理性一直以来都是一个备受争议的话题，也是探索替代方案的一个动机。在反向梯度传播算法中一个明显不符合生物学的问题是突触权重根据后面层的误差调整，这在生物学的理论中是不合理的[14] [15]。另一个问题是在前向传播和反向传播过程中，权重矩阵是共享的[16] [17]，此外反向传播是线性计算的且在计算前向传播时必须停止反向传播，反之亦然[18]。因此，寻找更加合理的反向传播替代方案已经成为 DNN (Deep Neural Network)领域急需解决的重要问题，也是迫在眉睫的要求。

2019 年，Wan-Duo Kurt Ma 等人提出的一种代替反向传播的方法[11]，引用了 HSIC 度量，使用抽样来测量两个分布依赖的强弱。该方法避免了梯度弥散、梯度爆炸现象的发生，可以在某层没有梯度的情况下跨层优化，可以同时并行优化多个层，但是在复杂的卷积神经网络中对于小物体检测的效果和传统反向传播训练算法相比还有一定差距。因此，本文提出一种改进的 HSIC-Bottleneck 方法来代替传统反向传播训练，基于希尔伯特空间核方法，将原始数据映射为再生核希尔伯特空间(Reproducing Kernel Hilbert Space, RKHS)中的核函数，然后再构造协方差算子来描述条件独立性，根据条件独立性的定理可以得到度量独立性和条件独立性的目标函数。在实际操作中，根据样本数据来构造经验条件协方差算子，通过 RKHS 中的内积运算可得到以 Gram 矩阵表示的估计函数。

信息论(Information Theory) [19]是学习理论和大量研究的基础。信息瓶颈(Information Bottleneck, IB)原则[20]概括了最小充分统计的概念，表示了最优化的平衡预测输出所需的信息与保留的关于输入的信息之间的关系，最优解可由下式得到：

$$\min_{P_{O_i|X},P_{Y|O_i}} I\left(X;O_i\right)-\beta I\left(O_i;Y\right)_i$$

其中 $X$、$Y$ 分别表示输入和标签，$O_i$ 表示在第 $i$ 个隐藏层的输出，$\beta$ 表示拉格朗日乘数，$I(X;O_i)$ 与 $I(O_i;Y)$ 分别表示 $X$ 与 $O_i$ 之间以及 $O_i$ 与 $Y$ 之间的互信息。从公式可看出，IB 主要保留了在压缩输入数据的特征信息时，隐藏层中关于标签的输出信息，即在保留预测所需的重要信息的同时去除无关的信息，达到平衡和消除冗余信息的目的。

实际中，由于很多原因导致 IB 难以计算。如果输入信号是连续的(如语音信号)，除非向网络中添加噪声信号，否则互信息 $I(X;O_i)$ 是无限的，因此许多算法将输入数据进行分箱操作，这样不会将数据扩展到高维，但这样会由于分箱的规则不同导致得到的结果也不同。额外的影响因素如离散和连续数据之间以及离散数据和差分熵之间的不同。本文使用 HSIC 代替在 IB 目标中的互信息，与互信息估计不同的是 HSIC 采用关于时间复杂度 $O(l^2)$ 的鲁棒计算方法，其中 $l$ 代表输入数据的数量。

HSIC 是 RKHS [12]的数据分布之间的交叉协方差算子的 Hilbert-Schmidt 范数，如下式所示：

$$\begin{aligned}\mathrm{HSIC}(P_{MN},H,G)&=\|C_{mn}\|^2\\&=E_{MNM'N'}\left[k_m(m,m')k_y(y,y')\right]+E_{MM'}\left[k_m(m,m')\right]E_{NN'}\left[k_n(n,n')\right]\\&\quad-2E_{MN}\left\{E_{M'}\left[k_x(m,m')\right]E_{N'}\left[k_n(n,n')\right]\right\}\end{aligned}$$

其中 $m\in M,n\in N$ 为输入数据，$N$ 为样本标签，$k_m,k_n$ 代表核函数，$H,G$ 表示希尔伯特空间，$E_{MN}$ 表示 $MN$ 的期望，由上式可推出以下表达式：

$$\mathrm{HSIC}(P_{MN},H,G)=(l-1)^{-1}tr(K_MHK_NH)$$

其中，$l$ 代表样本数量，$K_M\in R^{l\times l},K_N\in R^{l\times l},K_{M_{ij}}=k(m_i,m_j),K_{N_{ij}}=k(n_i,n_j),H\in R^{l\times l}$ 是一个中心对称的和幂等矩阵，$tr$ 表示矩阵的迹，这样一来计算代价只与样本的数量有关，尤其适合计算高维的小样本数据。

在一个由 $h$ 个隐藏层构成的全连接网络中，隐藏层的输出矩阵维度为 $(1,d_i)$，其中 $i\in\{1,\cdots,h\}$，$d_i$ 表示第 $i$ 个隐藏层单元数量，则每个 batch-size 的隐藏层输出矩阵大小为 $(b,d_i)$，其中 $b$ 为 batch-size 大小，在应用 IB 原则计算目标函数过程中，使用 HSIC 代替互信息可得：

$$Z_i^*=\arg\min_{Z_i}\mathrm{HSIC}(Z_i,X)-\beta\mathrm{HSIC}(Z_i,Y)$$

其中 $X$ 为输入数据，$Y$ 为标签数据，$\beta$ 是表示拉格朗日乘数，则 HSIC 的各项可如下表示：

$$\mathrm{HSIC}(Z_i,X)=(l-1)^{-1}tr(K_{Z_i}HK_XH)$$

$$\mathrm{HSIC}(Z_i,Y)=(l-1)^{-1}tr(K_{Z_i}HK_YH)$$

上式表明了最佳的隐藏层输出 $Z_i$ 在不依赖于输入的冗余信息和与输出具有最大相关性之间找到了平衡。理想情况下当 $Z_i^*$ 收敛时，预测标签所需的信息会被保留，且消除了导致过拟合的冗余信息。

## 3. 改进的稠密卷积神经网络

在传统 CNN 前馈网络中，浅层提取的特征较为粗略，检测到的是类似边缘的一些特征，将浅层特征可视化结果如图 1 所示。

在中间部分的卷积层提取到的特征就较为抽象，可以检测到部分的物体，例如面部器官或比较高级的纹理特征，将提取到的特征可视化结果如图 2 所示。

而在最后部分的卷积层提取到的特征就更加抽象，可能检测到完整的物体，例如人脸或更加高级的纹理特征，将提取到的特征可视化结果如图 3 所示。

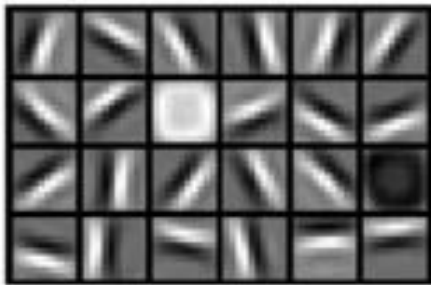

**Figure 1.** Visualization of shallow features
**图 1.** 浅层特征可视化效果图

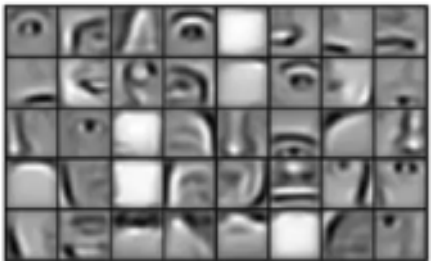

**Figure 2.** Visualization of middle layer features
**图 2.** 中间层特征可视化效果图

**Figure 3.** Visualization of the final layer features
**图 3.** 最后层特征可视化效果图

本文采用稠密连接的卷积神经网络，将相同尺度的特征图在 channel 维度进行拼接。使用多个稠密卷积模块相连，在每个稠密连接模块中有 5 层卷积，在拼接时分配给每层的输出特征不同的权重项。在前向传播过程中，稠密连接模块中的每一层都与其它所有层相连，即每一层都将之前所有层输出的图像特征连接起来作为自己的输入，并将自己的输出传递给之后的所有层，增强了图像特征在各个层之间的流动，充分利用了提取到的图像特征，稠密连接模块的结构如图 4 所示。

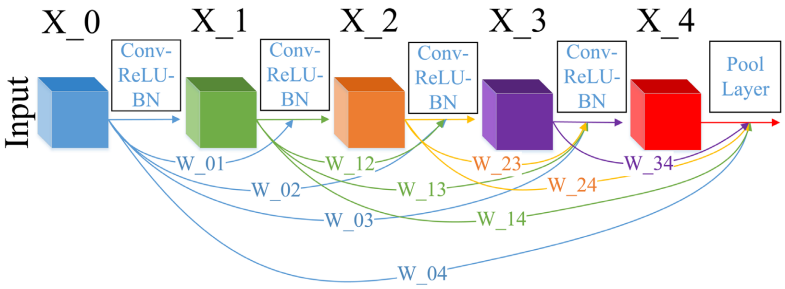


**Figure 4.** Diagram of the structure of the dense connection block
**图 4.** 稠密连接模块结构示意图

由于浅层卷积层得到的特征图较大，感受野较小，对小物体的检测比较敏感，但对物体的整体特征表达不够完整，而较深的卷积层特征图尺寸较小，感受野较大，可以很好的表达物体的整体特征，但细节特征表达有所缺失。因此，本文采用 FPN 将不同尺寸的特征图进行融合，对语义强但分辨率低的高层次特征进行上采样，并与高分辨率特征相结合，生成高分辨率和语义强的特征表示，充分的利用提取到的特征，并引入 attention 机制，在拼接时赋予之前层一定的权重值，这样每层的输入均为前面所有层的输出，加强了对小物体检测的敏感度同时还不丢失整体特征的表达。将前馈网络结构进行可视化展示，结果如图 5 所示。

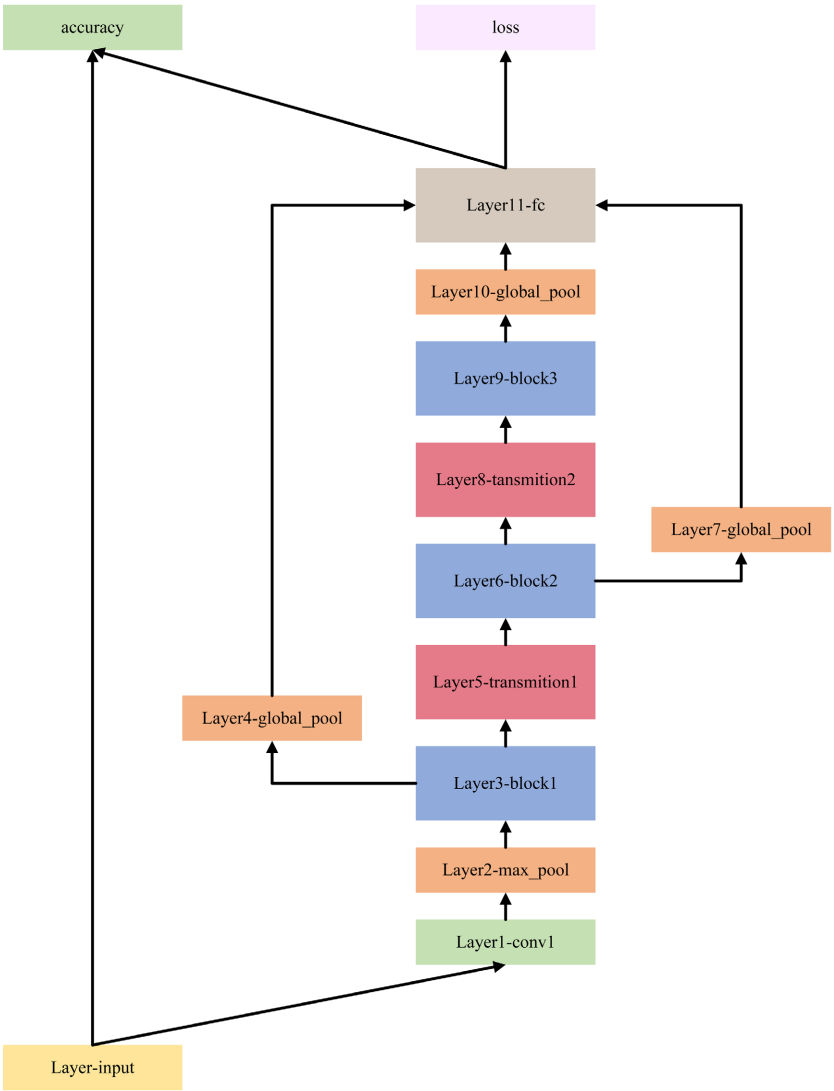


**Figure 5.** Diagram of the structure of the feed-forward network
**图 5.** 前馈网络结构图

由图 5 可知，网络共包含 3 个 block，其中 block 表示稠密连接模块，三个稠密连接模块的输出通过 FPN 进行融合，然后经过全局平均池化降维成二维向量。layer_11 为注意力机制模块，经过全连接层后可计算出每个通道维度的注意力得分，以达到对重要信息更加关注的效果，同时可去除冗余信息。其中，预训练网络为除 layer_11 外的隐藏层。

## 4. 人脸表情分类实验

本文所做的实验是基于 Python 语言的 Tensorflow 框架，采用 Tensorboard 将结果进行可视化，硬件平台 CPU 采用 Intel Core i7-9700k，GPU 采用单块 NVIDIA GeForce RTX 2080，显存为 8 GB。

### 4.1. 数据集预处理

本文采用 2013 年 Kaggle 比赛用的 FER2013 公开数据集，数据集以 csv 文件的形式保存，数据信息包含表情分类的标签、像素值、用途(训练、验证、测试)，数据样本示例如图 6 所示。

| | emotion | pixels | Usage |
|---|---|---|---|
| 0 | 0 | 70 80 82 72 58 58 60 63 54 58 60 48 89 115 121... | Training |
| 1 | 0 | 151 150 147 155 148 133 111 140 170 174 182 15... | Training |
| 2 | 2 | 231 212 156 164 174 138 161 173 182 200 106 38... | Training |
| 3 | 4 | 24 32 36 30 32 23 19 20 30 41 21 22 32 34 21 1... | Training |
| 4 | 6 | 4 0 0 0 0 0 0 0 0 0 0 0 0 3 15 23 28 48 50 58 84... | Training |

**Figure 6.** Example of data sample
**图 6.** 数据样本示例

将数据按用途分为训练集、开发集和测试集三部分，其中训练集数据 28708 条，开发集和测试集各 3589 条，训练集、开发集、测试集数据样本示例如图 7 所示。

| | pixels | emotion |
|---|---|---|
| 0 | 70 80 82 72 58 58 60 63 54 58 60 48 89 115 121... | 0 |
| 1 | 151 150 147 155 148 133 111 140 170 174 182 15... | 0 |
| 2 | 231 212 156 164 174 138 161 173 182 200 106 38... | 2 |
| 3 | 24 32 36 30 32 23 19 20 30 41 21 22 32 34 21 1... | 4 |
| 4 | 4 0 0 0 0 0 0 0 0 0 0 0 0 3 15 23 28 48 50 58 84... | 6 |

| | pixels | emotion |
|---|---|---|
| 0 | 170 118 101 88 88 75 78 82 66 74 68 59 63 64 6... | 0 |
| 1 | 7 5 8 6 7 3 2 6 5 4 4 5 7 5 5 5 6 7 7 7 10 10 ... | 5 |
| 2 | 232 240 241 239 237 235 246 117 24 24 22 13 12... | 6 |
| 3 | 200 197 149 139 156 89 111 58 62 95 113 117 11... | 4 |
| 4 | 40 28 33 56 45 33 31 78 152 194 200 186 196 20... | 2 |

| | pixels | emotion |
|---|---|---|
| 0 | 254 254 254 254 254 249 255 160 2 58 53 70 77 ... | 0 |
| 1 | 156 184 198 202 204 207 210 212 213 214 215 21... | 1 |
| 2 | 69 118 61 60 96 121 103 87 103 88 70 90 115 12... | 4 |
| 3 | 205 203 236 157 83 158 120 116 94 86 155 180 2... | 6 |
| 4 | 87 79 74 66 74 96 77 80 80 84 83 89 102 91 84 ... | 3 |

**Figure 7.** Example of training set, validation set and test set data samples
**图 7.** 训练集、验证集、测试集数据样本示例

训练集、开发集、测试集数据标签的分布直方图如图 8 所示。

将数据中的像素值还原为图像，可得到 48 × 48 的灰度图，共有 7 种表情：愤怒、厌恶、恐惧、高兴、悲伤、惊讶、中性，对应图像示例如图 9 所示。

### 4.2. 实验结果分析

在本文实验中比较了在使用相同参数情况下的传统反向传播训练的网络与非反向传播的网络的性能，传统网络使用交叉熵损失函数和 Adam 优化器，并且测试了采用不同数量稠密卷积模块的非反向传

播网络的性能，并比较了在使用不同激活函数时的性能，测试了在不同的学习率下的收敛速度及最终分类结果，最后比较了在使用相同参数的情况下，本文提出的算法与目前最先进的三种算法的结果，验证了算法的性能[21]。在训练时，HSIC-Bottleneck 的拉格朗日乘数 $\beta$ 根据经验设置为 100。

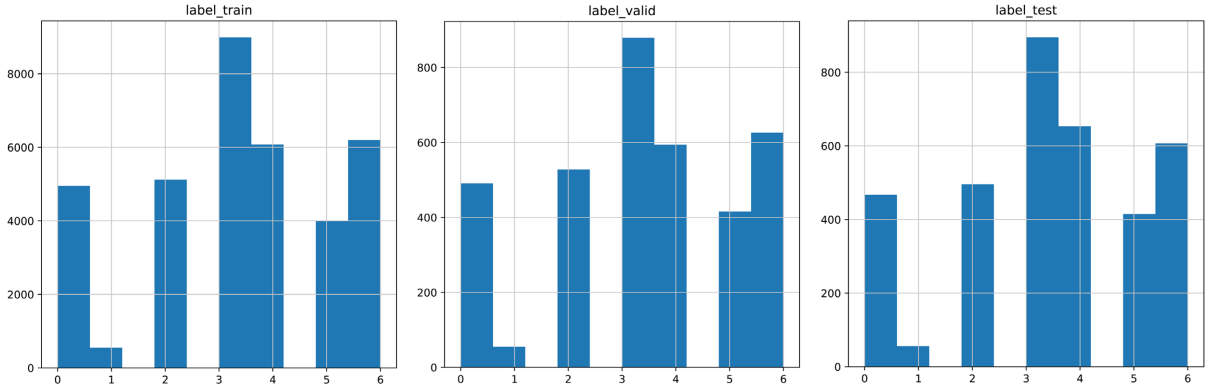


**Figure 8.** Histogram of label distribution for training set, validation set, and test set data
**图 8.** 训练集、验证集、测试集数据标签分布直方图

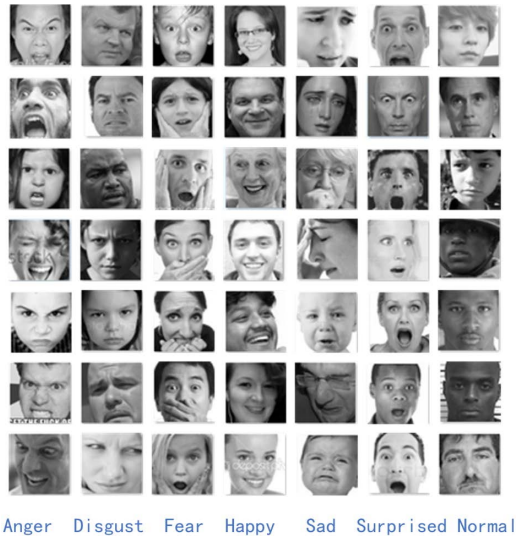


**Figure 9.** Example of 7 expressions in FER2013 expression library
**图 9.** FER2013 表情库 7 种表情示例

为了验证本文提出的非反向传播算法，使用深造训练的网络和传统反向传播训练的网络准确率和损失值变化如图 10 和图 11 所示。

在图 10、图 11 中，横坐标表示训练步数，共训练 10,000 步，纵坐标分别表示准确率和损失值，最终准确率达到了 0.9946。从图 10、图 11 中可以看出，传统 BP 算法在 10,000 时还未达到收敛，而非反向传播算法在 2500 步时达到收敛，非反向传播的网络具有更高的准确率，且收敛速度更快，这是因为本文提出的模型可以去除冗余的特征信息，且使得每层的特征与输出的相关性更强，在减少计算量的同时加快了模型收敛速度。在传统的反向传播训练中，计算复杂度很高，基于后面的梯度顺序向前计算所有层，往往需要高级的算力。相比之下，提出的 HSIC-Bottleneck 可以单独训练每一层，允许每层单独优化，并不向前传递梯度，实现并行计算，更好的提高了计算效率。在使用不同数量的稠密卷积模块测试时的准确率变化如图 12 所示。

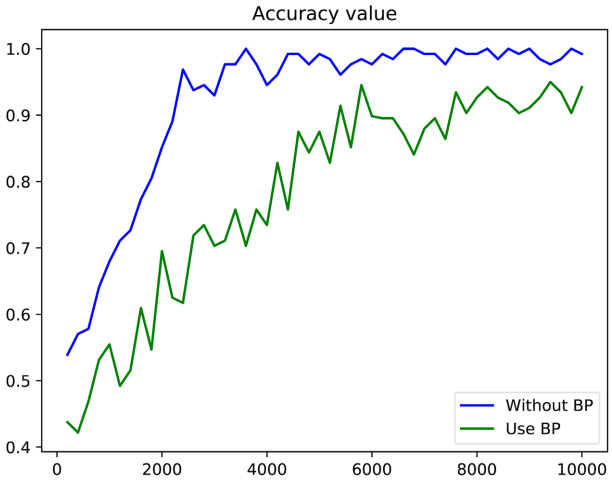


**Figure 10.** Accuracy curves of non-back propagation networks and BP networks with the same structure and parameters
**图 10.** 具有相同结构及参数的非反向传播网络与 BP 网络准确率变化曲线

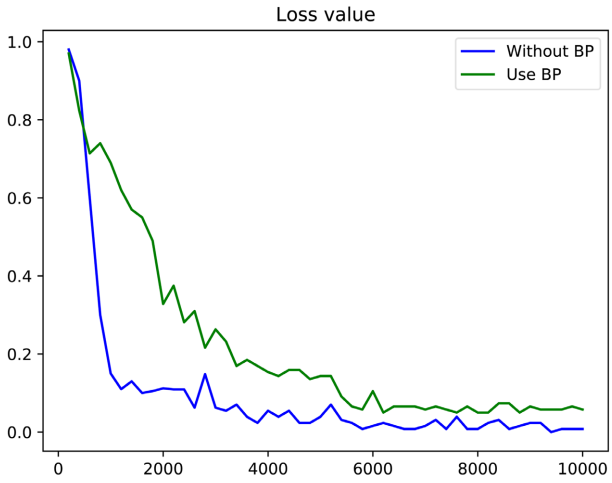


**Figure 11.** Curves of loss values of non-back propagation networks and BP networks with the same structure and parameters
**图 11.** 具有相同结构及参数的非反向传播网络与 BP 网络损失值变化曲线

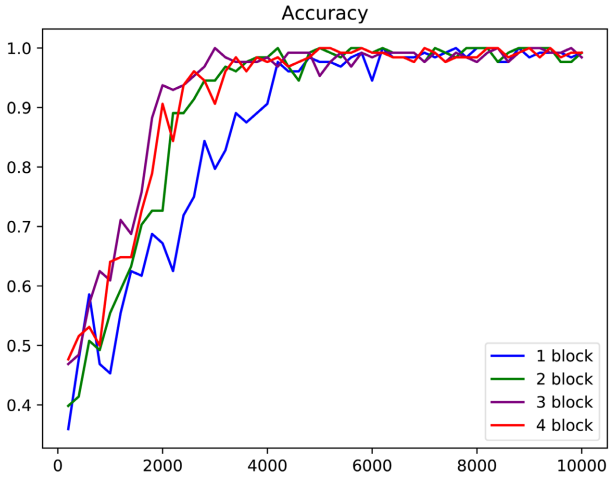


**Figure 12.** Accuracy curves of different number of dense convolution blocks
**图 12.** 不同数量稠密卷积模块准确率变化曲线

在图 12 中横坐标代表训练步数，共训练 10,000 步，纵坐标代表准确率。从图中可以看出，在使用 3 个稠密卷积模块时效果最好，在准确率更高的同时收敛速度更快，在使用 4 个稠密卷积模块时会使参数量增加，导致模型的收敛速度降低。在使用不同的激活函数测试时的结果如图 13 所示。

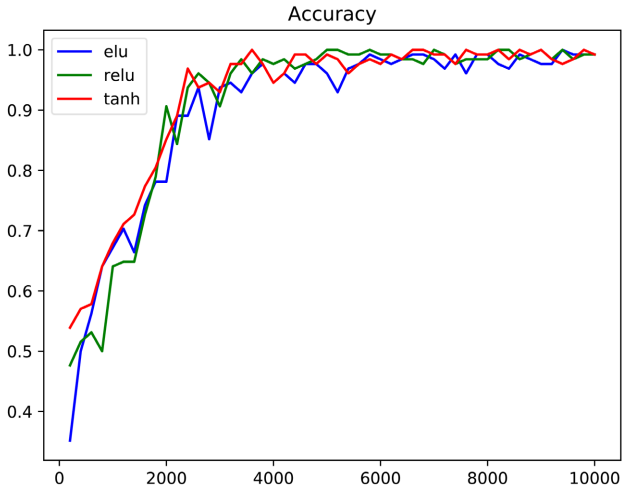


**Figure 13.** Accuracy curves when using different activation functions
**图 13.** 使用不同激活函数时准确率变化曲线

从图 13 中可以看出，使用 tanh 函数和 elu 函数得到的结果几乎一样，但不如 relu 函数表现好。非反向传播算法在使用不同的学习率时，模型在训练过程中的表现如图 14 所示。

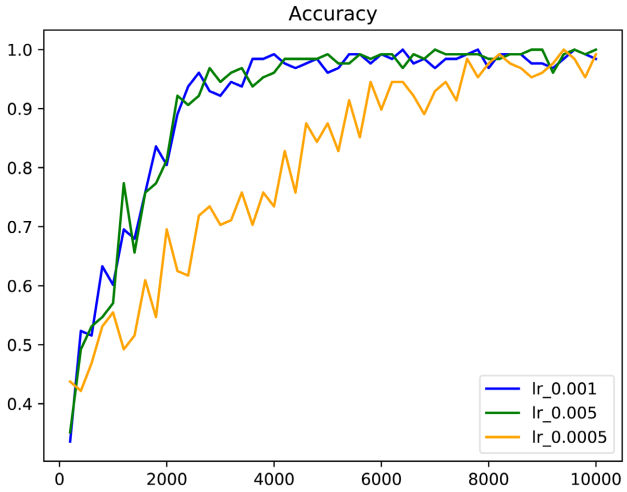


**Figure 14.** Performance of non-back propagation algorithms when using different learning rates
**图 14.** 非反向传播算法在使用不同学习率时的表现

从图 14 可知，学习率为 0.001 与 0.005 时模型的收敛速度几乎相同，但当学习率设置为 0.005 时，模型的表现最优，当学习率为 0.0005 时，模型收敛速度明显较慢。学习率设置的越小可使模型的最终结果越接近最优解，学习率设置的越大则可使模型迭代的速度更快，因此，本文通过多组多次的实验，得到了权衡模型的收敛速度和模型的表现能力的学习率。

本文将基于非反向传播的人脸表情识别算法与 Alexnet，Inception v2，Resnet 50 进行了比较，学习率均为 0.005，batch size 均为 128，在使用相同的环境及时的结果如图 15 所示。

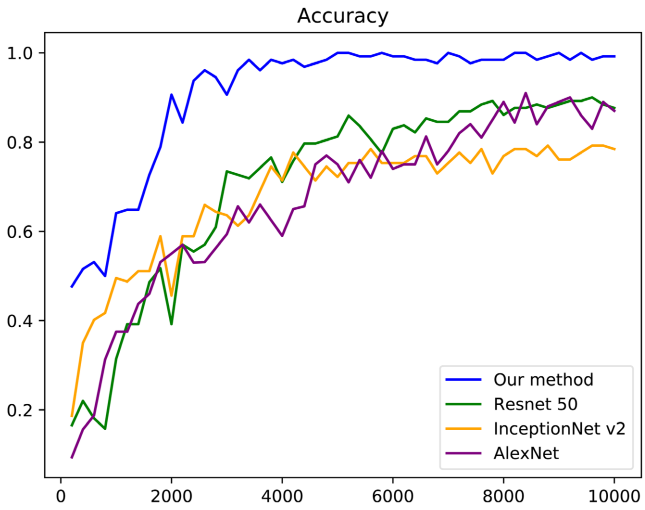


**Figure 15.** Accuracy curves of different models using the same environment and parameters
**图 15.** 使用相同的环境及参数时不同模型的准确率变化曲线

从图 15 中可看出，本文提出的算法的收敛速度明显比其他算法快，与反向传播相比，HSIC-Bottleneck 更好地分离了单个神经元表示中的隐藏信号，这表明 HSIC-Bottleneck 目标有助于使提取到的特征的分布更加独立，更容易与其标签相关联，且最终的准确率与其他反向传播算法的准确率相差不多。其中每种算法迭代一步的运行时间及在测试集上的最终结果如表 1 所示。

**Table 1.** Detection speed and average accuracy of different models on the test set
**表 1.** 不同模型在测试集上的检测速度及平均准确率

| Method | Per step running time (s) | Number of parameters (M) | The results in testset |
|---|---|---|---|
| Alexnet [22] | 0.243178 | 60 | 0.89 |
| Inception v2 [23] | 0.565114 | 11 | 0.93 |
| Resnet 50 [24] | 0.180212 | 24 | 0.95 |
| Our method | 0.049286 | 7 | 0.97 |

从表 1 中可看出，本文提出的算法在达到与传统反向传播算法的准确率的同时大幅提高了运行速度，并且在模型收敛速度上更快、计算量更小、内存占用更小，证明了本文提出的算法性能更好，优化方法可行。

## 5. 结束语

本文提出了一种基于非反向传播稠密卷积神经网络的人脸表情识别算法，采用 HSIC-Bottleneck 代替传统的梯度反向传播，首先训练一个类似对输入变量编码的网络，从而更容易得到与输出相关的所需信息，使用 HSIC-Bottleneck 作为训练目标，在没有反向传播的情况下训练深度神经网络，然后用训练好的网络在深造训练期间为分类器提供更良好的特征，进一步改善网络算法的性能，使用 Adam 优化器，但没有反向传播。HSIC-Bottleneck 训练网络，可以删掉输入信息中的冗余，并加强和标签数据的相关性，优化了隐藏层的输出信息，加快了模型的收敛速度，避免了反向传播中出现的梯度弥散、梯度爆炸问题，大幅减少了分类任务中的计算量。前馈网络采用基于 attention 机制的稠密卷积神经网络结构，更加充分地利用了提取到的特征，在对小物体保持敏感的同时也能很好的学习物体的整体特征。实验结果表明，

较其它传统深度学习算法，非反向传播的稠密神经网络算法在人脸表情识别任务中的训练速度、准确率和计算量有巨大优势。

本文提出的非反向传播神经网络，属于端到端任务，能够自动提取输入数据的特征，不需要人工干预，在人脸表情识别方向取得了很好的效果，其检测结果远远高于一般机器学习的方法。但是端到端的学习任务往往需要大量的有标签数据，这对于缺少标签数据的表情识别领域提出了很大的挑战，同时也是下一步要研究的重点。

## 参考文献


[1] Mehrabiau, A. (1968) Communication without Words. *Psychology Today*, **2**, 53-56.

[2] Ekman, P. (1972) Universals and Cultural Differences in Facial Expressions of Emotion. *Nebraska Symposium of Motivation*, Vol. 19, 207-283.

[3] Wan, S. and Aggarwal, J.K. (2013) A Scalable Metric Learning-Based Voting Method for Expression Recognition. 10*th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition* (*FG*), Shanghai, 22-26 April 2013, 1-8.

[4] Ali, H., Hariharan, M., Yaacob, S., *et al*. (2015) Facial Emotion Recognition Using Empirical Mode Decomposition. *Expert Systems with Applications*, **42**, 1261-1277. https://doi.org/10.1016/j.eswa.2014.08.049

[5] Liu, C. and Wechsler, H. (2002) Gabor Feature Based Classification Using the Enhanced Fisher Linear Discriminant Model for Face Recognition. *IEEE Transactions on Image Processing*, **11**, 467-476. https://doi.org/10.1109/TIP.2002.999679

[6] Shin, Y.S. (2006) Recognizing Facial Expressions with PCA and ICA onto Dimension of the Emotion. *Joint IAPR International Workshops on Statistical Techniques in Pattern Recognition* (*SPR*) *and Structural and Syntactic Pattern Recognition* (*SSPR*), Hong Kong, 17-19 August 2006, 916-922. https://doi.org/10.1007/11815921_101

[7] Lajevardi, S.M. and Hussain, Z.M. (2010) Emotion Recognition from Color Facial Images Based on Multilinear Image Analysis and Log-Gabor Filters. 2010 25*th International Conference of Image and Vision Computing New Zealand* (*IVCNZ*), Queenstown, 8-9 November 2010, 1-6. https://doi.org/10.1109/IVCNZ.2010.6148802

[8] Widrow, B., Greenblatt, A., Kim, Y., *et al*. (2013) The No-Prop Algorithm: A New Learning Algorithm for Multilayer Neural Networks. *Neural Networks*, **37**, 180-186. https://doi.org/10.1016/j.neunet.2012.09.020

[9] Yosinski, J., Clune, J., Bengio, Y., *et al*. (2014) How Transferable Are Features in Deep Neural Networks?

[10] Kleinberg, J.M. (1997) Two Algorithms for Nearest-Neighbor Search in High Dimensions. 29*th ACM Symposium on Theory of Computing*, El Paso, 4-6 May 1997, 599-608. https://doi.org/10.1145/258533.258653

[11] Ma, W.D.K., Lewis, J.P. and Kleijn, W.B. (2019) The HSIC Bottleneck: Deep Learning without Back-Propagation. *Proceedings of the AAAI Conference on Artificial Intelligence*, **34**, 5085-5092. https://doi.org/10.1609/aaai.v34i04.5950

[12] Gretton, A., Bousquet, O., Smola, A., *et al*. (2005) Measuring Statistical Dependence with Hilbert-Schmidt Norms. *Algorithmic Learning Theory*, 16*th International Conference*, *ALT* 2005, Singapore, 8-11 October 2005, 63-77.

[13] Talpur, N., Abdulkadir, S.J., Alhussian, H., *et al*. (2022) A Comprehensive Review of Deep Neuro-Fuzzy System Architectures and Their Optimization Methods. *Neural Computing and Applications*, **34**, 1837-1875.

[14] Lillicrap, T.P., Cownden, D., Tweed, D.B., *et al*. (2016) Random Synaptic Feedback Weights Support Error Backpropagation for Deep Learning. *Nature Communications*, **7**, 13276. https://doi.org/10.1038/ncomms13276

[15] Xiao, W., Chen, H., Liao, Q., *et al*. (2018) Biologically-Plausible Learning Algorithms Can Scale to Large Datasets.

[16] Grossberg, S. (1987) Competitive Learning: From Interactive Activation to Adaptive Resonance. *Cognitive Science*, **11**, 23-63. https://doi.org/10.1111/j.1551-6708.1987.tb00862.x

[17] Lillicrap, T.P., Cownden, D., Tweed, D.B., *et al*. (2014) Random Feedback Weights Support Learning in Deep Neural Networks.

[18] Bengio, Y., Lee, D.H., Bornschein, J., *et al*. (2015) Towards Biologically Plausible Deep Learning.

[19] Bach, F. (2022) Information Theory with Kernel Methods. *IEEE Transactions on Information Theory*, **69**, 752-775.

[20] Sun, Q., Li, J., Peng, H., *et al*. (2022) Graph Structure Learning with Variational Information Bottleneck. *Proceedings of the AAAI Conference on Artificial Intelligence*, **36**, 4165-4174. https://doi.org/10.1609/aaai.v36i4.20335

[21] Cong, S. and Zhou, Y. (2023) A Review of Convolutional Neural Network Architectures and Their Optimizations. *Artificial Intelligence Review*, **56**, 1905-1969. https://doi.org/10.1007/s10462-022-10213-5

[22] Krizhevsky, A., Sutskever, I. and Hinton, G. (2012) ImageNet Classification with Deep Convolutional Neural Networks. *Communications of the ACM*, **60**, 84-90.

[23] Ioffe, S. and Szegedy, C. (2015) Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift.

[24] Szegedy, C., Vanhoucke, V., Ioffe, S., *et al*. (2016) Rethinking the Inception Architecture for Computer Vision. 2016 *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Las Vegas, 27-30 June 2016, 2818-2826. https://doi.org/10.1109/CVPR.2016.308