

文章编号: 2095-2163(2022)08-0110-05

中图分类号: TP391

文献标志码: A

# 基于深度学习的口罩佩戴识别技术研究

王俊秀<sup>1,2</sup>, 路旭鹏<sup>1</sup>, 郁晓庆<sup>2</sup>

(1 太原工业学院 计算机工程系, 太原 030008; 2 中北大学 大数据学院, 太原 030051)

**摘要:** 在公共场所佩戴口罩, 是防止新型冠状病毒传染的最主要手段, 在必要的场所, 每个人都必须佩戴口罩以进行自我保护。在人群相对集中的公共场所, 相互之间不可避免地存在遮挡干扰, 从而产生了小范围内的复杂干扰识别问题。如果使用单一的卷积神经网络对口罩佩戴进行识别, 有可能造成提取关键特征信息时聚焦度欠缺, 出现特征提取不足等问题。因此本文提出一种两渠道卷积神经网络的佩戴口罩识别方法。在卷积神经网络的基础上, 通过 2 个输入渠道, 分别对眼睛区域和眼睛以下的区域, 进行特征提取; 最后通过基于决策层的信息融合方法, 将 2 个渠道的识别结果加以融合, 从而得到最终的识别结果, 其平均识别准确率达到 98.8%。经过实验验证, 该方法在佩戴口罩的识别上, 取得了较好的识别准确率。

**关键词:** 新型冠状病毒; 口罩识别; 深度学习; 卷积神经网络; 融合技术

## Research on recognition technology of mask wearing based on deep learning

WANG Junxiu<sup>1,2</sup>, LU Xupeng<sup>1</sup>, YU Xiaoqing<sup>2</sup>

(1 Department of Computer Engineering, Taiyuan Institute of Technology, Taiyuan 030008, China;

2 School of Data Science, North University of China, Taiyuan 030051, China)

**[Abstract]** Wearing a mask in public is an important means to prevent the new coronavirus infection. When necessary, everyone should wear a mask to protect themselves. In the public where the crowd is relatively concentrated, there will inevitably be occlusion and interference between each other, which is a complex interference identification problem in a small area. Using a single convolutional neural network to identify people wearing masks may lead to insufficient focus when extracting key feature information, and the extracted features are inadequate. Therefore, this paper proposes a mask recognition method based on two-channel convolutional neural network. On the basis of the convolutional neural network, two input channels are used to extract features from the eyes and the area below the eyes. After that, the decision level fusion method is adopted to make voting decision based on the recognition results of two channels and the final recognition results are obtained. The obtained average recognition accuracy is 98.8%. The experimental results show that the method achieves good accuracy in mask wearing recognition.

**[Key words]** novel coronavirus; mask recognition; deep learning; convolutional neural network; fusion technology

## 0 引言

口罩佩戴识别是人脸识别的延伸应用。人脸识别是模式识别范畴的主要研究问题, 早在 20 世纪 90 年代, 就已作为独立课题受到广泛关注。人脸识别方法是多种方法的融合, 其中, 包括知识、模板匹配和统计学等<sup>[1]</sup>。对于复杂条件下的人脸检测问题, 张志伟<sup>[2]</sup>研究了可变光照条件下的人脸识别技术。苏岑等人<sup>[3]</sup>根据主成分分析算法系统设计了人脸表情的识别方法, 识别率达到 89.52%。人脸识别主要分为静态人脸识别和动态人脸识别两大类。近年来随着深度学习的发展, 将人脸识别与深度学习相结合成为一种趋势。深度神经网络中, 最普遍的应用是卷积神经网络(Convolutional Neural Network, CNN)<sup>[4]</sup>。卷积

神经网络能够直接将图像的每个像素数据均作为输入, 而且通过使用包括局部感受野、参数共享、稀疏连接、以及下采样的技术方法, 充分挖掘输入数据的特征并实现自主学习, 同时证明也对图像的所有其它形态的变换都有健壮性。卷积神经网络在人脸识别的有关研究中, 已经取得了非常好的识别效果<sup>[5-6]</sup>。近年来, 研究学界一直致力于提高卷积神经网络在人脸识别应用中的准确度。为了进一步提高深度网络在人脸识别上的准确率, Kim 等人<sup>[7]</sup>采取了初始化权重、重构网络结构等多种学习策略。李江等人<sup>[8]</sup>也指出了在深度网络模型的全连接层采用 Dropout 技术, 来减少由于训练样本不足所产生的过拟合问题。Wen 等人<sup>[9]</sup>使用构造 softmax 与 center 联合的损失函数, 来减少类内距离、并增大类间特征分散程度。Wang

**基金项目:** 山西省自然科学基金面上青年基金项目(201801D221209)。

**作者简介:** 王俊秀(1987-), 女, 硕士, 讲师, 主要研究方向: 计算机视觉、人工智能; 路旭鹏(1987-), 男, 本科, 工程师, 主要研究方向: 大数据、人工智能; 郁晓庆(1982-), 女, 博士后, 副教授, 主要研究方向: 图像处理、大数据、深度学习。

**收稿日期:** 2022-04-07

等人<sup>[10]</sup>采用了 Face R-CNN 框架进行人脸检测。

目前,已经有学者研究人脸口罩佩戴检测算法。肖俊杰<sup>[11]</sup>基于 YOLOv3 框架和 YCrCb 椭圆肤色模型,实现了人脸是否佩戴口罩和口罩佩戴是否规范的检测。邓黄潇<sup>[12]</sup>基于 RetinaNet 网络和迁移学习方法,对人们是否佩戴口罩进行检测。牛作东等人<sup>[13]</sup>通过增加自注意力机制和改进 RetinaFace 人脸识别算法,实现了人脸口罩检测。目前,卷积神经网络模型在进行人脸口罩佩戴识别时,因提取关键特征信息时聚焦性不够,影响了识别的准确率。针对此问题,本文采用聚焦机制的两渠道卷积神经网络方法,对口罩佩戴情况进行识别,并在自建数据集上进行实验分

析。实验表明,结果识别准确率已达到 99.4%。

### 1 基于卷积神经网络的佩戴口罩识别

#### 1.1 卷积神经网络

卷积神经网络是深度学习神经网络的一种,由输入层、卷积层、池化层和激活函数构成,卷积网络模型如图 1 所示。其特殊之处在于,卷积神经网络同时拥有卷积层和池化层。卷积层和池化层作为特征向量提取器,解决了特征向量选择的难题,并且能够减少网络模型的相关参数和复杂程度。这里,对卷积神经网络各组成部分的功能原理可给出阐释分述如下。

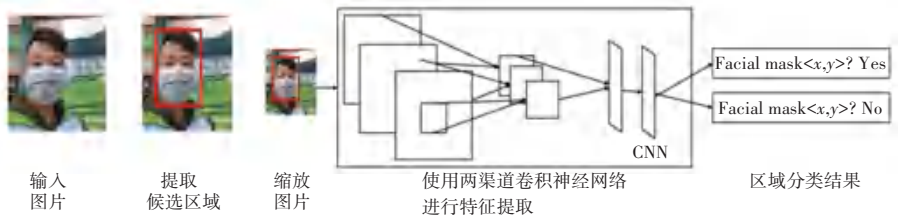


图 1 卷积神经网络模型

Fig. 1 Convolutional neural network model

(1)卷积层。在神经网络模型训练过程中,网络中的参数数量过多,是训练困难的一个重要原因。卷积神经网络通过随机初始化权重、权值共享的形式,减少了网络中的参数,增加了网络的泛化功能。在网络模型的训练中,可以训练成能够检测形状和边缘的滤波器。

(2)池化层。为减少模型中参数数量以及网络模型的训练时间,池化层将卷积层的输出进行特征组合。池化层通过计算卷积层输出的局部区域值,达到减少特征向量、防止过拟合和降低图像表达维度的目的。

常用的池化层计算方法有:最大池化和平均池化。其中,最大池化矩阵是选定区域内的最大值,认为该值可以代表区域特征;平均池化则选取整个区域特征的平均值,作为该区域的特征。

(3)激活函数。卷积神经网络中激活函数的主要作用,是使网络模型可以逼近任意分类函数。例如, Sigmoid 函数、Tanh 函数和 Relu 函数等,都是卷积神经网络常用的激活函数。Sigmoid 函数和 Tanh 函数的优势是能解决非线性问题,缺点是计算较为复杂,易出现梯度消失等问题。Relu 函数的优点是梯度稳定,计算相对节省时间。

#### 1.2 改进的卷积神经网络

为挖掘出不同形象的人物佩戴口罩的特征,本方法采用不同人物佩戴口罩的图像作为模型的输入,进行口罩佩戴的识别。选取合适的数据集,进行统一规格的预处理,使其大小符合卷积神经网络模型的输入。对每张图像取眼部特征和口罩特征。具体模型如图 2 所示。由图 2 可知,各部分的研究设计可做剖析论述如下。

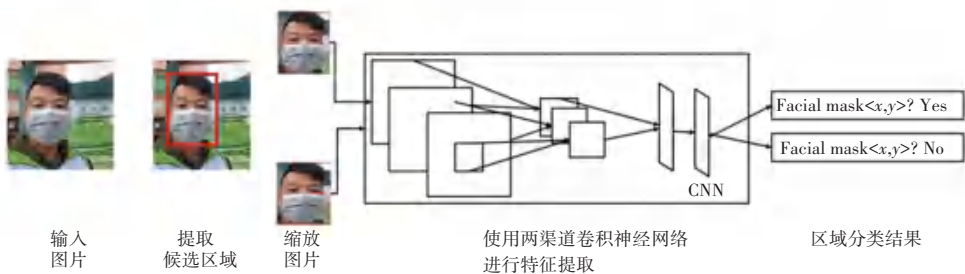


图 2 两渠道卷积神经网络模型

Fig. 2 Two-channel convolutional neural network model

(1)眼部区域特征提取。将去除无关背景的固定宽和高的区域,定义眼部所在的区域。其中,眼部区域所在范围的比例因子见表1。表1中, $x_{eyes}$ 、 $y_{eyes}$ 表示眼睛左上角坐标, $w_{eyes}$ 、 $h_{eyes}$ 表示眼睛的宽和高, $w_{face}$ 、 $h_{face}$ 表示脸部区域的宽和高。

(2)口罩区域特征提取。由于口罩区域范围较大,口罩形状较为规整,能够提取更多有效信息。同样,去除无关干扰背景后,用固定宽和高来获取口罩所在区域。其中,口罩的比例因子见表2。表2中, $x_{mask}$ 、 $y_{mask}$ 是口罩的左上角坐标, $w_{mask}$ 是口罩的宽度, $h_{mask}$ 是口罩的高度。

表1 眼部 Region of Interest 比例因子

Tab. 1 Proportion factor of Region of Interest related to the eyes

部位	$x_{eyes}$	$y_{eyes}$	$w_{eyes}$	$h_{eyes}$
眼睛	$\frac{1}{6 * w_{face}}$	$\frac{1}{4 * h_{face}}$	$\frac{17}{24 * w_{face}}$	$\frac{1}{4 * h_{face}}$

表2 口罩 Region of Interest 比例因子

Tab. 2 Proportion factor of Region of Interest related to the mask

部位	$x_{mask}$	$y_{mask}$	$w_{mask}$	$h_{mask}$
口罩	$\frac{1}{6 * w_{face}}$	$\frac{1}{4 * h_{face}}$	$\frac{23}{24 * w_{face}}$	$\frac{3}{5 * h_{face}}$

(3)两渠道决策层卷积神经网络。卷积神经网络在进行模型训练时,特征向量被压缩后会忽略部分图像信息,造成部分关键信息的丢失。但若只对关键区域进行图像数据特征提取,就有可能产生过拟合的问题,对特征信息提取不够充分。为解决上述问题,本文采用不同焦点的图像作为输入,即聚焦眼部区域和聚焦口罩区域或嘴部区域。通过2个渠道分别进行网络模型的训练,并对2个网络模型的识别结果进行融合,作为最后的识别结果。两渠道融合方法属于决策层融合,采用多数投票法作为最后判别结果。



(a) 预处理前图像



(b) 预处理后图像

图4 图像预处理前、后对比

Fig. 4 Comparison of images before and after preprocessing

为消除无关信息对口罩佩戴的识别影响,对无效背景区域进行消除。采用级联检测技术,从原始图像中获取眼部数据和口罩部分数据,对得到的数

据进行归一化处理,并将训练集输入到网络模型中。图5为原始图像和获取的眼部区域及口罩区域图像的示例。

$$c(x) = \text{larg} \max_j f(c_i^j(x)) \quad (1)$$

$$f(c_i) = (c_1 c_2 \cdots c_M) + (c_1 c_2 \cdots c_M + \cdots + c_1 c_2 \cdots c_M) + \cdots + (c_1 \cdots c_n c_{n+1} \cdots c_M + \cdots + c_1 \cdots c_{n-1} c_n \cdots c_M) \quad (2)$$

其中, $c_i$ 是卷积神经网络模型; $c_i(x)$ 是样本 $x$ 的在卷积神经网络 $c_i$ 的输出结果; $c_i^j(x)$ 是卷积神经网络 $c_i$ 在类别 $j$ 上的输出概率; $c_i^j(x) \in [0,1]$ ; $f(c_i)$ 是多个基本卷积神经网络识别结果进行“少数服从多数”的运算表示; $M = 2n$ ; $c(x)$ 结果为预测得票最多的标记。

## 2 实验结果与分析

### 2.1 数据预处理

实验所用自建的数据集,训练集、验证集和测试集各总量分别是900、300、300,数据中口罩佩戴分布均衡。训练集、验证集和测试集口罩佩戴分布如图3所示。

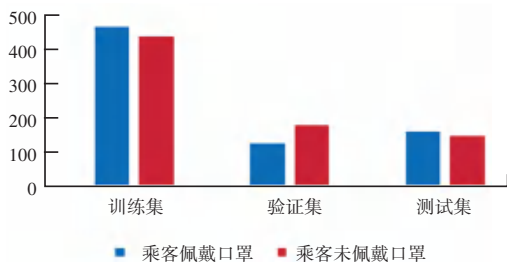


图3 实验数据集分布情况

Fig. 3 Distribution of experimental data sets

采用灰度化及高斯滤波法消除图像中无关的信息,保留有用信息,缩减无关信息的干扰。同时采用伽马变换,减少光纤对输入图像的影响<sup>[14]</sup>。预处理前后图像对比如图4所示。



图 5 原始图像以及获取的眼部区域和口罩区域图像

Fig. 5 Original images and the corresponding images of the eye area and mask area

### 2.2 多种模型对比分析

在模型训练过程中,分别将聚焦人眼图像和聚焦口罩图像输入到卷积神经网络模型中。同时采用随机梯度下降的方法进行优化处理,用 Xavier 均匀初始化,激活函数使用梯度稳定的 ReLU,并把学习率和学习率衰减分别设为  $1e-4$  和  $(1e-4)/epochs$ 。采用 2 个输入,分别输入到该卷积神经网络模型中,训练得到 2 个模型  $c_1, c_2$ 。模型评价指标是准确率 (Accuracy),记正确识别到戴口罩为 TP,正确识别到没戴口罩为 TN,错误识别到戴口罩为 FP,错误识别到没戴口罩为 FN,则 Accuracy 计算公式可写为:

$$Accuracy = \frac{TP + TN}{TP + FP + FN + FP} \times 100\% \quad (3)$$

表 3、表 4 分别是在训练集上的测试结果。其中,  $c_1$  模型对于口罩佩戴的识别准确率为 71%,  $c_2$  模型对于口罩佩戴的识别准确率为 87%。采用聚焦口罩图像作为输入建立的模型  $c_2$ ,对于口罩佩戴的识别准确率,要高于  $c_1$ 。

表 3 训练集在  $c_1$  模型的结果

Tab. 3 Results of  $c_1$  model in the training set

待识别	佩戴口罩	未佩戴口罩
佩戴口罩	0.75	0.25
未戴口罩	0.33	0.67

表 4 训练集在  $c_2$  模型的结果

Tab. 4 Results of  $c_2$  model in the training set

待识别	佩戴口罩	未佩戴口罩
佩戴口罩	0.89	0.11
未佩戴口罩	0.15	0.85

根据模型  $c_1, c_2$ 、即不同焦点训练得到的卷积神经网络模型,结合两渠道神经网络融合技术,对  $c_1, c_2$  在决策层采用“少数服从多数”的方法进行判别,得到最终用于识别口罩佩戴的两渠道卷积神经网络。其融合公式如下:

$$\begin{cases} c(x) = \mathop{\text{larg}}\limits_j \max f(c_j(x)) \\ f(c_i) = (c_1 c_2) + (c_1 \bar{c}_2 + \bar{c}_1 c_2) \end{cases} \quad (4)$$

改进后的卷积神经网络模型,对于口罩佩戴的识别准确率见表 5,平均识别准确率为 99.4%,识别环境

为白天常规环境。其识别准确率优于任意一种单输入的卷积神经网络模型,分类效果得到明显提高。

表 5 改进后卷积神经网络模型识别准确率

Tab. 5 Recognition accuracy of improved convolutional neural network model

待识别	佩戴口罩	未佩戴口罩
佩戴口罩	0.996	0.040
未佩戴口罩	0.008	0.992

### 2.3 模型抗性分析

为检验基于两渠道融合技术的深度卷积神经网络模型对人脸佩戴口罩识别效果的抗干扰能力,实验分别在白天环境、夜晚环境、常规无遮挡环境和佩戴复杂有遮挡环境中进行。表 6 给出了模型在多种环境下的识别效果。实验结果表明,本文提出的口罩识别模型在不同环境下,具有较高的识别准确率,能够满足防疫控制检验口罩佩戴情况的要求。

表 6 改进后卷积神经网络模型在不同环境下识别准确率

Tab. 6 Recognition accuracy of improved convolutional neural network model in different environments

待识别	Accuracy
白天常规环境	0.994
白天复杂环境	0.993
夜晚常规环境	0.987
夜晚复杂环境	0.980

## 3 结束语

疫情期间佩戴口罩,能有效预防病毒的传播,保护好自己和身边的人。本文提出的基于深度神经网络的两渠道模型,分别训练不同焦点作为输入的卷积神经网络,进行特征提取,并用融合技术进行最后的识别判别,平均识别准确率达到 98.8%。在未来的工作中,将着重研究如下方向内容:

(1) 尽可能多地收集样本,包含不同发型、不同装扮、不同外部环境的人群。

(2) 由于静态图片作为输入,容易出现误判。

因此,在公共场所摄像头的协助下,可以利用前、后帧之间的连续性,对输入进行多次判别,提高识别准确率。

(下转第 118 页)