

结合 LBP 特征和深度学习的人脸表情识别

张广世¹, 葛广英¹, 朱荣华¹, 孙 群²

(1. 聊城大学 物理科学与信息工程学院, 山东 聊城 252059;

2. 聊城大学 机械与汽车工程学院, 山东 聊城 252059)

摘要: 对于人脸表情识别, 传统方法是先提取图像特征, 再使用机器学习方法进行识别, 这种方法不但特征提取过程复杂且泛化能力也差; 为了达到更好的人脸表情识别效果, 文中提出一种结合特征提取和卷积神经网络的人脸表情识别方法; 首先使用基于 Haar-like 特征的 AdaBoost 算法对于数据库原始图片进行人脸区域检测, 然后提取人脸区域局部二值模式 (Local Binary Patterns, LBP) 特征图, 将其尺寸归一化后输入到改进的 LeNet-5 神经网络模型中进行识别; 在 CK+ 和 JAFFE 数据集上采用 10 折交叉验证方法进行实验, 分别为 98.19% 和 96.35% 的准确率; 实验结果表明该方法与其他主流方法相比在人脸表情识别上有一定的先进性和有效性。

关键词: 图像处理; LBP 特征; 人脸检测; 卷积神经网络; 人脸表情识别

Facial Expression Recognition Based on LBP Features and DeepLearning

Zhang Guangshi¹, Ge Guangying¹, Zhu Ronghua¹, Sun Qun²

(1. College of Physics and Information Engineering, Liaocheng University, Liaocheng 252059, China;

2. College of Mechanical and Automotive Engineering, Liaocheng University, Liaocheng 252059, China)

Abstract: For facial expressions recognition, the traditional method is to execute feature extraction and recognize by machine learning. This method not only has complex feature extraction process but also poor generalization. In order to achieve better facial expression recognition, the paper proposes a facial expression recognition method combining feature extraction and convolutional neural network. Firstly, the AdaBoost algorithm based on Haar-like feature is used to detect the face region of the original image of the database, and then extract the local binary patterns (LBP) feature map of the face region, normalize the size and input it into the improved LeNet-5 network to recognize. The recognition rate is 98.19% and 96.35% respectively in the CK+ and JAFFE database with 10-fold Cross-validation method. The experimental results show that this method has certain advancement and effectiveness in facial expression recognition compared with other mainstream methods.

Keywords: image processing; LBP feature; face detection; convolutional neural network; facial expression recognition

0 引言

人脸表情在人与人交流过程中扮演着重要的角色, 是进行情感信息交流的主要方式。随着大数据与人工智能时代的来临, 如何在人机交互中实现计算机对人脸表情的识别, 成为了当下的一个热门研究领域。

现如今人脸表情识别大多分为两个方向, 先提取人脸表情特征再使用分类算法进行识别的传统方法和深度学习下的人脸表情识别方法。在提取特征方面, 众多学者提出了 LBP 特征^[1]、尺度不变特征变换 (scale-invariant feature transform, SIFT)^[2]、灰度共生矩阵 (gray-level co-occurrence matrix, GLCM)^[3] 和 Gabor 小波变换^[4] 等特征提取方式, 在分类算法上则有支持向量机 (support vector machine, SVM)^[5]、K 最近邻分类^[6] (k-NearestNeighbor,

KNN)、随机森林^[7] (random forest, RF) 等分类方式。但这些方法所提取特征受人干扰因素过大, 会造成人脸表情信息的丢失, 导致最终分类准确率不高, 除此以外, 这些方法对其他数据的泛化能力也较低。在 2012 年, Krizhevsky 使用卷积神经网络 (convolutional neural networks, CNN) 模型在 ImageNet 竞赛中取得冠军, 其 Top-5 错误率为 17%, 远超过传统方法下的图像分类正确率, 这一现象引发了研究人员的广泛关注, 从此深度学习被广泛应用于图像分类问题中。例如 Tang^[8] 将 CNN 与 SVM 相结合, 在 FER2013 数据集上达到了 71.2% 的识别率。Jeon^[9] 等结合方向梯度直方图 (histogram of oriented gradient, HOG) 和 CNN 来进行人脸表情识别, 取得了较好的识别效果。本文也将使用 CNN 的方式来识别人脸表情, 首先提取数据库图片 LBP 特征, 将其尺寸归一化之后输入到改进的 LeNet-5 神经网络中进行识别, 实验结果表明, 所提方法在 CK+ 和 JAFFE 数据库中可获得非常好的识别效果。

1 系统流程

文中系统流程如图 1 所示, 主要过程包括: 1) 图像预

收稿日期: 2019-07-09; 修回日期: 2019-07-26。

基金项目: 中央引导地方科技发展专项资金计划。

作者简介: 张广世 (1993-), 男, 山东聊城人, 硕士研究生, 主要从事图像处理方向的研究。

葛广英 (1964-), 男, 山东聊城人, 教授, 硕士生导师, 主要从事图像处理、物联网方向的研究。

处理。采用基于 Haar-like 特征的 AdaBoost 算法来提取人脸区域, 然后提取人脸区域 LBP 特征并将其尺寸归一化, 以作为神经网络的输入; 2) 改进神经网络。原始的 LeNet-5 网络是用来识别手写体数字图片, 该种图片背景单一且较为简单, 因此原始网络不适合直接对复杂的人脸表情进行识别, 需要对网络的各种参数和结构进行调整; 3) 实验与评估。使用调整好的神经网络对输入图片进行训练和识别, 统计准确率来判断文中方法性能。

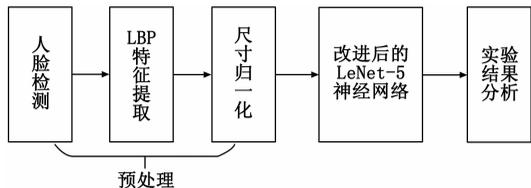


图 1 系统流程图

2 图像预处理

CK+与 JAFFE 数据集中的原始图像包含了人脸区域和背景, 需要经过预处理去除冗余信息, 才适合作为神经网络的输入图像, 本文中的预处理为: 人脸检测、LBP 特征提取和尺寸归一化。

2.1 人脸检测

本文采用基于 Haar-like 特征的 Adaboost 算法检测人脸区域, 该算法运算速度和正确率都很高, 可满足实时检测的要求。该方法基本思想是使用 Adaboost 算法将基于图像特征的弱分类器训练为强分类器, 再将强分类器组合成级联强分类器来检测人脸区域, 所检测出的人脸图像如图 2 所示。



图 2 人脸检测示意图

2.2 LBP 特征提取

LBP 是一种描述纹理的算法, 具有旋转和灰度不变性^[10]等特点, 被广泛地应用于纹理分析、图像匹配等领域。常用的 LBP 有两种: 基本 LBP 算法和圆形 LBP 算法。基本 LBP 算法是将中心点像素值和中心点 8 邻域像素值作比较, 如果 8 邻域像素值小于中心像素值, 则标记为 0, 否则标记为 1, 然后从左上角开始顺时针将二值化后的 8 邻域像素值组合成一串 8 位二进制数字, 然后将其转换为 10 进制数字, 由该数字来代替中心像素值。圆形 LBP 算法则是将基本 LBP 算法的 3×3 邻域扩展到任意圆形邻域, 该邻域可由参数 (P,R) 表示, P 为邻域像素个数, R 为该邻域半径, 该算

法改善了基本 LBP 算法无法识别大尺寸纹理特征的缺点。本文中所提取到的人脸表情 LBP 特征图如图 3 所示。



图 3 LBP 算法示意图

图 3 中可以看出处理后的图像可明显突出表情特征。

2.3 尺寸归一化

通过双线性插值算法将已经检测出来的人脸 LBP 特征图尺寸进行调整, 使得所有的神经网络输入图片尺寸相同, 避免了输入到神经网络后出现的尺寸不匹配问题。尺寸归一化算法为双线性插值算法, 其基本思想是将待求像素点周围 4 个点的像素值进行加权平均, 最后计算出待求点的像素值。该方法可一定程度上避免失真, 保持图像的清晰度。尺寸归一化之后的图像如图 4 所示。

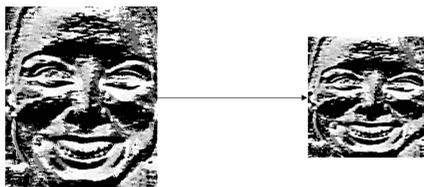


图 4 尺寸归一化示意图

3 神经网络模型

3.1 卷积神经网络

卷积神经网络是前馈式的神经网络, 可从输入数据中自动提取特征, 具有很强的学习和表达能力, 非常适合图像分类问题。该网络通常包含以下部分: 卷积层、池化层、全连接层^[11]。

一般来说, 卷积层和输入图像直接相连, 通过使用不同的卷积核将输入图像转变为抽象程度更高的图像特征并传给下一层, 卷积层的计算过程如式 (1) 所示:

$$I_j^l = f(\sum_i I_i^{l-1} \otimes k_{ij}^l + b_j^l) \quad (1)$$

式 (1) 中, I_j^l 为第 l 层第 j 个通道的输出, $f(\cdot)$ 为激活函数, I_i^{l-1} 为第 $l-1$ 层第 i 个特征图, k_{ij}^l 为卷积核, \otimes 为卷积运算, b_j^l 为相应的偏置。

通常在卷积层之后会再加入池化层, 池化层不但可以对特征降维, 而且可以较好地保持特征的尺度不变性, 同时也能降低神经网络的运算量。池化层计算公式如式 (2) 所示:

$$I_j^l = f(\beta_j \text{down}(I_j^{l-1}) + b_j^l) \quad (2)$$

其中: I_j^l 和 I_j^{l-1} 分别表示当前层输出特征图和上一层输出特征图, β_j 为当前池化层的权重值, b_j^l 为该层相应的偏置项。

表 1 改进后的 LeNet-5 网络结构

层数	C1	S2	C3	S4	C5	S6	F7	F8	F8
网络类型	卷积	池化	卷积	池化	卷积	池化	全连接	全连接	全连接
卷积核尺寸	9×9	3×3	5×5	3×3	3×3	3×3	—	—	—
步长	2	2	1	2	1	2	—	—	—
输出特征图尺寸	61×61	30×30	26×26	13×13	11×11	5×5	1×1	1×1	1×1
填充方式	全 0 填充	无	无	全 0 填充	无	无	—	—	—

式 (2) 中, $down(I_j^{l-1})$ 表示池化函数, 通过移动内核将特征图 I_j^{l-1} 划分为多个不重叠的子特征图, 然后对每个子特征图的像素进行数值运算, 常见的有求平均值和最大值。最大池化与平均池化相比, 优点在于可更多的保留图像纹理特征, 考虑到对人脸表情进行识别, 要更多关注不同表情之间的细微纹理差别, 所以在该层中采取最大池化方式。

全连接层的作用是将图像的二维特征组合为一维特征, 通常放在卷积层和池化层之后。该层输出公式如式 (3) 所示。

$$I^l = f(\omega^l I^{l-1} + b^l) \quad (3)$$

其中: I^l 和 I^{l-1} 分别为输出和输入, ω^l 为全连接层的权重值, b^l 为全连接层的相应的偏置项。

3.2 改进的 LeNet-5 网络模型

LeNet-5 模型最初是用来识别手写体数字, 其准确率达到了 98% 以上, 是一种经典的神经网络模型, 具体结构如图 5 所示。

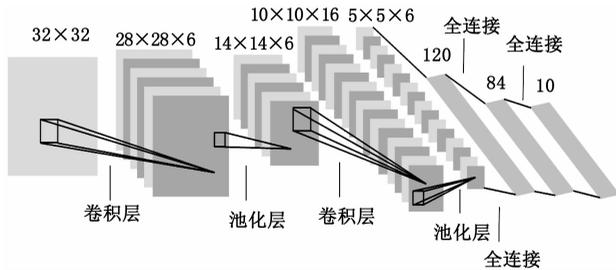


图 5 Lenet-5 网络结构图

使用 LeNet-5 模型对人脸表情进行识别, 发现损失值收敛速度过慢, 准确率较低。经分析之后对网络结构进行改进, 具体如下:

1) 原始 LeNet-5 是用来进行手写体数字识别, 其图片相对简单, 尺寸也较低, 为 32×32; 而本文所进行的人脸表情识别图像复杂, 且不同表情之间差异较小, 使用低尺寸图片会使得图像损失特征信息, 所以本文将输入图像尺寸调整为 128×128。

2) 调整卷积核的个数。卷积核的个数影响到之后输出的特征图个数, 考虑到人脸表情过于复杂, 本文将增加卷积核的个数以提高网络分辨表情的能力, 将前 6 层的卷积核个数分别设置为 16、16、32、32、64 和 64, 同时将后 3 层全连接层的输出节点个数分别设置为 128, 64 和 7。

3) 调整卷积核的尺寸大小。卷积核的尺寸与提取到的特征质量有关, 原始的 LeNet-5 采用 5×5 的卷积核, 用

来提取人脸图像特征效果不好, 本文将调整卷积核大小, 具体为表 1 所示。

4) 将 sigmoid 激活函数替换为 Leaky ReLU 激活函数。原有 sigmoid 函数计算量过大, 并且在输入过大或者过小时, 输出会接近 +1 和 -1, 斜率很低, 在使用梯度下降法时梯度下降过慢, 会严重降低训练速度, 不适合当前网络结构。所采用的 Leaky ReLU 函数与当前卷积神经网络常用的 ReLU 激活函数相比, 其优点在于可以避免当输入为负值时导致权重无法更新的问题。

5) 在原有的前 4 层网络之后, 新加入第 5 层卷积层和第 6 层池化层。本文网络结构输入图像尺寸为 128×128, 经过前 4 层网络之后特征图尺寸为 11×11, 需要继续降低尺寸, 减少计算的复杂度。

6) 在第一层全连接层之后加入 Dropout 函数, Dropout 可用来解决网络模型中数据过拟合的问题, 也可以增强网络模型的泛化能力。

7) 使用学习率指数衰减法对网络结构进行优化。使用该方法可提高神经网络训练速度, 使训练过程中损失曲线更加稳定的下降, 减少曲线的震荡次数。

改进之后的网络模型结构如表 1 所示。表 1 结构包含卷积层、池化层和全连接层, 输入层和 softmax 层未列于此表中。

4 实验结果与分析

4.1 实验数据与环境

本文将分别在 CK+ 与 JAFFE 公开数据集上进行测试, 由于 JAFFE 数据集不包含轻蔑类型的表情, 因此本文将去除 CK+ 数据集中的轻蔑类型, 最终将表情分类为生气、厌恶、恐惧、悲伤、快乐、惊讶、中性 7 种人脸表情。由于 CK+ 和 JAFFE 数据集原始表情数量较少, 本文将原始人脸表情图像以仿射变换方式对样本数量进行扩充。扩充后的 CK+ 数据集图像为 6476 张, JAFFE 数据集图像为 4689 张。人脸表情图像样例如图 6 所示。

本文方法使用硬件信息如下: GPU 为 NVIDIA GeForce GTX1660 6G, CPU 为 InterCorei3 8100 3.6 GHz, 内存为 8 G; 软件信息如下: 操作系统为 Window 10, 安装 Python3.6.5 和 TensorFlow 深度学习框架。神经网络中所设置超参数如下: 初始学习率为 0.0005, 学习率衰减指数为 0.96, 动量为 0.9, 正则化系数为 0.001, dropout 设置为 0.5, Epoch 为 20。



图 6 CK+与 JAFFE 数据集人脸表情样例

4.2 输入图像和网络结构之间的交叉对比

为证明本文方法的有效性, 并探索何种方式才能得出最高的正确率, 本文将使用 10 次 10 折交叉验证方法, 对不同的输入图像输入到不同的网络中得到的准确率做出了统计, 具体如表 2 所示。

表 2 CK+与 JAFFE 数据集实验准确率

输入图像类型	CK+		JAFFE	
	LeNet-5	Pro-LeNet-5	LeNet-5	Pro-LeNet-5
原数据集图像	5/%	5/%	5/%	5/%
基本 LBP	85.31	94.58	82.67	92.61
圆形 LBP(8,3)	87.57	95.91	85.23	93.76
圆形 LBP(8,5)	88.19	96.38	86.93	95.84
圆形 LBP(8,7)	87.23	98.19	87.59	96.35
圆形 LBP(8,7)	86.36	95.64	84.29	94.37

表 2 中第 1 列为不同的输入图像类型, 第 2 列和第 3 列为使用原始 LeNet-5 网络和改进之后的 LeNet-5 网络对 CK+数据集上的输入图像进行识别的准确率, 第 4 列和第 5 列为使用原始 LeNet-5 网络和改进之后的 LeNet-5 网络对 JAFFE 数据集上的输入图像进行识别的准确率。

从表 2 可以看出, 使用原始 LeNet-5 网络来进行人脸表情识别准确率较低, 即使输入图像为 LBP 特征图像, 也只提升了 2%~3% 左右的准确率, 其原因在于原始的网络结构并不适合人脸表情识别, 需要做出针对性的调整才能使准确率得到提高。由实验数据可知将数据集图像所提取的圆形 LBP(8, 5) 特征图输入到改进的 LeNet-5 网络中, 其识别率可达到最高, 在 CK+ 和 JAFFE 数据集可以分别达到 98.19% 和 96.35% 的准确率。

为进一步查看本文中方法性能, 使用上述方法统计不同表情的准确率, 如表 3 所示。

表 3 不同表情识别准确率

	CK+/%	JAFFE/%
生气	97.61	92.14
厌恶	98.05	95.69
恐惧	99.14	97.86
快乐	100	99.17
中性	96.37	93.48
悲伤	96.18	97.86
惊讶	100	98.31

由表 3 可知, 本文方法对 CK+数据集不同表情识别率均达到了 95% 以上, 其中高兴和惊讶的识别率为 100%; 在 JAFFE 数据集准确率稍低, 但也有 5 种表情识别率在 95% 以上, 足以证明本文方法的有效性。

为了更加直观地观察神经网络的性能, 本文绘制了 JAFFE 和 CK+数据集的损失值和准确率曲线, 如图 7 所示。

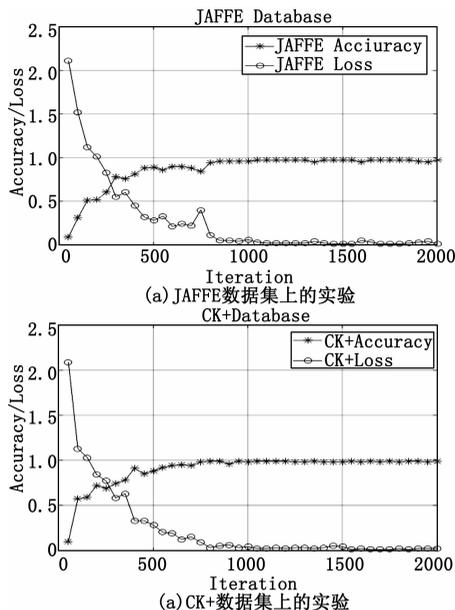


图 7 不同数据集损失和准确率曲线图

从图 7 中可以看出, 随着迭代次数的增加, JAFFE 和 CK+数据集的损失值和准确率均趋于平缓, 且损失值可以迅速下降, 说明本文所设计的网络模型较为合理。

4.3 与其他方法的对比

表 4 和表 5 展示了使用其他方法在 JAFFE 和 CK+表情数据集的识别率对比。

表 4 JAFFE 数据集实验准确率

方法	JAFFE/%
LBP+SRC ^[12]	87.42
SVM ^[13]	91.7
AR ^[14]	86.45
本文方法	96.35

表 5 CK+数据集实验准确率

方法	CK+/%
CNN ^[15]	92.15
CNN ^[16]	94.49
多通道 CNN ^[17]	94.58
SA ^[18]	96.5
本文方法	98.19

由表 4 可知, 使用深度学习对人脸表情进行识别要比传统方法的准确率高。这是因为传统方法所提取特征难以

完全描述人脸表情特点，进而在分类算法中准确率难以得到提高。

由表 5 可知，虽然同样采用了深度学习的方法来进行人脸识别，但本文方法的准确率是最高的。文献 [15] 和文献 [16] 所使用卷积神经网络结构较为基础，没有针对人脸表情识别问题对网络结构做出针对性的调整，且训练样本较少，最终准确率较低；文献 [17] 虽然对网络结构做出了改进，但其特征提取能力不如本文方法，使得准确率不高。文献 [18] 使用 HOG 提取特征，再使用主成分分析法降维，最终将特征向量送入深度稀疏编码网络中进行识别。其总体准确率尚可，但单一表情识别率如恐惧和悲伤不如本文方法准确率高。

5 结语

本文结合 LBP 特征提取与改进后的 Lenet-5 网络对人脸表情进行识别，其结果要优于主流方法。为更好地提取到图像特征，将原始图像的 LBP 特征输入到网络模型中去，并对比了输入哪一种 LBP 特征图准确率更高；将 Lenet-5 网络进行以适用于人脸表情识别，包括输入图像尺寸的调整、卷积核的调整、层数的调整和加入神经网络优化算法等。

文章下一步计划是将本方法应用于不同的数据集中以查看方法性能，此外还需要加强神经网络泛化能力。

参考文献:

[1] Ding Y, Zhao Q, Li B, et al. Facial expression recognition from image sequence based on LBP and Taylor expansion [J]. IEEE Access, 2017, 5: 19409 - 19419.

[2] Li H, Ding H, Huang D, et al. An efficient multimodal 2D+3D feature-based approach to automatic facial expression recognition [J]. Computer Vision and Image Understanding, 2015, 140: 83 - 92.

[3] 李蕊. 局部遮挡条件下的鲁棒表情识别方法研究与系统实现 [D]. 北京: 北京工业大学, 2015

[4] Zhou J, Zhang S, H MEI, et al. A method of facial expression recognition based on Gabor and NMF [J]. Pattern Recognition and Image Analysis, 2016, 26 (1): 119 - 124.

.....
(上接第 156 页)

参考文献:

[1] 侯浪. 三菱 Q PLC 应用技术研究 [D]. 武汉: 武汉理工大学, 2008.

[2] 程维湖. 电气自动控制中 PLC 技术的应用 [J]. 山东工业技术, 2019 (18): 155 - 185.

[3] 李锋. PLC 在电气自动化控制中的应用探讨 [J]. 机电信息, 2019 (17): 16 - 17.

[4] 杨剑, 袁浩. 基于三菱 Q 系列 PLC 工业生产控制系统的设计 [J]. 实验室科学, 2010, 13 (2): 162 - 164.

[5] 郭家星. 三菱 Q 系列 PLC 与安规测试仪的串口通信 [J]. 机械工程与自动化, 2013 (5): 185 - 186.

[6] 三菱电机自动化 (上海) 有限公司. Q 系列串行通信模块用户手册 (基础篇) [M]. 上海: 三菱电机自动化 (上海) 有限

[5] Hsieh C C, Hsieh M H, Jiang M K, et al. Effective semantic features for facial expressions recognition using SVM [J]. Multimedia Tools and Applications, 2016, 75 (11): 6663 - 6682.

[6] Wang X H, Liu A, Zhang S Q. New facial expression recognition based on FSVM and KNN [J]. Optik, 2015, 126 (21): 3132 - 3134.

[7] Pu X, Fan K, Chen X, et al. Facial expression recognition from image sequences using twofold random forest classifier [J]. Neurocomputing, 2015, 168: 1173 - 1180.

[8] Tang Y. Deep learning using linear support vector machines [J/OL]. arXiv: 1306. 0239 [2015-02-21]. <https://arxiv.org/abs/1603.0239>.

[9] Jeon J, Park J C, Jo Y J, et al. A real-time facial expression recognizer using deep neural network [A]. International Conference on Ubiquitous Information Management and Communication [C]. New York, ny: ACM, 2016: 1 - 4.

[10] 王景中, 李萌. 基于 LBP 和 PCA 机器学习的手势识别算法 [J]. 计算机测量与控制, 2015, 23 (4): 1320 - 1322, 1326.

[11] 杜洋, 王展青. 基于改进卷积神经网络的手写数字识别 [J]. 计算机测量与控制, 2018, 26 (7): 256 - 261.

[12] 赵栋杰. 改进的 LBP 算子和稀疏表达分类在人脸表情识别上的应用 [J]. 电子设计工程, 2016, 24 (20): 174 - 177.

[13] 钟志鹏, 张立保. 基于多核学习特征融合的人脸表情识别 [J]. 计算机应用, 2015, 35 (s2): 245 - 249.

[14] 苏志铭, 陈靓影. 基于自回归模型的动态表情识别 [J]. 计算机辅助设计与图形学学报, 2017, 29 (6): 1085 - 1092.

[15] 卢官明, 何嘉利, 闫静杰, 等. 一种用于人脸表情识别的卷积神经网络 [J]. 南京邮电大学学报 (自然科学版), 2016, 36 (1): 16 - 22.

[16] 徐鹏, 薄华. 基于卷积神经网络的人脸表情识别 [J]. 微型机与应用, 2015, 34 (12): 45 - 47.

[17] 张金刚, 方圆, 袁豪, 等. 一种识别表情序列的卷积神经网络 [J]. 西安电子科技大学学报, 2018, 45 (1): 150 - 155.

[18] Zeng N, Zhang H, Song B, et al. Facial expression recognition via learning deep sparse autoencoders [J]. Neurocomputing, 2018, 273: 643 - 649.

.....

[7] 郝智刚. 试论 PLC 与智能仪表的通讯设计与实现 [J]. 山东工业技术, 2018 (20): 158.

[8] 三菱电机自动化 (中国) 有限公司. MELSEC-Q/L 结构体编程手册 (特殊指令篇) [M]. 上海: 三菱电机自动化 (中国) 有限公司, 2010: 5 - 58 - 5 - 106.

[9] 张磊. AIBUS 现场通信总线的 PLC 工控设计方案 [J]. 现代制造, 2006 (9): 62 - 63.

[10] 潘丰, 王胜阳, 刘凯, 等. 基于 ModBus 的远程数据采集器设计 [J]. 河南科技学院学报 (自然科学版), 2019, 47 (2): 65 - 71.

[11] 三菱电机自动化 (中国) 有限公司. GX Works2 操作手册 (结构化工程篇) [M]. 上海: 三菱电机自动化 (中国) 有限公司, 2010: 12 - 2 - 12 - 12.