



(12) 发明专利申请

(10) 申请公布号 CN 113449661 A

(43) 申请公布日 2021. 09. 28

(21) 申请号 202110758045.5

(22) 申请日 2021.07.05

(71) 申请人 东南大学

地址 211189 江苏省南京市江宁区东南大学路2号

(72) 发明人 郑文明 魏梦婷 宗源 赵力

(74) 专利代理机构 南京苏高专利商标事务所
(普通合伙) 32204

代理人 王安琪

(51) Int. Cl.

G06K 9/00 (2006.01)

G06N 3/08 (2006.01)

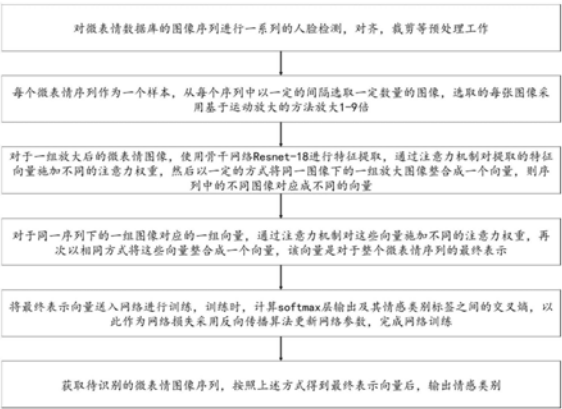
权利要求书2页 说明书5页 附图3页

(54) 发明名称

一种基于注意力机制的自适应微表情识别方法

(57) 摘要

本发明公开了一种基于注意力机制的自适应微表情识别方法,包括如下步骤:对微表情数据库的图像序列进行一系列的预处理工作;每个微表情序列作为一个样本,从每个序列中以一定的间隔选取一定数量的图像,选取的每张图像采用基于运动放大的方法放大1-9倍;在一组不同的放大图像间添加注意力权重,并将其最终整合成一个特征向量;对于同一序列下的一组图像对应的一组向量,通过注意力机制对这些向量施加不同的注意力权重,再次以相同方式将这些向量整合成一个向量;将最终表示向量送入网络进行训练;获取待识别的微表情图像序列,按照上述方式得到最终表示向量后,输出情感类别。本发明能够获得更高的识别准确率。



1. 一种基于注意力机制的自适应微表情识别方法,其特征在于,包括如下步骤:

(1) 获取一个微表情数据库,对数据库中的所有图像序列进行人脸检测,人脸校准,面部区域裁剪的工作,最终获得只包含面部区域的图像,数据库中的一个受试者作为测试集,剩下的最为训练集;

(2) 从预处理后的每个微表情图像序列中以一定的间隔选取一定数量的图像,每张图像采用基于运动放大的方法依次放大1-9倍;

(3) 对单张图像的一组放大图像使用Resnet-18骨干网络进行特征提取,得到每个放大图像对应的特征向量,对这一组特征向量使用注意力机制,注意力机制由一个全连接层和一个sigmoid函数组成,添加了注意力权重的特征向量之后被整合为一个向量;

(4) 经过上述步骤,序列中的每张图像都对应有一个特征向量,对同一序列中的这些特征向量再次使用注意力机制,添加了注意力权重的特征向量再次被整合成为一个向量;

(5) 将经过上述步骤得到的特征向量经过一个全连接层,全连接层的输出维度与数据库的表情类别数一致,在全连接层后添加softmax层,计算训练数据对应的softmax层输出以及其情感类别标签之间的交叉熵,将该交叉熵作为网络损失,采用反向传播算法更新网络参数,完成网络训练;

(6) 将测试数据集以同样的方式进行预处理,得到其对应的特征向量,并将特征向量输入训练好的网络,输出中维数最大的类别即为识别的情感类别。

2. 如权利要求1所述的基于注意力机制的自适应微表情识别方法,其特征在于,步骤(2)中所述放大图像的表示为:

$$I_{\text{magnified}} = I_{\text{onset}} + (I_{\text{seleted}} - I_{\text{onset}}) \times \text{amp}_i$$

式中, $I_{\text{magnified}}$ 表示放大后图像, I_{onset} 表示微表情序列的起始帧, I_{seleted} 表示选取的图像帧, amp_i 表示放大因子,值设置为1-9。

3. 如权利要求1所述的基于注意力机制的自适应微表情识别方法,其特征在于,步骤(3)中所述使用Resnet-18骨干网络进行特征提取得到方法图像对应的特征向量的计算方法为:

$$F_i = f(\text{conv}(I_{\text{magnified}}, W_c) + b_c)$$

式中, F_i 表示放大图像对应的特征向量, $I_{\text{magnified}}$ 表示骨干网络的输入, W_c 、 b_c 分别为权重参数、偏置参数,有一部分通过网络训练得到,另外一部分为骨干网络原始参数不参与更新,conv()表示卷积函数, ϕ 为激活函数,且 $\phi(\cdot) = \max(0, \cdot)$ 。

4. 如权利要求1所述的基于注意力机制的自适应微表情识别方法,其特征在于,步骤(3)中所述对每张放大图像施加注意力权重的计算方法为:

$$\alpha_i = \sigma(F_i^T q^0), F_i \in [F_1, F_2, \dots, F_9]$$

式中, α_i 为单张放大图像的注意力权重值, F_i^T 表示放大图像经过骨干网络特征提取后得到的特征向量, q^0 表示全连接层的参数,通过网络训练得到, σ 表示sigmoid函数。

5. 如权利要求1所述的基于注意力机制的自适应微表情识别方法,其特征在于,步骤(3)中所述将添加注意力权重的特征向量整合为一个特征向量的计算方法为:

$$F_m = \frac{\sum_{i=1}^9 \alpha_i F_i^T}{\sum_{i=1}^9 \alpha_i},$$

式中, F_m 表示单张图像整合后得到的特征向量, α_i 为单张放大图像的注意力权重值, F_i^T 表示放大图像经过骨干网络特征提取后得到的特征向量。

6. 如权利要求1所述的基于注意力机制的自适应微表情识别方法, 其特征在于, 步骤(4)中所述对序列中每张图像对应的整合后的特征向量施加注意力权重值的计算方法为:

$$\beta_j = \sigma(F_{m_j}^T p^0), F_{m_j} \in [F_{m_1}, F_{m_2}, \dots, F_{m_n}]$$

式中, β_j 表示每张图像对应的整合后的特征向量添加的注意力权重值, F_{m_j} 表示单张图像整合后得到的特征向量, p^0 为另一全连接层的参数, 通过网络训练得到, σ 表示sigmoid函数。

7. 如权利要求1所述的基于注意力机制的自适应微表情识别方法, 其特征在于, 步骤(4)中所述添加了注意力权重的特征向量再次被整合成为一个向量的计算方法为:

$$F_v = \frac{\sum_{j=1}^n \beta_j F_{m_j}^T}{\sum_{j=1}^n \beta_j}$$

式中, F_v 表示对整个微表情序列的最终表示, β_j 表示每张图像对应的整合后的特征向量添加的注意力权重值, F_{m_j} 表示单张图像整合后得到的特征向量。

8. 如权利要求1所述的基于注意力机制的自适应微表情识别方法, 其特征在于, 步骤(5)中所述网络损失的计算方法为:

$$\text{Loss} = - \sum_{i=1}^c p_i \log(q_i)$$

式中, c 表示情感类别数, p_i 为训练数据库中存储的微表情序列对应的情感类别, q_i 表示 F_v 输入到最后的全连接层和softmax层之后, softmax层输出的情感类别, 即预测的情感类别。

一种基于注意力机制的自适应微表情识别方法

技术领域

[0001] 本发明涉及微表情数据情感识别技术领域,尤其是一种基于注意力机制的自适应微表情识别方法。

背景技术

[0002] 微表情是指当一个人试图掩藏自己的真实情绪时,不经意泄露出的表情状态。它的持续时间一般在0.065-0.5秒之间,且强度微弱,肉眼难以捕捉。相比于有意识的表情,微表情更能反映人们的真实感受和动机,应用前景广阔。针对微表情微弱的特性,许多研究方法首先采用某种方法对表情强度进行放大,之后基于放大的表情进行识别。而在实际强度放大过程中,不同受试者之间,同一受试者的不同表情之间存在很大差异,对于有的微表情,较小的放大倍数就可以产生比较明显的效果,而对于另外一些微表情,放大倍数则要设置的比较大才能看到效果,传统的方法没有考虑到这一问题,直接对所有的微表情设置同一个放大强度,这极大可能会影响后续的识别性能。

[0003] 注意力机制是一种将同等看待变为差别对待的方式,结合深度学习的学习过程,可以自动地为有利于提高分类性能的特征赋予更大的注意力权重,在图像识别领域应用广泛。

发明内容

[0004] 本发明所要解决的技术问题在于,提供一种基于注意力机制的自适应微表情识别方法,能够获得更高的识别准确率。

[0005] 为解决上述技术问题,本发明提供一种基于注意力机制的自适应微表情识别方法,包括如下步骤:

[0006] (1) 获取一个微表情数据库,对数据库中的所有图像序列进行人脸检测,人脸校准,面部区域裁剪的工作,最终获得只包含面部区域的图像,数据库中的一个受试者作为测试集,剩下的最为训练集;

[0007] (2) 从预处理后的每个微表情图像序列中以一定的间隔选取一定数量的图像,每张图像采用基于运动放大的方法依次放大1-9倍;

[0008] (3) 对单张图像的一组放大图像使用Resnet-18骨干网络进行特征提取,得到每个放大图像对应的特征向量,对这一组特征向量使用注意力机制,注意力机制由一个全连接层和一个sigmoid函数组成,添加了注意力权重的特征向量之后被整合为一个向量;

[0009] (4) 经过上述步骤,序列中的每张图像都对应有一个特征向量,对同一序列中的这些特征向量再次使用注意力机制,添加了注意力权重的特征向量再次被整合成为一个向量;

[0010] (5) 将经过上述步骤得到的特征向量经过一个全连接层,全连接层的输出维度与数据库的表情类别数一致,在全连接层后添加softmax层,计算训练数据对应的softmax层输出以及其情感类别标签之间的交叉熵,将该交叉熵作为网络损失,采用反向传播算法更

新网络参数,完成网络训练;

[0011] (6)将测试数据集以同样的方式进行预处理,得到其对应的特征向量,并将特征向量输入训练好的网络,输出中维数最大的类别即为识别的情感类别。

[0012] 优选的,步骤(2)中所述放大图像的表示为:

$$[0013] \quad I_{\text{magnified}} = I_{\text{onset}} + (I_{\text{selected}} - I_{\text{onset}}) \times \text{amp}_i$$

[0014] 式中,对于一个微表情序列样本, $I_{\text{magnified}}$ 表示放大后图像, I_{onset} 表示微表情序列的起始帧, I_{selected} 表示选取的图像帧, amp_i 表示放大因子,值设置为1-9。

[0015] 优选的,步骤(3)中所述使用Resnet-18骨干网络进行特征提取得到方法图像对应的特征向量的计算方法为:

$$[0016] \quad F_i = f(\text{conv}(I_{\text{magnified}}, W_c) + b_c)$$

[0017] 式中, F_i 表示放大图像对应的特征向量, $I_{\text{magnified}}$ 表示骨干网络的输入, W_c 、 b_c 分别为权重参数、偏置参数,有一部分通过网络训练得到,另外一部分为骨干网络原始参数不参与更新, $\text{conv}()$ 表示卷积函数, ϕ 为激活函数,且 $\phi(\cdot) = \max(0, \cdot)$ 。

[0018] 优选的,步骤(3)中所述对每张放大图像施加注意力权重的计算方法为:

$$[0019] \quad \alpha_i = \sigma(F_i^T q^0), \quad F_i \in [F_1, F_2, \dots, F_9]$$

[0020] 式中, α_i 为单张放大图像的注意力权重值, F_i^T 表示放大图像经过骨干网络特征提取后得到的特征向量, q^0 表示全连接层的参数,通过网络训练得到, σ 表示sigmoid函数。

[0021] 优选的,步骤(3)中所述将添加注意力权重的特征向量整合为一个特征向量的计算方法为:

$$[0022] \quad F_m = \frac{\sum_{i=1}^9 \alpha_i F_i^T}{\sum_{i=1}^9 \alpha_i},$$

[0023] 式中, F_m 表示单张图像整合后得到的特征向量, α_i 为单张放大图像的注意力权重值, F_i^T 表示放大图像经过骨干网络特征提取后得到的特征向量。

[0024] 优选的,步骤(4)中所述对序列中每张图像对应的整合后的特征向量施加注意力权重值的计算方法为:

$$[0025] \quad \beta_j = \sigma(F_{m_j}^T p^0), \quad F_{m_j} \in [F_{m_1}, F_{m_2}, \dots, F_{m_n}]$$

[0026] 式中, β_j 表示每张图像对应的整合后的特征向量添加的注意力权重值, F_{m_j} 表示单张图像整合后得到的特征向量, p^0 为另一全连接层的参数,通过网络训练得到, σ 表示sigmoid函数。

[0027] 优选的,步骤(4)中所述添加了注意力权重的特征向量再次被整合成为一个向量的计算方法为:

$$[0028] \quad F_v = \frac{\sum_{j=1}^n \beta_j F_{m_j}^T}{\sum_{j=1}^n \beta_j}$$

[0029] 式中, F_v 表示对整个微表情序列的最终表示, β_j 表示每张图像对应的整合后的特征向量添加的注意力权重值, F_{m_j} 表示单张图像整合后得到的特征向量。

[0030] 优选的,步骤(5)中所述网络损失的计算方法为:

$$[0031] \quad \text{Loss} = - \sum_{i=1}^c p_i \log(q_i)$$

[0032] 式中, c 表示情感类别数, p_i 为训练数据库中存储的微表情序列对应的情感类别, q_i 表示 F_v 输入到最后的全连接层和 softmax 层之后, softmax 层输出的情感类别, 即预测的情感类别。

[0033] 本发明的有益效果为: 本发明提供了一种基于注意力机制的自适应微表情识别方法, 该方法针对传统微表情放大过程中由于个体差异导致的效果不理想的问题, 采用深度神经网络结合注意力机制的方法, 使网络在训练过程中能够自动为合适的放大强度分配高注意力权重, 从而提高模型的识别准确率。

附图说明

[0034] 图1为本发明的方法流程示意图。

[0035] 图2为本发明一个微表情序列样本的选取及放大过程示意图。

[0036] 图3为本发明建立的在放大图像间添加注意力机制的结构示意图。

[0037] 图4为本发明建立的在特征向量间添加注意力机制的结构示意图。

[0038] 图5为本发明建立的两种注意力机制结合卷积神经网络进行训练的流程示意图。

具体实施方式

[0039] 如图1所示, 一种基于注意力机制的自适应微表情识别方法, 包括如下步骤:

[0040] (1) 获取一个微表情数据库, 对数据库中的所有图像序列进行人脸检测, 人脸校准, 面部区域裁剪的工作, 最终获得只包含面部区域的图像, 数据库中的一个受试者作为测试集, 剩下的最为训练集;

[0041] (2) 从预处理后的每个微表情图像序列中以一定的间隔选取一定数量的图像, 每张图像采用基于运动放大的方法依次放大1-9倍。整个选取及放大过程如图2所示。

[0042] 其中, 所述放大图像的表示为

$$[0043] \quad I_{\text{magnified}} = I_{\text{onset}} + (I_{\text{selected}} - I_{\text{onset}}) \times \text{amp}_i$$

[0044] 式中, $I_{\text{magnified}}$ 表示放大后图像, I_{onset} 表示微表情序列的起始帧, I_{selected} 表示选取的图像帧, amp_i 表示放大因子, 值设置为1-9。

[0045] (3) 对单张图像的一组放大图像使用 Resnet-18 骨干网络进行特征提取, 得到每个放大图像对应的特征向量, 对这一组特征向量使用注意力机制, 注意力机制由一个全连接层和一个 sigmoid 函数组成, 添加了注意力权重的特征向量之后被整合为一个向量, 具体如图3所示。

[0046] 其中, 所述使用 Resnet-18 骨干网络进行特征提取得到方法图像对应的特征向量的计算方法为:

$$[0047] \quad F_i = f(\text{conv}(I_{\text{magnified}}, W_c) + b_c)$$

[0048] 式中, F_i 表示放大图像对应的特征向量, $I_{\text{magnified}}$ 表示骨干网络的输入, W_c 、 b_c 分别为权重参数、偏置参数, 有一部分通过网络训练得到, 另外一部分为骨干网络原始参数不参与更新, $\text{conv}()$ 表示卷积函数, ϕ 为激活函数, 且 $\phi(\cdot) = \max(0, \cdot)$ 。

[0049] 所述对每张放大图像的特征向量施加注意力权重的计算方法为:

$$[0050] \quad \alpha_i = \sigma(F_i^T q^0), \quad F_i \in [F_1, F_2, \dots, F_9]$$

[0051] 式中, α_i 为单张放大图像的注意力权重值, F_i^T 表示放大图像经过骨干网络特征提取后得到的特征向量, q^0 表示全连接层的参数, 通过网络训练得到, σ 表示sigmoid函数。

[0052] 所述将添加注意力权重的特征向量整合为一个特征向量的计算方法为:

$$[0053] \quad F_m = \frac{\sum_{i=1}^9 \alpha_i F_i^T}{\sum_{i=1}^9 \alpha_i},$$

[0054] 式中, F_m 表示单张图像整合后得到的特征向量, α_i 为单张放大图像的注意力权重值, F_i^T 表示放大图像经过骨干网络特征提取后得到的特征向量。

[0055] (4) 经过上述步骤, 序列中的每张图像都对应有一个特征向量, 对同一序列中的这些特征向量再次使用注意力机制, 添加了注意力权重的特征向量再次被整合成为一个向量, 具体如图4所示。

[0056] 其中, 所述对序列中每张图像对应的整合后的特征向量施加注意力权重值的计算方法为:

$$[0057] \quad \beta_j = \sigma(F_{m_j}^T p^0), \quad F_{m_j} \in [F_{m_1}, F_{m_2}, \dots, F_{m_n}]$$

[0058] 式中, β_j 表示每张图像对应的整合后的特征向量添加的注意力权重值, F_{m_j} 表示单张图像整合后得到的特征向量, p^0 为另一全连接层的参数, 通过网络训练得到, σ 表示sigmoid函数。

[0059] 所述添加了注意力权重的特征向量再次被整合成为一个向量的计算方法为:

$$[0060] \quad F_v = \frac{\sum_{j=1}^n \beta_j F_{m_j}^T}{\sum_{j=1}^n \beta_j}$$

[0061] 式中, F_v 表示对整个微表情序列的最终表示, β_j 表示每张图像对应的整合后的特征向量添加的注意力权重值, F_{m_j} 表示单张图像整合后得到的特征向量。

[0062] (5) 将经过上述步骤得到的特征向量经过一个全连接层, 全连接层的输出维度与数据库的表情类别数一致, 在全连接层后添加softmax层, 计算训练数据对应的softmax层输出以及其情感类别标签之间的交叉熵, 将该交叉熵作为网络损失, 采用反向传播算法更新网络参数, 完成网络训练, 具体如图5所示。

[0063] 其中, 所述网络损失的计算方法为:

$$[0064] \quad \text{Loss} = - \sum_{i=1}^c p_i \log(q_i)$$

[0065] 式中, c 表示情感类别数, p_i 为训练数据库中存储的微表情序列对应的情感类别, q_i 表示 F_v 输入到最后的全连接层和softmax层之后, softmax层输出的情感类别, 即预测的情感类别。

[0066] (6) 将测试数据集以同样的方式进行预处理, 得到其对应的特征向量, 并将特征向量输入训练好的网络, 输出中维数最大的类别即为识别的情感类别。

[0067] 下面以CASME II微表情数据库的情感识别人物为例进行验证,验证结果如表1所示:

[0068] 表1

[0069]	方法	类别	准确率(%)
	LBP-TOP+AdaBoost	5	43.78
	SIP+MOP	5	45.75
	CNN+LSTM	5	60.98
	TSCNN-I	5	74.05
	Graph-TCN	5	73.98
	AIAN	5	75.40

[0070] 其中LBP-TOP+AdaBoost、SIP+MOP为基于传统的手工提取表情特征与机器学习方法结合的微表情情感识别方法,CNN+LSTM为未进行表情放大,直接使用卷积神经网络进行特征提取和识别的方法,TSCNN-I,Graph-TCN分别为基于欧拉视频放大和运动放大的方法对微表情进行情感识别的模型,AIAN为基于注意力机制的自适应微表情识别模型。

[0071] 实验结果表明,基于所使用的注意力机制结合卷积神经网络的模型取得了最高的微表情识别准确率。在CASME II微表情数据的情感识别任务中,本发明提出的模型相对于其它传统的方法起到了更好的识别结果。相较于其他方法,本发明成功地增加了神经网络模型的鲁棒性,提高了微表情情感识别的准确率。

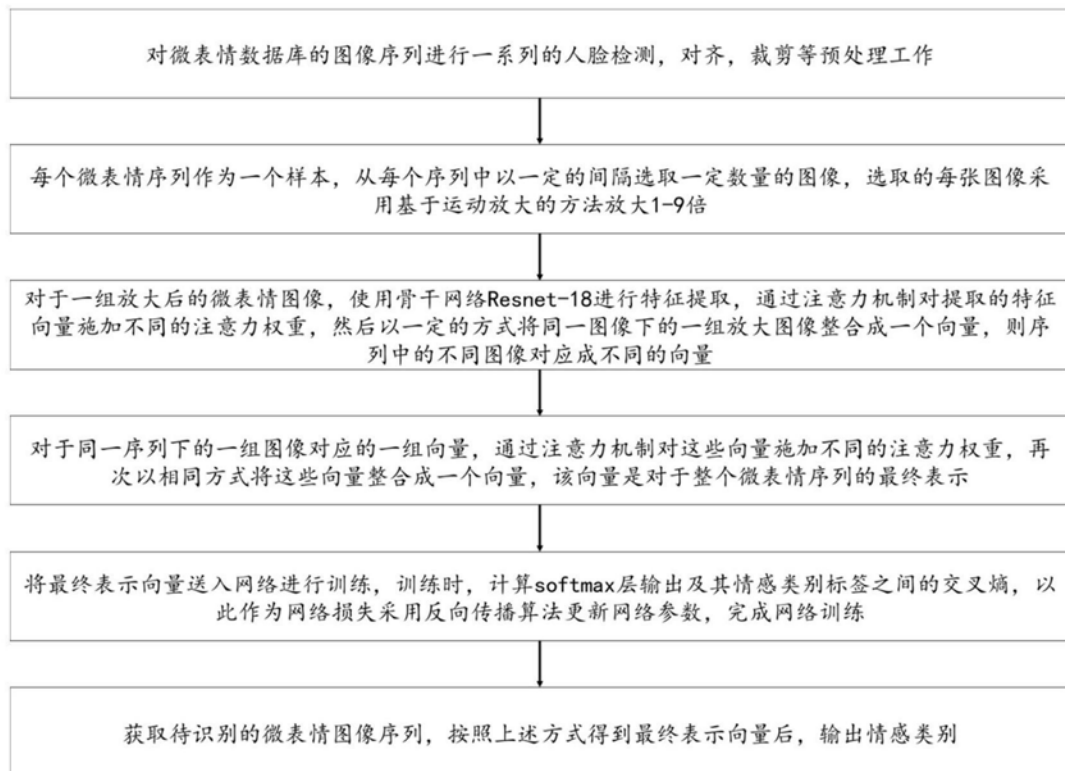


图1

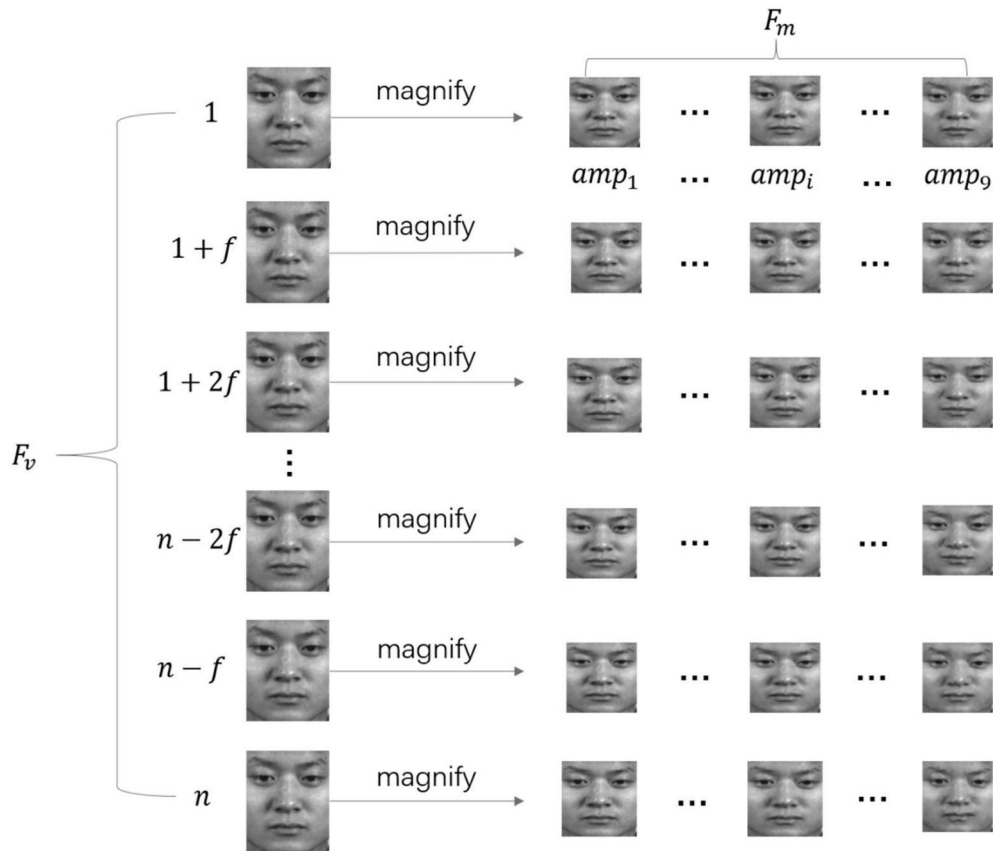


图2

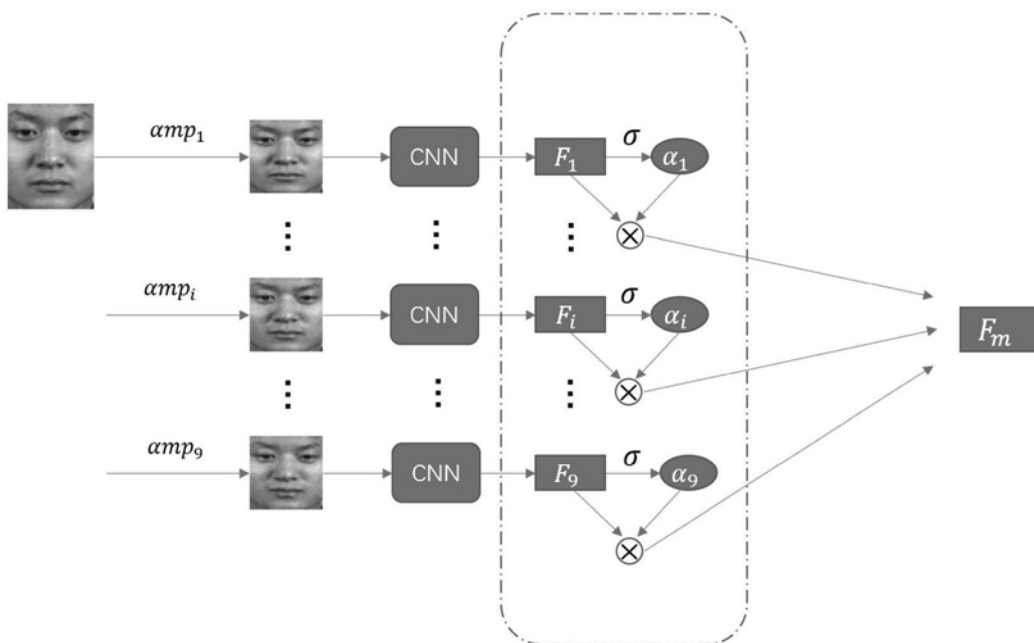


图3

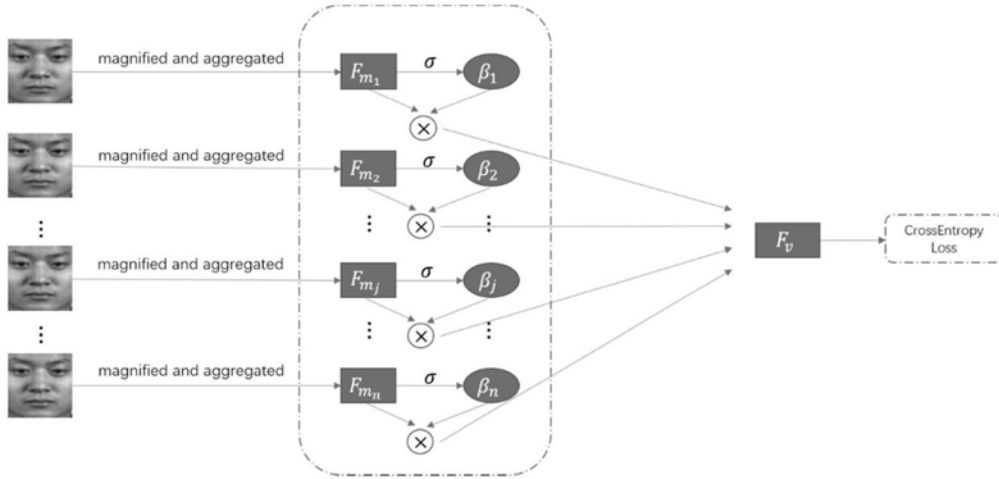


图4

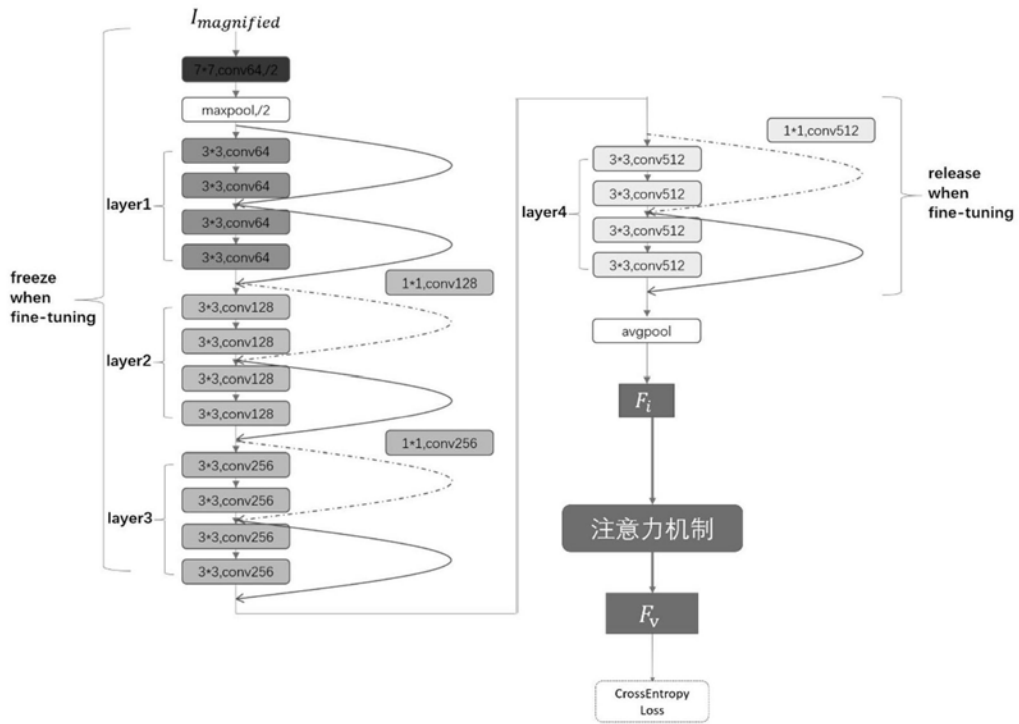


图5