

# 基于深度学习的抗年龄干扰人脸识别

何星辰<sup>1</sup> 郭勇<sup>1</sup> 李奇龙<sup>1</sup> 高唱<sup>2</sup>

**摘要** 随着年龄的增长,人脸的形状、纹理等特征会随之发生较明显的改变而造成显著的类内干扰,这使得人脸识别的性能大大降低.为了解决上述问题,本文基于深度卷积神经网络将年龄估计任务和人脸识别任务相结合,提出了一种抗年龄干扰的人脸识别新方法 AD-CNN (Age decomposition convolution neural network),首先将卷积块注意力模型 (Convolutional block attention module, CBAM) 嵌入到残差网络中以学习更具有代表性的面部特征,随后利用线性回归指导年龄估计任务,提取出年龄干扰因子,通过多层感知机将整个面部特征与年龄干扰特征投影到同一线性可分空间,最后从面部稳定的特征中将年龄干扰分离,得到与年龄无关的面部特征,并采用改进后的角度损失函数基于年龄无关的身份特征进行人脸识别任务,从而达到抑制年龄干扰的目的.本文在 MORPH 和 FGNET 数据集上的识别正确率分别达到了 98.93% 和 90.0%,充分证实了本文所提方法的先进性和有效性.

**关键词** 人脸识别, 年龄干扰, 深度学习, 年龄估计, 卷积神经网络注意力模型

**引用格式** 何星辰, 郭勇, 李奇龙, 高唱. 基于深度学习的抗年龄干扰人脸识别. 自动化学报, 2022, 48(3): 877-886

**DOI** 10.16383/j.aas.c190256

## Age Invariant Face Recognition Based on Deep Learning

HE Xing-Chen<sup>1</sup> GUO Yong<sup>1</sup> LI Qi-Long<sup>1</sup> GAO Chang<sup>2</sup>

**Abstract** Facial appearances such as shape and texture are subject to significant intra-class variations caused by the aging process over time, resulting in the performance reduction of face recognition. To overcome this problem, this paper proposes a novel method (age decomposition convolution neural network, AD-CNN) based on deep convolution neural network to learn age-invariant face features. Firstly, the AD-CNN utilizes convolutional block attention module (CBAM) to extract facial features and estimates age factors by linear regression. Then, the facial features and age factors are projected into the same linear separable space by multi-layer perceptron. Finally, the age-invariant face features can be obtained by separating age factors from the whole facial features. Here, the improved angle loss function is considered to guide the training process. The proposed AD-CNN achieves 98.93%, and 90.0% recognition accuracy on MORPH and FGNET datasets, respectively, which demonstrates the AD-CNN with a great potential for age-invariant face recognition.

**Key words** Face recognition, age interference, deep learning, age estimation, convolution neural network attention model

**Citation** He Xing-Chen, Guo Yong, Li Qi-Long, Gao Chang. Age invariant face recognition based on deep learning. *Acta Automatica Sinica*, 2022, 48(3): 877-886

人脸识别是一个极具挑战但又非常重要的问题,因为其广泛的应用场景而受到越来越多的关注,目前一些一般人脸识别方法<sup>[1-6]</sup>在几个基准测试中已经取得了理想的成绩,但由于人脸本身存在的较大的类内干扰(如表情、光照、年龄等)使得这项工

作面临极大的困难.其中年龄变化所引起的类内变化仍然是许多实际应用的主要瓶颈,如寻找多年前的失踪人口、排查潜逃的犯罪嫌疑人、证件的注册及认证工作等.相较于其他干扰,年龄相关的变化取决于内因(如压力和疾病)和外因(如生活环境和生活方式),以及随着年龄的增长,脸部的纹理和形状也会发生显著改变,这使得学习不变的模式变得困难.因此,基于相同年龄的的一般人脸识别模型在处理跨年龄识别任务时无法得到理想的结果,2015年Chen等<sup>[7]</sup>的研究显示在跨年龄识别任务中一般人脸识别模型的正确率会降低13%以上.生成模型和判别模型是现有的两种抗年龄干扰的人脸识别方法.生成模型通过将面部合成到一个或多个年龄类别来进行建模,然后利用人工面部表示进行识

收稿日期 2019-03-28 录用日期 2019-09-24

Manuscript received March 28, 2019; accepted September 24, 2019

国家自然科学基金(41574136)资助

Supported by National Natural Science Foundation of China (41574136)

本文责任编辑 刘青山

Recommended by Associate Editor LIU Qing-Shan

1. 成都理工大学信息科学与技术学院 成都 610051 2. 成都理工大学地球物理学院 成都 610051

1. College of Information Science and Technology, Chengdu University of Technology, Chengdu 610051 2. College of Geophysics, Chengdu University of Technology, Chengdu 610051

别<sup>[8-10]</sup>, 随着深度学习的迅速发展, 生成模型在抗年龄干扰识别上更有前景<sup>[11-13]</sup>. 然而生成模型的识别过程通常分为两步, 无法以端到端的方式优化识别性能, 并且通过生成模型生成的面部图像通常不稳定, 会引入额外的噪声到合成图相中从而对识别过程形成干扰. 因此, 对判别模型的研究成为近些年来的热点, 本文本文提出的深度模型也是属于判别模型的范畴.

判别模型旨在对人脸特征进行分离, 从剔除年龄相关特征, 提取年龄无关的身份特征的角度提高人脸识别的稳定性. Ling 等<sup>[14]</sup> 使用梯度方向金字塔 (Gradient orientation pyramid, GOP) 作为年龄不变的特征并采用支持向量机 (Support vector machine, SVM) 作为面部识别的分类器. 许多判别模型在设计适当的特征和有效的匹配系统两个方面做了深入研究, 但大多数设计的特征仍包含干扰信息, 为了将特定于人的身份因素与年龄因素分开, Gong 等<sup>[15-16]</sup> 提出了隐藏因子分析方法 (Hidden factor analysis, HFA), 将人的面部图像表示为在衰老过程中稳定的身份特定成分和反映衰老效果的其他成分的组合, 表示身份的特征成分用于年龄不变的面部识别, 他们采用期望最大化 (Expectation maximization, EM) 算法来估计模型参数, 并进一步提出最大熵特征. Li 等<sup>[17]</sup> 提出了区域模式选择 (Local patterns selection, LPS) 作为跨年龄人脸识别的新特征描述符. 文献 [18] 提出了一种称为跨年龄参考编码 (Cross age reference code, CARC) 的编码框架, CARC能够使用年龄不变的参考空间对人脸图像的低级特征进行编码. 深度学习因其在原始特征中直接学习一系列非线性特征映射函数的卓越性能而受到广泛关注, 相比之下, 深度学习模型通常通过使用单个或几个具有集中特定损失函数的年龄模型来处理年龄差异<sup>[19-22]</sup>. Wen 等<sup>[21]</sup> 提出一种潜在因子引导卷积神经网络 (Latent factor guided convolutional neural network, LF-CNN) 模型来学习年龄不变的深层特征. Xu 等<sup>[22]</sup> 提出了一种非线性方法来分离整个特征以获得身份特征, 并提出称为耦合自动编码器网络 (Coupled auto-encoder network, CAN) 的神经网络模型, 它利用两个浅层神经网络桥接两个自动编码器以适应老化和衰老过程. Wang 等<sup>[20]</sup> 提出正交嵌入 CNN (Orthogonal embedding convolutional neural network, OE-CNN) 模型以将深度面部表征分解成两个正交分量以表示年龄和身份特定特征. 本文基于深度学习提出了一种新的抗年龄干扰的人脸识别模型 AD-

CNN (Age decomposition convolution neural network), 如图 1 所示. 给出面部图像年龄标签和身份标签, AD-CNN同时实现年龄估计任务和人脸识别任务, 第 1 个全连接层输出包括年龄信息在内的所有面部特征, 第 2 个全连接层用于输出通过年龄回归任务所提取到的年龄特征, 并将其送入含有一层隐藏层的多层感知机 (Multilayer perceptron, MLP) 网络最终得到年龄相关因子, 通过从整个特征中减去年龄因子获得年龄不变特征. 此外, AD-CNN 在基线网络中采用了空间注意力机制, 重新建模了特征提取过程中通道和空间的相互依赖关系, 增强了特征提取性能.

## 1 抗年龄干扰的深度人脸识别模型

跨年龄识别任务的困难在于年龄的变化会导致较大的类内变化 (如面部形态和纹理), 但通过观察我们可以发现, 同一年龄段的不同人的面部图像通常具有相同的特征 (如皱纹和皮肤), 此外同一个人在不同的年龄段始终会有如性别等部分特征保持相对稳定的状态. 因此, 本文的目标是基于深度学习设计一种有效的方法, 以便从参有年龄信息的混合的特征中有效地学习年龄不变的成分.

Gong 等<sup>[15]</sup> 和 Wen 等<sup>[21]</sup> 的研究通过迭代和学习的方法将所提取到的混合特征线性分解为年龄特征和抗年龄干扰的身份特征两部分; Wang 等<sup>[20]</sup> 则将混合特征正交分解为年龄特征和身份特征的乘积, 并分别对这两部分赋予年龄标签和身份标签进行联合学习. 受上诉方法的启发, 本文延用了文献 [12, 18] 的思想, 将深度卷积神经网络提取到的原始混合特征看作是年龄特征与年龄无关的身份特征在高维空间的线性组合, 并利用多层感知机能够学习任何映射函数的这一数学特性<sup>[23]</sup> 在线性空间对混合特征进行了分离. 具体来说, 首先年龄估计任务可以从混合特征中提取出年龄特征, 随后通过含有一层隐藏层的多层感知机将年龄特征映射回原始混合特征所在的高维线性空间, 最后从混合特征中减去年龄特征带来的干扰得到年龄无关的身份特征从而完成身份分类的任务, 整个过程可表示为

$$\mathbf{y} = f(\mathbf{t} - h(\mathbf{x})) \quad (1)$$

$$f(\mathbf{x}) = \text{relu}(W_1\mathbf{x} + b_1) \quad (2)$$

$$h(\mathbf{x}) = \text{relu}(W_3[\text{relu}(W_2\mathbf{x} + b_2)] + b_3) \quad (3)$$

其中,  $\mathbf{t}$ ,  $\mathbf{x}$ ,  $\mathbf{y}$  都是 128 维向量,  $\mathbf{t}$  表示输入图像的特征,  $\mathbf{x}$  表示从年龄估计中获得的年龄特征,  $\mathbf{y}$  表示与年龄无关的人脸潜在特征,  $W_1$ ,  $W_2$ ,  $W_3$  和

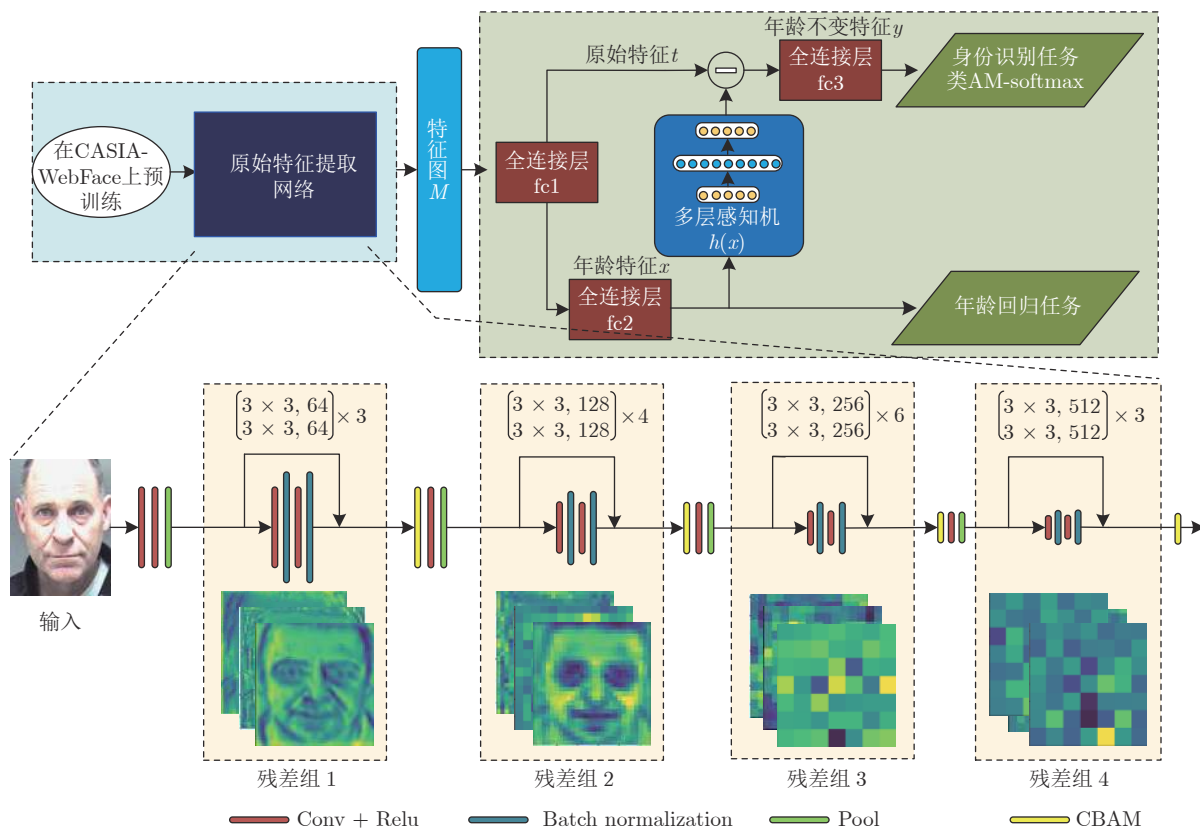


图 1 AD-CNN 模型流程图

Fig. 1 The architecture of the proposed AD-CNN

$b_1, b_2, b_3$  分别表示全连接层 fc3 和多层感知机网络的权重参数和偏置项,  $relu$  表示非线性激活函数.  $h(x)$  是获得年龄因子的函数, 全连接层 fc3 的输出对应式 (1) 中的  $y$ , 它能更好地处理整个面部特征、年龄特征和不受年龄干扰的人脸特征之间的关系. 在模型中, 人脸识别任务和年龄估计任务同时进行.

### 1.1 CNN 原始特征提取网络

随着 CNN 层数逐渐加深, 所能提取到图像的特征也越来越丰富, 但如果简单地增加深度, 会带来梯度弥散或梯度爆炸的问题. He 等<sup>[24]</sup> 在 2015 年提出了建立在快捷连接和恒等映射基础上的残差网络 (Residual networks, ResNets) 结构, 解决了 CNN 随着深度增加而带来的网络退化问题.

该网络的基本组成单元为残差块, 如图 2 所示, 在残差块中, 上层输出做为下层输入的一部分, 即在原始卷积层外部加上了跃层连接支路, 使得原始映射可以表示为  $H(x) = F(x) + x$ , 此时可以将对  $H(x)$  的学习转换为对残差函数  $F(x) = H(x) - x$  的学习, 使得网络更容易拟合. 本文采用了类似 ResNets-34<sup>[24]</sup> 的模型作为 CNN 原始特征提取网络, 该模型由 4 组残差块组成, 每个组的输入到输出接有

一额外的支路映射 shortcut, 每个组中分别包含 3, 4, 6, 3 个残差块, 每个残差块由 2 维卷积层 (Conv), 批量归一化层 (Batch normalization)<sup>[25]</sup>, 非线性激活层 (Relu) 组成, 并且相同的组中残差块的输出通道数相同, 分别为 64, 128, 256 和 512. 每个残差块组之间连接有池化层 (Pool) 和卷积块注意力模型 (Convolutional block attention module, CBAM) 以增强模型的鲁棒性及特征提取能力. 整个基线网络的计算过程可以表示为

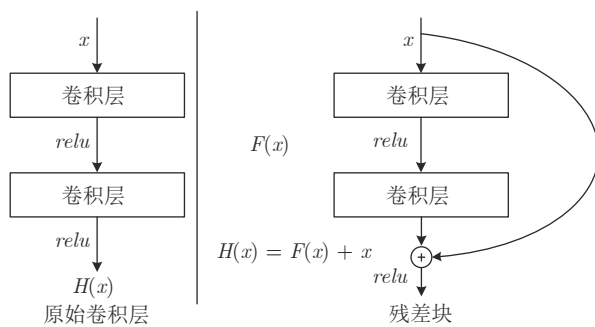


图 2 残差块的结构

Fig. 2 The structure diagram of residual block

$$\begin{cases} x_{3\text{out}} = M_{\text{CBAM}}(g(x_{1\text{in}}) + x_{3\text{in}} + \\ \quad F(x_{3\text{in}}, W_c^3)) \\ x_{7\text{out}} = M_{\text{CBAM}}(g(x_{3\text{out}}) + x_{7\text{in}} + \\ \quad F(x_{7\text{in}}, W_c^7)) \\ x_{13\text{out}} = M_{\text{CBAM}}(g(x_{7\text{out}}) + x_{13\text{in}} + \\ \quad F(x_{13\text{in}}, W_c^{13})) \\ x_{16\text{out}} = M_{\text{CBAM}}(g(x_{13\text{out}}) + x_{16\text{in}} + \\ \quad F(x_{16\text{in}}, W_c^{16})) \end{cases} \quad (4)$$

其中,  $x_{l\text{in}}$  与  $x_{l\text{out}}$  表示第  $l$  个残差块的输入和输出,  $F(\cdot)$  表示残差映射函数,  $g(x_{l\text{in}})$  表示每个残差组的 shortcut 支路映射,  $M_{\text{CBAM}}(\cdot)$  表示注意力模型,  $W_c^l$  表示第  $l$  个残差块的权重参数.

## 1.2 网络注意力模型

为了增强网络模型的特征表达能力, 提升特定区域的表征, 本文在特征提取过程中引入卷积块注意力模型 (CBAM)<sup>[26]</sup> 以提高深度模型对有用信息的敏感度. CBAM 模块将注意力机制同时运用在通道和空间两个维度上, 通过建模卷积特征通道和空间的相互依赖性来提高网络的特征提取能力. 在特征空间和通道进行校准的机制, 使网络从全局信息出发来提升有价值的特征并且抑制对任务无用的特征. CBAM 模块示意图如图 3 所示. 将一个经过卷积变换后的特征图:  $F \in \mathbf{R}^{H \times W \times C}$  作为输入,  $M$ ,  $H$ ,  $C$  代表特征图的三维信息, 特征图  $F$  经过 1 维通道注意力模块  $M_c \in \mathbf{R}^{1 \times 1 \times C}$  和 2 维空降注意力模块  $M_s \in \mathbf{R}^{H \times M \times 1}$  后得到最终输出  $F''$ , 整个过程可表示为

$$F' = M_c(F) \otimes F \quad (5)$$

$$F'' = M_s(F') \otimes F' \quad (6)$$

其中,  $\otimes$  表示 element-wise 的点乘操作, 将特征图  $F$  送入通道注意力模块  $M_c$ , 并将其结果与  $F$  进行点乘得到  $F'$ , 再将  $F'$  送入空间注意力模块  $M_s$ , 将其结果与  $F'$  进行点乘从而得到最终的输出  $F''$ .

通道注意力模块首先利用平均池化和最大池化操作聚合特征图在每个通道上的信息, 生成两个不

同的通道上下文描述符  $F_{\text{avg}}^c$  和  $F_{\text{max}}^c$ , 即

$$F_{\text{avg}}^c = \frac{1}{M \times H} \sum_{i=1}^M \sum_{j=1}^H u_c(i, j) \quad (7)$$

$$F_{\text{max}}^c = \max(u_c(i, j)), \quad i \in [1, M], \quad j \in [1, H] \quad (8)$$

其中,  $F_{\text{avg}}^c$  表示第  $c$  个通道上的平均池化特征,  $F_{\text{max}}^c$  表示第  $c$  个通道上的最大池化特征,  $M$ ,  $H$  分别表示特征图的长和宽,  $u_c$  表示第  $c$  个通道, 然后将两个描述符分别送入具有一个隐藏层的多层感知机网络 (MLP), 为了减少参数开销, 隐藏层的神经元个数设置为  $N/2$ ,  $N$  表示特征图通道的数量. 最后通过逐元素求和来合并输出特征权重参数, 计算过程表示为

$$M_c(F) = \sigma(\text{mlp}(P_{\text{Avg}}(F)) + \text{mlp}(P_{\text{Max}}(F))) = \sigma(w_1(w_0(F_{\text{avg}}^c)) + w_1(w_0(F_{\text{max}}^c))) \quad (9)$$

其中,  $P_{\text{max}}(\cdot)$  表示最大池化操作,  $P_{\text{avg}}(\cdot)$  表示平均池化操作,  $w_0 \in \mathbf{R}^{N/2 \times N}$ ,  $w_1 \in \mathbf{R}^{N \times N/2}$ , 表示多层感知机的权重, 并且在  $w_0$  后接有一个 ReLU 激活函数,  $\sigma$  表示 Sigmoid 激活函数.

空间注意力模块利用特征的空间关系生成空间注意力图层, 空间注意力集中在“哪里”是信息部分, 这是对通道注意力的补充. 为了计算空间注意力, 首先沿通道轴应用平均池化和最大池化操作, 并将它们连接起来以生成有效的特征描述符. 通过两个池化操作融合特征图的通道信息, 生成两个 2 维特征图  $F_{\text{avg}}^s \in \mathbf{R}^{H \times M \times 1}$ ,  $F_{\text{max}}^s \in \mathbf{R}^{H \times M \times 1}$ , 即

$$F_{\text{avg}}^s = \frac{1}{N} \sum_{c=1}^N s_{ij}(c) \quad (10)$$

$$F_{\text{max}}^s = \max(s_{ij}(c)), \quad c \in [1, N] \quad (11)$$

其中,  $N$  表示特征的通道数量,  $s_{ij}(c)$  表示第  $c$  个特征图中位于坐标  $(i, j)$  处的元素,  $F_{\text{avg}}^s$  和  $F_{\text{max}}^s$  分别表示每个通道上的平均融合特征和最大融合特征. 然后合并这两个 2 维特征图并送入卷积层, 最终得到空间维度上的注意力权重, 其计算过程为

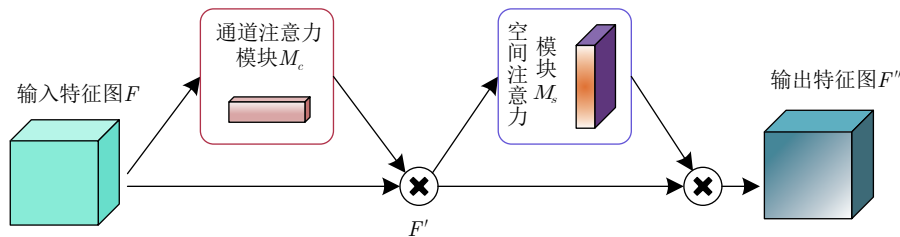


图 3 CBAM 注意力模块示意图

Fig. 3 The overview of CBAM attention module

$$Ms(F) = \frac{\sigma(f^{7 \times 7}([P_{avg}(F); P_{max}(F)]))}{\sigma(f^{7 \times 7}([F_s^{avg}; F_s^{max}]))} \quad (12)$$

其中,  $\sigma$  表示 sigmoid 激活函数,  $f^{7 \times 7}$  表示卷积核大小为 7 的卷积操作. 图 4 显示了两个注意力子模块内部的处理流程.

### 1.3 多任务联合学习方法

由于老化过程 (例如面部形貌变化、纹理变化等) 会引起不同年龄类别 (类内变化) 中相同个体的变化, 以及从 CNN 中提取的深度特征中不可避免地会出现无关成分, 这都会给人脸识别任务带来巨大挑战. 因此本文提出年龄估计任务和人脸识别任务联合学习的方法来克服上述困难.

为了挖掘年龄信息的内在表示, 本文将线性回归的方法应用在年龄估计任务中, 如图 1 所示, 将全连接层 fc2 所得的年龄特征  $X_i[x_1^i, x_2^i, \dots, x_{128}^i]$  作为输入, 回归损失表示为

$$L_{age} = \frac{1}{2M} \sum_{i=1}^M \|f(X_i) - z_i\|_2^2 \quad (13)$$

其中,  $M$  表示每一个批次中所含样本的数量,  $z_i$  表示第  $i$  个样本所对应的年龄标签,  $\|\cdot\|_2^2$  表示  $L_2$  范数,  $f(x)$  是关联  $X_i$  和  $z_i$  的映射函数, 本文使用的是线性多项式  $f(x) = \sum_{i=1}^{128} w_i \times x_i + b_i$ .

在进行面部识别时, 将分类边界投影到角度空间, 采用 A-Softmax<sup>[27]</sup> 类似的损失函数增大类间距离, 减小类内距离. 先对全连接层 fc3 输出的面部特征和权重向量做归一化处理, 使它们的  $L_2$  范数都为 1, 此时损失函数可写为

$$L_{id} = \frac{1}{M} \sum_{i=1}^M -\lg \left( \frac{e^{s \times \phi(\theta_{y_i, i})}}{e^{s \times \phi(\theta_{y_i, i})} + \sum_{j \neq y_i} e^{s \times \cos(\theta_{j, i})}} \right) \quad (14)$$

其中,  $\phi(\theta_{y_i, i}) = (-1)^k \cos(m\theta_{y_i, i}) - 2k$ ,  $\theta_{y_i, i}$  是第  $i$  个样本的面部特征  $F_i$  与第  $y_i$  个类别对应的权重向量之间的夹角,  $\theta_{y_i, i} \in [k\pi/m, (k+1)\pi/m]$ ,  $k \in [0,$

$m-1]$ ,  $m \geq 1$  是控制角度大小的超参数.  $s > 0$  是一个可调整的比例因子, 用于补偿 Softmax 的学习. 从几何角度来看, 式 (14) 增加了一个约束, 该约束保证特征  $x$  的角度与其相应的权重向量应小于特征  $x$  与任何其他权重向量之间的角度的  $1/m$ , 因此, 可以增加两个任意类之间的余量, 与原始的 A-Softmax 相比, 式 (14) 引入了一个新的自由变量  $s$  以补偿  $L_2$  规范化的损失.

综上, 合并式 (13) 和式 (14) 以形成多任务联合学习的损失函数, 即

$$L = L_{id} + \lambda L_{age} \quad (15)$$

其中,  $\lambda$  是平衡两损失的超参数.

## 2 实验

### 2.1 数据集

本文在 MORPH Album 2<sup>[28]</sup> 和 FGNET<sup>[29]</sup> 两个公共领域跨年龄人脸基准数据集上设置了实验, 以证明所提方法的有效性. MORPH Album 2 数据集包含了超过 78 000 幅人脸图像, 20 000 个身份标识 (id), 年龄范围从 16 岁到 77 岁, 平均每个身份 id 对应 4 幅图片, 每个 id 的年龄跨度较小. FGNET 数据集由来自于 82 个身份 id 的 1 002 幅图像组成, 样本数相对较小, 但平均每个身份 id 对应 12 幅人脸图像且同一人的年龄跨度较大. 两个数据集的年龄分布如图 5 所示.

### 2.2 数据处理

使用 MTCNN (Multitask cascaded convolutional neural network)<sup>[30]</sup> 检测图像中的面部和面部关键点, 然后应用相似性变换, 根据 5 个面部关键点 (两个眼睛、鼻子和两个嘴角) 进行人脸对齐, 并将面部贴片裁剪为  $130 \times 152$  像素. 为了避免出现过拟合, 训练时本文对输入图像进行随机裁剪和随机翻转的操作, 随机裁剪后图像大小为  $112 \times 112$  像素. 最后将图像归一化到  $[-1, 1]$  区间, 如图 6 所示. 图 6(a)

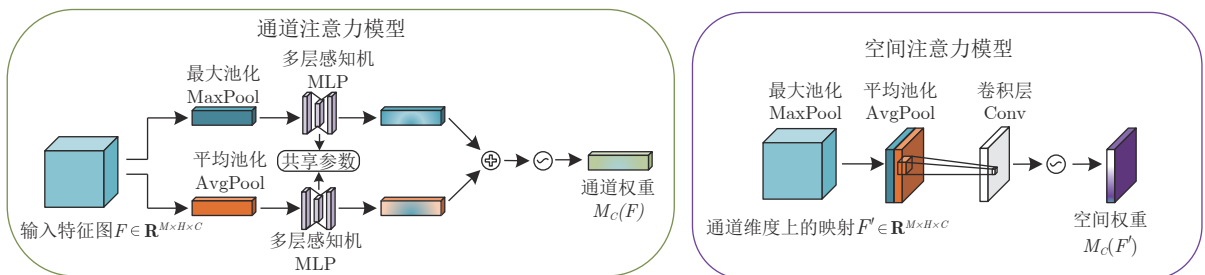


图 4 注意力子模块示意图  
Fig. 4 Diagram of each attention sub-module

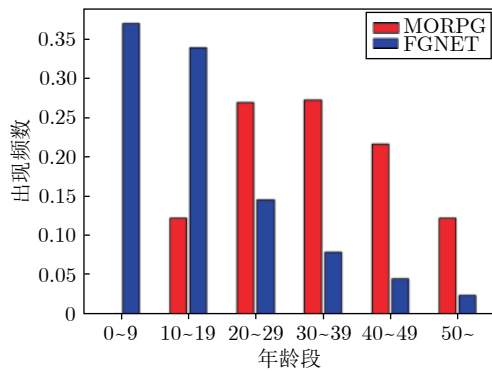


图 5 MORPH 和 FGNET 数据集的年龄分布

Fig.5 Age range distribution of MORPH and FGNET

和图 6(c) 为来自于数据集中的原始图像, 图 6(b) 和图 6(d) 为处理后的图像.



图 6 MORPH Album 2 中处理后的图像

Fig.6 Processed images of MORPH Album 2 dataset

### 2.3 实验设置

考虑到 MORPH Album 2 和 FGNET 的规模都不大, 为防止过拟合, 本文将 CNN 原始特征提取模型在 CASIA-WebFace<sup>[23]</sup> 上进行预训练, 实验时用跨年龄数据集进行微调. 根据文献 [27, 31] 的结果, 将式 (13) 中提到的超参数设置为 4, 32. 对于权重因子  $\lambda$ , 选取 0.1, 0.3, 0.5, 0.7 四个值分别进行实验. 所有模型均采用 Tensorflow 框架进行训练, 采用随机梯度下降 (Stochastic gradient descent, SGD) 算法进行优化. 训练批量大小设置为 16, 迭代次数设置为 25 个 epoch. 动量因子为 0.9, 初始学习率设置为 0.001, 并且当损失变得稳定时 (在第 8, 16 和 20 个迭代周期), 训练过程自适应地降低学习率 3 次.

### 2.4 FGNET 数据库上的实验结果

FGNET 包含来自 82 个人的 1002 幅面部图像, 年龄范围从儿童到老年人的跨度很大. 由于 FGNET 的规模较小, 本文采用文献 [13, 15] 中的留一法设置, 以便与之前的方法进行公平比较. 选取 1 幅图像作为测试样本, 剩下 1001 幅图像作为训练集对模型进行微调. 重复此过程 1 002 次, 对每个实验结果从 Top-1 到 Top- $n$  进行排序, 最后对

这些结果取平均值得到 Rank- $n$  识别率. 考虑到式 (15) 中调节两个损失函数平衡权重参数  $\lambda$  会影响模型性能, 本文在 FGNET 上设置了 4 组实验以探索  $\lambda$  合理的取值. 图 7 显示了  $\lambda$  取不同值时, 模型正确率的变化曲线, 当  $\lambda$  取值较小时, 年龄估算任务的重要性较低, 使得特征分离不充分, 随着  $\lambda$  逐渐增大, 年龄因素和身份因素的相关性随之减小, 识别准确率逐步上升;  $\lambda$  超过一定范围后, 模型对身份因素的表征能力会受到干扰, 从而使得识别精度降低. 实验表明, 将  $\lambda$  的值设为 0.5 能使模型得到较好的性能. 表 1 给出了本文方法与其他已发表方法的 Rank-1 识别率, 可以看到本文提出的基于深度学习的方法不仅与其他基于线性分析的方法相比有较大优势, 其识别结果也高于 CAN<sup>[22]</sup> 和 LF-CNN<sup>[21]</sup> 两种深度模型.

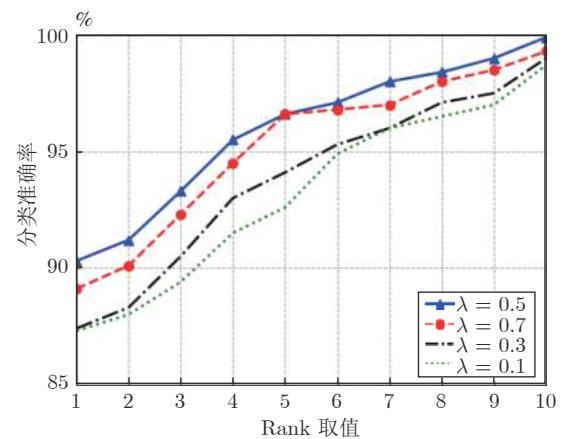


图 7 不同权重参数下的人脸分类准确率曲线图  
Fig.7 Face classification accuracy graph under different weight parameters

表 1 不同方法在 FGNET 数据库上的识别率  
Table 1 Recognition rate of different method on FGNET

方法	识别率 (%)
Li 等 <sup>[18]</sup> (2010)	47.5
HFA <sup>[19]</sup> (2013)	69.0
MEFA <sup>[16]</sup> (2015)	76.2
CAN <sup>[22]</sup> (2017)	86.5
LF-CNN <sup>[21]</sup> (2016)	88.1
本文方法	<b>90.0</b>

在 FGNET 中, 受试者的年龄分布不均匀, 并且与 MORPH 中的年龄分布也不同, 在 FGNET 中, 大约 61% 的样本不到 16 岁. 但对于 MORPH 数据集, 所有人都超过 16 岁. 因此, 有必要验证所提方法在不同年龄分布下的表现. 本文将 FGNET 中的图片按照年龄分成了 0~4, 5~10, 11~16, 17~24,

25 ~ 69 五个年龄段, 表 2 给出了各年龄段的 Rank-1 识别率.

表 2 本文方法在 FGNET 数据库上各个年龄段的识别正确率

Table 2 Performance of our method on different age groups on FGNET

年龄组	数量	原始特征提取网络 (%)	本文方法 (%)
0 ~ 4	193	60.40	67.30
5 ~ 10	218	86.86	89.12
11 ~ 16	201	92.43	95.81
17 ~ 24	182	94.63	98.01
25 ~ 69	208	99.09	99.54
0 ~ 16	612	80.30	84.43
17 ~ 69	390	97.01	98.87

可以看出, 本文方法在所有年龄组上均优于 CNN 基线模型, 表明年龄不变的特征能够成功地从原始特征中分离并良好地表征. 但我们注意到, 在较小的年龄段中识别结果较低, 这是因为在幼年时人的脸部轮廓和纹理变化较大, 与成年时期相比面部特征构成有所差异.

## 2.5 MORPH 数据库上的实验结果

MORPH Album 2 数据集包含了超过 78 000 幅人脸图像和 20 000 多个身份 id. 为了保证与其他方法进行对比的公平性, 本文依据文献 [13, 15] 对训练集和测试集的划分, 随机选取 10 000 个身份 id 作为训练集, 余下的 10 000 个身份 id 作为测试集. 测试集的每个身份 id 中选取年龄跨度最大的两幅照片, 年龄较小的 10 000 幅照片作为 gallery 集, 较大的作为 prob 集, 根据 FG-NET 的实验结果  $\lambda$  取 0.5, 实验结果如表 3 所示.

表 3 不同方法在 MORPH 数据库上的识别率

Table 3 Recognition rate of different method on MORPH

方法	识别率 (%)
HOG+HFA <sup>[15]</sup> (2013)	91.14
HLBP+CARC <sup>[18]</sup> (2014)	92.80
HOG+IFA <sup>[16]</sup> (2015)	92.26
MEFA <sup>[20]</sup> (2015)	93.80
LPS+HFA <sup>[17]</sup> (2016)	94.87
LF-CNNs基线模型 <sup>[21]</sup> (2016)	95.13
LF-CNN <sup>[21]</sup> (2016)	97.51
原始特征提取网络 (在CASIA上预训练)	74.40
原始特征提取网络 (用MORPH数据集微调)	96.77
MORPH微调后的原始特征提取网络+联合训练	97.10
本文方法	<b>98.93</b>

从表 3 中我们可以得出以下结论: 1) 没有微调的基线 CNN 的结果仅 74.73%, 不如表 3 中的其他结果, 这表明 CASIA-WebFace 数据集与 MORPH Album 2 数据集中的图片有明显不同, CASIA-WebFace 数据集不适合直接用来解决年龄不变问题. 2) 由 MORPH Album 2 微调的 CNN 基线模型可以达到 96.77% 的准确率, 这表明卷积神经网络内部具有非常强大的特征提取能力, 但原始特征提取网络所提取的特征本身包含了年龄因素的干扰, 所以与其他方法相比没有显著优势, 因此设计一个不受年龄因素干扰的卷积神经网络是有必要的. 3) 本文所提出的网络模型在由 MORPH Album 2 上的准确率达到 98.93%, 相比于 CNN 基线模型有明显的提升, 这表明使用估计任务来指导年龄不变的人脸识别任务是有用的, 通过从原始特征中减去年龄特征, 成功的得到了抗年龄干扰的身份特征. 4) 本文方法是建立在人脸分类和年龄回归联合训练基础之上的, 而多任务联合训练也会影响模型正确率, 因此为了进一步评估年龄因素对人脸识别带来的干扰, 验证正确率的提升是本文的特征分离方法带来的而不是联合训练所导致, 本文设计了一组对比实验, 在用 MORPH 数据集对原始特征提取网络微调后, 分别用本文所提的方法和仅基于联合训练的方法继续训练, 结果显示仅基于联合训练的方法正确率为 97.10%. 图 8 给出了训练过程中年龄估计和面部识别的曲线图, 年龄估计采用训练损失表示. 从图中可以看出仅采用联合训练的方式无法给识别正确率带来明显提升, 正确率曲线在短时间内达到饱和. 采用本文方法时, 正确率有较大幅度上涨, 并且人脸识别率曲线与年龄回归损失曲线发生显著变化的时间段基本吻合, 说明随着网络对年龄的估计越来越准确, 原始特征的分离也就越彻底, 在排除了年龄干扰后, 人脸识别率随之提升.

最后, 图 9 显示了失败检索的一些示例. 第 1 行为待测试的人脸图像, 第 2 行对应识别结果, 第 3 行对应真实的结果. 虽然在这些情况下 Rank-1 检索不正确, 但待测图像与识别结果的视觉相似性高于真实图像.

## 3 结束语

本文提出了一种基于深度学习的抗年龄干扰的人脸识别方法, 该方法将面部识别任务、年龄估计任务和网络注意力机制结合到同一个深度模型中, 使用 Resnet-34-CBAM 提取整张人脸的特征, CBAM 注意力模型使得网络所提取到的特征更具有针对性和完整性, 随后从整个特征中减去由年龄

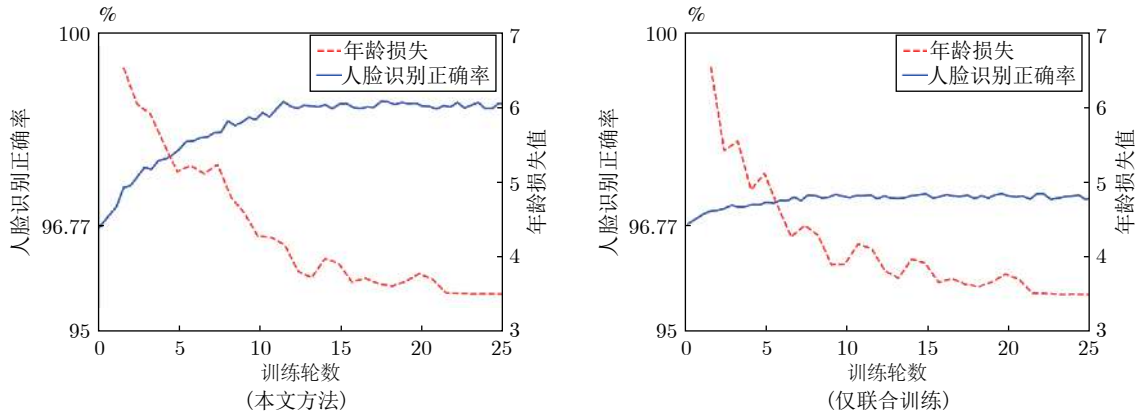


图 8 识别正确率与年龄估计值的变化曲线

Fig.8 The performance of age estimation and cross-age face recognition rate



图 9 MORPH Album 2 中部分识别错误的人脸图像

Fig.9 Some examples of failed retrievals in MORPH Album 2 dataset

估计任务所得到的年龄因子以消除掉年龄因素对人脸识别带来的干扰,最后使用基于角度损失函数的人脸识别方法对只包含身份信息特征进行识别.

本文方法在 MORPH 和 FGNET 两个数据库上得到了理想的实验结果,充分证明了该方法的有效性,此外,本文所引入的年龄因子也可替换为其他干扰



因子(如表情、光照等), 有望用来解决类似的类内干扰较大的人脸识别问题。

## References

- Schroff F, Kalenichenko D, Philbin J. FaceNet: A unified embedding for face recognition and clustering. In: Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston, USA: IEEE, 2015. 815–823
- Wang H, Wang Y T, Zhou Z, Ji X, Gong D H, Zhou J C, et al. CosFace: Large margin cosine loss for deep face recognition. In: Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE, 2018. 5265–5274
- Zhao J, Xiong L, Cheng Y, Cheng Y, Li J S, Zhou L, et al. 3D-aided deep pose-invariant face recognition. In: Proceedings of the 27th International Joint Conference on Artificial Intelligence. Stockholm, Sweden: AAAI Press, 2018. 1184–1190
- Hu Yang, Zhang Dong-Bo, Duan Qi. An improved rotation-invariant HDO local description for object recognition. *Acta Automatica Sinica*, 2017, **43**(4): 665–673  
(胡扬, 张东波, 段琪. 目标鲁棒识别的抗旋转HDO局部特征描述. *自动化学报*, 2017, **43**(4): 665–673)
- Wang Cun-Rui, Zhang Qing-Ling, Duan Xiao-Dong, Wang Yuan-Gang, Li Ze-Dong. Research of face ethnic features from manifold structure. *Acta Automatica Sinica*, 2018, **44**(1): 140–159  
(王存睿, 张庆灵, 段晓东, 王元刚, 李泽东. 基于流形结构的人脸民族特征研究. *自动化学报*, 2018, **44**(1): 140–159)
- Wang Yu, Shen Xuan-Jing, Chen Hai-Peng. Video face recognition based on modified Fisher criteria and multi-instance learning. *Acta Automatica Sinica*, 2018, **44**(12): 2179–2187  
(王玉, 申铨京, 陈海鹏. 基于改进的Fisher准则的多示例学习视频人脸识别算法. *自动化学报*, 2018, **44**(12): 2179–2187)
- Chen B C, Chen C S, Hsu W H. Face recognition and retrieval using cross-age reference coding with cross-age celebrity dataset. *IEEE Transactions on Multimedia*, 2015, **17**(6): 804–815
- Geng X, Zhou Z H, Smith-Miles K. Automatic age estimation based on facial aging patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2007, **29**(12): 2234–2240
- Lanitis A, Taylor C J, Cootes T F. Toward automatic simulation of aging effects on face images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2002, **24**(4): 442–455
- Park U, Tong Y Y, Jain A K. Age-invariant face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2010, **32**(5): 947–954
- Zhang Z F, Song Y, Qi H R. Age progression/regression by conditional adversarial autoencoder. In: Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, USA: IEEE, 2017. 4352–4360
- Antipov G, Baccouche M, Dugelay J L. Face aging with conditional generative adversarial networks. In: Proceedings of the 2017 IEEE International Conference on Image Processing (ICIP). Beijing, China: IEEE, 2017. 2089–2093
- Duong C N, Quach K G, Luu K, Le T H N, Savvides M. Temporal non-volume preserving approach to facial age-progression and age-invariant face recognition. In: Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV). Venice, Italy: IEEE, 2017. 3755–3763
- Ling H B, Soatto S, Ramanathan N, Jacobs D W. Face verification across age progression using discriminative methods. *IEEE Transactions on Information Forensics and Security*, 2010, **5**(1): 82–91
- Gong D H, Li Z F, Lin D H, Liu J Z, Tang X O. Hidden factor analysis for age invariant face recognition. In: Proceedings of the 2013 IEEE International Conference on Computer Vision. Sydney, Australia: IEEE, 2013. 2872–2879
- Gong D H, Li Z F, Tao D C, Liu J Z, Li X L. A maximum entropy feature descriptor for age invariant face recognition. In: Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston, USA: IEEE, 2015. 5289–5297
- Li Z F, Gong D H, Li X L, Tao D C. Aging face recognition: A hierarchical learning model based on local patterns selection. *IEEE Transactions on Image Processing*, 2016, **25**(5): 2146–2154
- Li Z F, Park U, Jain A K. A discriminative model for age invariant face recognition. *IEEE Transactions on Information Forensics and Security*, 2011, **6**(3): 1028–1037
- Lin L, Wang G R, Zuo W M, Feng X C, Zhang L. Cross-domain visual matching via generalized similarity measure and feature learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, **39**(6): 1089–1102
- Wang Y T, Gong D H, Zhou Z, Ji X, Wang H, Li Z F, et al. Orthogonal deep features decomposition for age-invariant face recognition. In: Proceedings of the 15th European Conference on Computer Vision. Munich, Germany: Springer, 2018. 764–779
- Wen Y D, Li Z F, Qiao Y. Latent factor guided convolutional neural networks for age-invariant face recognition. In: Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, USA: IEEE, 2016. 4893–4901
- Xu C F, Liu Q H, Ye M. Age invariant face recognition and retrieval by coupled auto-encoder networks. *Neurocomputing*, 2017, **222**: 62–71
- Yi D, Lei Z, Liao S C, Li S Z. Learning face representation from scratch [Online], available: <https://arxiv.org/abs/1411.7923>, November 28, 2014
- He K M, Zhang X Y, Ren S Q, Sun J. Deep residual learning for image recognition. In: Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, USA: IEEE, 2016. 770–778
- Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In: Proceedings of the 32nd International Conference on International Conference on Machine Learning. Lille, France: JMLR.org, 2015. 448–456
- Woo S, Park J, Lee J Y, Kweon I S. CBAM: Convolutional block attention module. In: Proceedings of the 15th European Conference on Computer Vision. Munich, Germany: Springer, 2018. 3–19
- Liu W Y, Wen Y D, Yu Z D, Li M, Raj B, Song L. SphereFace: Deep hypersphere embedding for face recognition. In: Proceedings of the 2017 IEEE Conference on Computer Vision and Pat-

tern Recognition (CVPR). Honolulu, USA: IEEE, 2017. 6738–6746

- 28 Ricanek K, Tesafaye T. MORPH: A longitudinal image database of normal adult age-progression. In: Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition (FGR06). Southampton, UK: IEEE, 2006. 341–345
- 29 Liang Y X, Liu L B, Xu Y, Xiang Y, Zou B J. Multi-task GLOH feature selection for human age estimation. In: Proceedings of the 18th IEEE International Conference on Image Processing. Brussels, Belgium: IEEE, 2011. 565–568
- 30 Zhang K P, Zhang Z P, Li Z F, Qiao Y. Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE Signal Processing Letters*, 2016, **23**(10): 1499–1503
- 31 Wang F, Xiang X, Cheng J, Yuille A L. NormFace: L2 Hypersphere embedding for face verification. In: Proceedings of the 2017 ACM on Multimedia Conference. Mountain View, USA: ACM, 2017. 1041–1049



**何星辰** 成都理工大学信息科学与技术学院硕士研究生. 主要研究方向为图像处理, 计算机视觉与模式识别.

E-mail: hxc\_cdut@163.com

**(HE Xing-Chen** Master student at the College of Information Science and Technology, Chengdu University of Technology. His research interest covers image processing, computer vision, and pattern recognition.)

His research interest covers image processing, computer vision, and pattern recognition.)



**郭勇** 成都理工大学信息科学与技术学院教授. 主要研究方向为图像处理, 模式识别, 灾害预警与救援技术. 本文通信作者.

E-mail: guoy@cdut.edu.cn

**(GUO Yong** Professor at the College of Information Science and Technology, Chengdu University of Technology. His research interest covers image processing, pattern recognition, information and disaster early warning, and rescue technology. Corresponding author of this paper.)



**李奇龙** 成都理工大学信息科学与技术学院硕士研究生. 主要研究方向为计算机视觉, 自然语言处理.

E-mail: 15008294254@163.com

**(LI Qi-Long** Master student at the College of Information Science and Technology, Chengdu University of Technology. His research interest covers computer vision and natural language processing.)



**高唱** 成都理工大学地球物理学院硕士研究生. 主要研究方向为机器学习. E-mail: gaochang0708@163.com

**(GAO Chang** Master student at the College of Geophysics, Chengdu University of Technology. Her main research interest is machine learning.)

learning.)