

基于深度神经网络的伪装人脸识别

张润生^{1,2,3},贺超^{1,2,3},况朝青^{1,2,3}

(1.重庆邮电大学 通信与信息工程学院,重庆 400065;2.重庆高校市级光通信与网络重点实验室,重庆 400065;
3.泛在感知与互联重庆市重点实验室,重庆 400065)

摘要:伪装人脸识别在刑侦安防领域有着巨大的应用价值。针对现阶段对伪装人脸识别的研究较少、算法鲁棒性不强等缺点,提出了基于深度神经网络的伪装人脸识别算法。改进了SqueezeNet网络模型,并将其与FaceNet网络架构进行结合,用于人脸图像的身份识别。通过在训练数据集中引入伪装人脸图像,让网络学习到伪装的特征。实验结果表明,该算法识别准确率接近90%,相较于其他网络模型,具有更好的识别效果。

关键词:深度神经网络;伪装人脸识别;SqueezeNet网络模型;FaceNet网络架构

中图分类号: TN911.73;TP391.4

文献标识码: A

DOI:10.16157/j.issn.0258-7998.191314

中文引用格式: 张润生,贺超,况朝青. 基于深度神经网络的伪装人脸识别[J]. 电子技术应用, 2020, 46(5): 27-30.

英文引用格式: Zhang Runsheng, He Chao, Kuang Chaoqing. Disguised face recognition based on deep neural network[J]. Application of Electronic Technique, 2020, 46(5): 27-30.

Disguised face recognition based on deep neural network

Zhang Runsheng^{1,2,3}, He Chao^{1,2,3}, Kuang Chaoqing^{1,2,3}

(1.School of Communication and Information Engineering, Chongqing University of Posts and Telecommunications,
Chongqing 400065, China;
2.Optical Communication and Networks Key Laboratory of Chongqing, Chongqing 400065, China;
3.Ubiquitous Sensing and Networking Key Laboratory of Chongqing, Chongqing 400065, China)

Abstract: Camouflage face recognition has great application value in the field of criminal investigation and security. Aiming at the shortcomings of few researches on camouflage face recognition and weak robustness at present, a camouflaged face recognition algorithm based on deep neural network is proposed. The SqueezeNet network model has been improved and combined with the FaceNet network architecture for identity recognition of face images. By adding camouflage face images in the training data set, the network can learn the characteristics of the camouflages. The experimental results show that the recognition accuracy of the algorithm is close to 90%, which is better than other network models.

Key words: deep neural network; disguised face recognition; SqueezeNet network model; FaceNet network architecture

0 引言

近年来,视频监控领域下的人脸识别得到了广泛关注,通过监控抓捕到犯罪嫌疑人的新闻时有出现,大大提高了案件的侦破率。但是监控拍摄到的图像中很可能会存在遮挡,使得识别率下降,错失抓捕嫌疑人的机会。遮挡一般分为两种,即自然遮挡和人为伪装^[1]。自然遮挡包括树叶、栏杆等,人为伪装包括帽子、墨镜、围巾等。通过伪装,犯罪分子可以逃避监控的追踪,增大了案件的侦破难度。

针对这些问题,文献[2]提出了一种基于遮挡模式的稀疏表示分类的方法,构建的解析词典与测试图像具有相同的遮挡,提高了分类性能;利用稀疏字典学习的判别性来处理人脸识别问题中的连续遮挡。文献[3]使用Gabo小波、PCA和SVM来解决遮挡检测问题,将人脸图

像分成两个相等的分量,从每个分量中提取Gabor小波特征,用于降维主成分分析,最后使用局部二值模式来完成识别过程;在识别期间,权重被分配给测试图像的每个局部区域,与给定未被遮挡的训练示例的每个区域的可能性成比例。文献[4]在人脸图像的每个点上找到最大匹配区域,提取其傅里叶幅度谱作为特征,最后使用余弦相似度进行识别。但是这些方法都是针对一定类型的遮挡通过建模来完成识别的,泛化性能较差;而神经网络能够通过大量数据的训练来学习到相关特征,获得更好的识别性能,在信号调制、计算机视觉、文本分析、故障检测等领域均有广泛的应用^[5-8]。

本文提出了一种基于深度神经网络的伪装人脸识别算法,将SqueezeNet网络模型与FaceNet网络架构进行结合,并对SqueezeNet网络进行改进;构建伪装人脸数

人工智能 Artificial Intelligence

据集，在训练集中引入伪装人脸图像，使神经网络能够学习到伪装人脸的特征。实验表明，该方法能够提升识别的准确率。

1 网络架构

FaceNet 网络架构是 Google 公司于 2015 年所提出的^[9]，对于不同角度和姿态的照片，利用相同人脸具有高内聚性，不同人脸具有低耦合性，提出使用 CNN+Triplet 损失的方法，在 LFW 数据集上准确率达到 99.63%，在 YouTube 人脸数据集上准确率达到 95.12%，比以往准确度提升了近 30%。

1.1 FaceNet 网络架构

本文所使用的 FaceNet 网络架构如图 1 所示，在基于传统的深度神经网络的人脸识别基础上，对输出的特征向量进行处理，并提出了效果更好的损失函数，使模型的学习过程更高效，以此来获得更高的识别准确率。

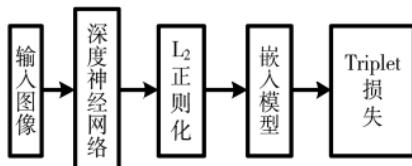


图 1 FaceNet 网络架构

FaceNet 的原理是把人脸图像之间的差异转化为空间距离的远近，通过判断距离与阈值的关系，从而区分出不同的身份。通过将输入图像送入到深度神经网络提取特征，再把特征向量经过 L_2 正则化后，嵌入到一个超球面上，图像在超球面上都将映射为一个点，不同点之间的距离就代表不同图像之间的差异程度。Triplet 损失函数通过选取来自超球面上合适的点组成 triplet 对，通过训练来改变不同图像映射在超球面上的位置，使得来自同一身份的人脸图像在超球面上的距离减小，来自不同身份的人脸图像在超球面上的距离增大，从而能够通过划分合适的阈值来实现对人脸图像的身份识别。

1.2 Triplet 损失

在 FaceNet 网络架构中，对深度神经网络输出的特征向量的处理，是其能够获得优异性能的重要原因之一。对于输入图像 x ，经过深度神经网络后，其输出特征向量为 $f(x)$ ，再经过嵌入模型 $f(x) \in \mathbb{R}^d$ 后，将 x 映射到 d 维的欧几里得空间中，通过 L_2 正则化使得 $\|f(x)\|_2=1$ ，以此来将其嵌入到 d 维的超球面上。Triplet 损失函数在训练数据集中选取合适的目标图像 x_i^a 、来自同一身份的正类图像 x_i^p 以及来自不同身份的负类图像 x_i^n 构成 triplet 对，通过训练使得类内距离减小，类间距离增大，即：

$$\left\| x_i^a - x_i^p \right\|_2^2 + \alpha < \left\| x_i^a - x_i^n \right\|_2^2, \forall (x_i^a, x_i^p, x_i^n) \in T \quad (1)$$

其中， α 为设定的阈值； T 为训练集中可能的 triplet 对的

集合，数量为 N 。那么，损失函数为：

$$L = \sum_i^N \left[\left\| f(x_i^a) - f(x_i^p) \right\|_2^2 - \left\| f(x_i^a) - f(x_i^n) \right\|_2^2 + \alpha \right]_+ \quad (2)$$

而在训练过程中，选择符合式(1)的 triplet 对来训练对网络的收敛没有任何增益，因此，需要选择违反式(1)的 triplet 对来使网络收敛，即：

$$\left\| f(x_i^a) - f(x_i^p) \right\|_2^2 > \left\| f(x_i^a) - f(x_i^n) \right\|_2^2 \quad (3)$$

训练模型如图 2 所示。通过在多个 batch 上挑选符合式(3)的 triplet 对来进行训练，能够避免在训练早期陷入局部最小值，同时能够加快网络的收敛。

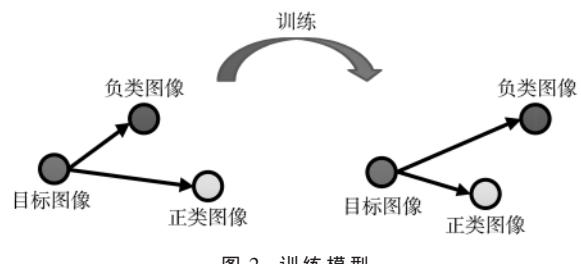


图 2 训练模型

2 深度神经网络模型

本文所采用的神经网络模型为改进的 SqueezeNet 网络模型。SqueezeNet 网络是于 2016 年提出的一种轻量化的网络模型^[10]，相较于 AlexNet 具有更少的参数同时，还能够保持相同等级的准确度。本文对 SqueezeNet 网络进行改进，并与 FaceNet 结合，用于伪装人脸的身份识别。

2.1 SqueezeNet 网络模型简介

SqueezeNet 网络是针对在较少的参数数量又要保证一定的准确度的情况下所提出的。根据卷积层的参数=输入通道数×输出通道数×卷积核尺寸，提出了 3 个策略来尽可能减少卷积层中的参数数量，分别为：(1) 使用尽可能多的 1×1 的卷积核来替换 3×3 的卷积核；(2) 减少 3×3 卷积核的输入通道数；(3) 将池化操作滞后，在网络后期使用池化层。

其中，策略(1)和(2)是为了减少卷积层的参数数量，而策略(3)是为了在参数数量受限的情况下，尽可能提升识别准确度，因为池化层会大幅减少输出特征图的尺寸，而将池化层滞后能够获得更大的输出特征图，从而保留更多的卷积层提取到的特征。

根据这些策略，制定了 Fire 模块来进行网络的搭建。Fire 模块的结构如图 3 所示，包含 2 个部分：压缩块和扩展块。其中，压缩块内均为 1×1 的卷积核，并具有超参数 $s_{1 \times 1}$ ，表示 1×1 卷积核的个数；扩展块内包含 1×1 和 3×3 的卷积核，并具有超参数 $e_{1 \times 1}$ 和 $e_{3 \times 3}$ ，分别表示 1×1 卷积核的个数以及 3×3 卷积核的个数。为了保证 1×1 和 3×3 的卷积核的输出特征图具有相同的尺寸，在计算过程中，对 1×1 的卷积核采用 0 填充扩展为 3×3 的

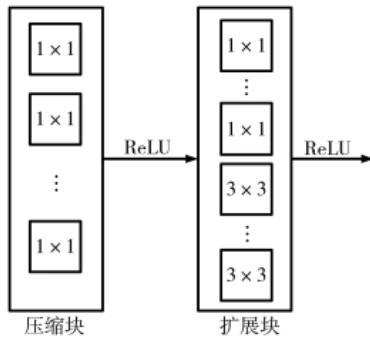


图 3 Fire 模块结构

卷积核。

2.2 改进的 SqueezeNet 网络结构

本文所使用的神经网络模型 SqueezeNet-D 网络为 SqueezeNet 网络的改进版。考虑到本文的目标,进行了如下改进:由于本文的输入图像是检测后裁剪的,尺寸相对较小,因此在 CONV1 层采用较小的 5×5 的卷积核替代 7×7 的卷积核来提取特征,这样能减小训练中过拟合的风险;将池化层进一步滞后,以此来获得更大的输出特征图,从而保留更多的特征;并将网络与 FaceNet 架构进行结合,添加全连接层,将网络的输出维度变为 128,从而更好地进行后续的正则化、嵌入等运算。

网络结构如图 4 所示。以 160×160 的 3 通道输入图像为例,经过具有 96 个 5×5 卷积核、步长为 2 的卷积层后,输出特征图大小为 $79 \times 79 \times 96$;池化层的参数含义和卷积层相同。之后经过 2 组连续的 4 个 Fire 模块和池化层,其中 $s_{1 \times 1}$ 、 $e_{1 \times 1}$ 和 $e_{3 \times 3}$ 的含义已在上文给出;末端再通过卷积层和池化层将输出的尺寸转为 1×1 ,输出通道数为 1000,最后在网络末端连接一个全连接层,将输出维度转换为 128。

3 实验

本实验所使用的计算机配置为: Intel-酷睿 i5-8500 的 CPU, 8 GB DDR3 内存, GTX1660Ti 的 GPU, WIN10 操作系统, 程序在 TensorFlow 框架下运行。

3.1 数据集及处理

本文所使用的数据集为 CASIA-FaceV5 亚洲人脸数据集,涵盖男女老少各个年龄层,包含 500 个人,每个人有 5 张样本图像。通过对每张图像进行处理,添加墨镜、帽子、围巾等配件,构建了相同图像数量的伪装人脸数

据集,效果如图 5 所示。使用 MTCNN 进行人脸检测及裁剪^[11],将图像中的人脸部分提取出来,分别裁剪出 64×64 、 128×128 和 160×160 这 3 种尺寸的图像。将数据集中编号为 000~399 共 400 个身份的图像作为训练集,400~499 共 100 个身份的图像作为测试集。

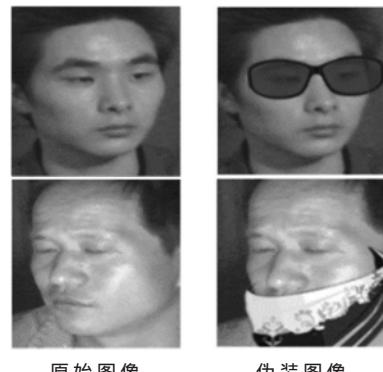


图 5 部分数据集

3.2 超参数设置及性能指标

本文的主要超参数设置如表 1 所示。

而对于性能指标,本文使用准确率来对网络的输出结果进行性能分析。由于测试图像对可能来自同一身份或不同身份,记输入的测试对为 (x_i, x_j) ,所有来自同一身份的测试对为 P_{same} ,所有来自不同身份的测试对为 P_{diff} ,阈值为 d ,网络计算出的测试对之间的欧氏距离之差为 $D(x_i, x_j)$;将网络正确识别来自同一身份的测试对的集合记为正确接受 $\text{TA}(d)$,错误识别来自同一身份的测试对的集合记为错误接受 $\text{FA}(d)$;将网络正确识别来自不同身份的测试对的集合记为正确拒绝 $\text{TR}(d)$,错误识别来自不同身份的测试对的集合记为错误拒绝 $\text{FR}(d)$,分别表示为:

$$\text{TA}(d) = \{(x_i, x_j) \in P_{\text{same}}, D(x_i, x_j) \leq d\} \quad (4)$$

$$\text{FA}(d) = \{(x_i, x_j) \in P_{\text{same}}, D(x_i, x_j) > d\} \quad (5)$$

$$\text{TR}(d) = \{(x_i, x_j) \in P_{\text{diff}}, D(x_i, x_j) > d\} \quad (6)$$

$$\text{FR}(d) = \{(x_i, x_j) \in P_{\text{diff}}, D(x_i, x_j) \leq d\} \quad (7)$$

那么,网络模型的识别准确率 $\text{VAL}(d)$ 为:

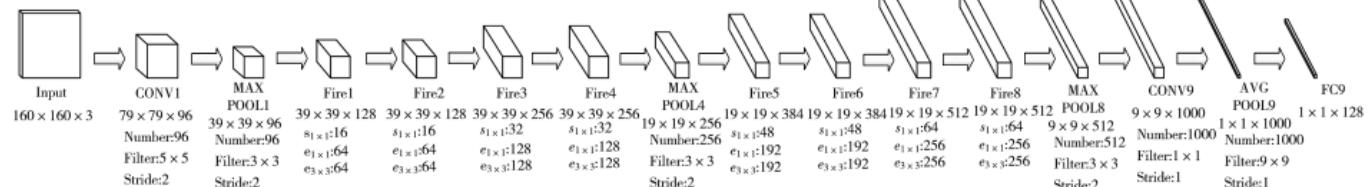


图 4 改进后的 SqueezeNet-D 网络

$$VAL(d) = \frac{|TA(d)| + |TR(d)|}{|P_{\text{same}}| + |P_{\text{diff}}|} \quad (8)$$

3.3 实验结果

本文在测试集中构建了6 000对测试对,其中来自同一身份和来自不同身份的测试对各3 000对,而各自又包含等量的原始-原始图像对、原始-伪装图像对和伪装-伪装图像对各1 000对。在训练时,分别对原始数据集和加入伪装图像后的伪装数据集进行训练,测试了原始的SqueezeNet模型和改进后的SqueezeNet-D的识别准确率,在不同图像尺寸下的测试结果如表2所示。

表2 不同情况下的识别率

网络模型	数据集	图像大小/pix	识别准确率/%
SqueezeNet	原始数据集	64×64	75.86
SqueezeNet	伪装数据集	64×64	79.62
SqueezeNet-D	原始数据集	64×64	76.73
SqueezeNet-D	伪装数据集	64×64	80.75
SqueezeNet	原始数据集	128×128	78.13
SqueezeNet	伪装数据集	128×128	84.23
SqueezeNet-D	原始数据集	128×128	79.15
SqueezeNet-D	伪装数据集	128×128	85.85
SqueezeNet	原始数据集	160×160	81.67
SqueezeNet	伪装数据集	160×160	87.63
SqueezeNet-D	原始数据集	160×160	83.01
SqueezeNet-D	伪装数据集	160×160	89.48

本文还测试了在FaceNet架构下,加入伪装后的数据集在不同网络模型下的准确率,在图像尺寸为160×160的情况下,测试结果如表3所示。

表3 不同网络模型的识别率

网络模型	识别准确率/%
SqueezeNet-D	89.48
SqueezeNet	87.63
AlexNet	77.98
LeNet-5	72.35

由上述结果可以看出,在相同图像尺寸及网络模型下,在训练数据集中加入伪装人脸图像后,识别准确率能够提升4%~6%;因为通过训练,使得网络能够学习到伪装的特征。而在相同情况下,增大图像的尺寸也能不同程度地提高识别准确率,因为图像包含的有用信息增多。同时,本文还对比了不同网络模型下的识别率大小;结果表明,经过改进后的SqueezeNet-D网络模型相比原始的模型准确率能够提升1%~2%,相较于LeNet-5网络模型以及AlexNet网络模型,准确率有较大的提升。

4 结论

本文提出了一种改进的SqueezeNet网络模型,并将其与FaceNet架构进行结合,应用于伪装人脸图像的身份识别。通过对原始数据集进行处理,构建了伪装人脸数据集,并在训练集中加入了带有伪装的人脸图像,使得

网络能够学习到伪装的特征。实验结果表明,相较于其他的神经网络模型,本文所提出的方法识别准确率更高,具有更好的性能。

参考文献

- [1] RAGASHE M U , GOSWAMI M M , RAGHUVANSI M M . Approach towards real time face recognition in streaming video under partial occlusion[C].2015 IEEE 9th International Conference on Intelligent Systems and Control(ISCO).IEEE, 2015 : 1–7.
- [2] WU C Y , DING J J . Occlusion pattern-based dictionary for robust face recognition[C].2016 IEEE International Conference on Multimedia and Expo(ICME).IEEE , 2016 : 1–6.
- [3] MIN R , HADID A , DUGELAY J L . Improving the recognition of faces occluded by facial accessories[C].Face and Gesture 2011.IEEE , 2011 : 442–447.
- [4] MC LAUGHLIN N , MING J , CROOKES D . Largest matching areas for illumination and occlusion robust face recognition[J]. IEEE Transactions on Cybernetics , 2016 , 47(3) : 796–808.
- [5] 姚宇晨,彭虎.基于深度学习的通信信号自动调制识别技术[J].电子技术应用,2019,45(2):12–15.
- [6] 仇晓松,邹旭东,王金戈,等.基于卷积神经网络的视觉位置识别方法[J].计算机工程与设计,2019,40(1):223–229.
- [7] 牛硕硕,柴小丽,李德启,等.一种基于神经网络与LDA的文本分类算法[J].计算机工程,2019,45(10):208–214.
- [8] 马辉,车迪,牛强,等.基于深度神经网络的提升机轴承故障诊断研究[J].计算机工程与应用,2019,55(16):123–129, 184.
- [9] SCHROFF F , KALENICHENKO D , PHILBIN J . Facenet : a unified embedding for face recognition and clustering[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition , 2015 : 815–823.
- [10] IANDOLA F N , Han Song , MOSKEWICZ M W , et al . SqueezeNet : AlexNet-level accuracy with 50x fewer parameters and<0.5 MB model size[J].arXiv Preprint , arXiv : 1602.07360[cs.CV] , 2016.
- [11] Zhang Kaipeng , Zhang Zhanpeng , Li Zhifeng , et al . Joint face detection and alignment using multitask cascaded convolutional networks[J].IEEE Signal Processing Letters , 2016 , 23(10) : 1499–1503.

(收稿日期:2019-12-02)

作者简介:

张润生(1994-),通信作者,男,硕士研究生,主要研究方向:深度学习与计算机视觉。

贺超(1990-),男,博士研究生,主要研究方向:光纤无线通信网络。

况朝青(1996-),男,硕士研究生,主要研究方向:深度学习与计算机视觉。

版权声明

经作者授权，本论文版权和信息网络传播权归属于《电子技术应用》杂志，凡未经本刊书面同意任何机构、组织和个人不得擅自复印、汇编、翻译和进行信息网络传播。未经本刊书面同意，禁止一切互联网论文资源平台非法上传、收录本论文。

截至目前，本论文已经授权被中国期刊全文数据库（CNKI）、万方数据知识服务平台、中文科技期刊数据库（维普网）、DOAJ、美国《乌利希期刊指南》、JST 日本科技技术振兴机构数据库等数据库全文收录。

对于违反上述禁止行为并违法使用本论文的机构、组织和个人，本刊将采取一切必要法律行动来维护正当权益。

特此声明！

《电子技术应用》编辑部

中国电子信息产业集团有限公司第六研究所