



(12) 发明专利申请

(10) 申请公布号 CN 115049957 A

(43) 申请公布日 2022.09.13

(21) 申请号 202210605395.2

G06V 10/764 (2022.01)

(22) 申请日 2022.05.31

G06T 7/30 (2017.01)

(71) 申请人 东南大学

G06T 7/11 (2017.01)

地址 211102 江苏省南京市江宁区东南大学路2号

G06N 3/04 (2006.01)

G06F 16/75 (2019.01)

G06F 16/71 (2019.01)

(72) 发明人 郑文明 魏梦婷 宗源 江星洵
刘佳腾 薛云龙

(74) 专利代理机构 南京苏高专利商标事务所
(普通合伙) 32204

专利代理师 冯艳芬

(51) Int.Cl.

G06V 20/40 (2022.01)

G06V 40/16 (2022.01)

G06V 10/82 (2022.01)

G06V 10/774 (2022.01)

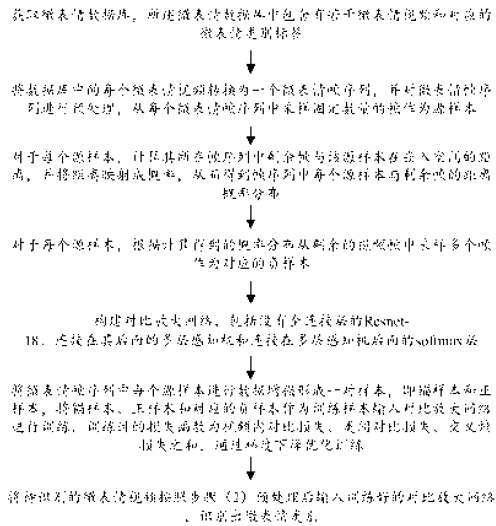
权利要求书3页 说明书7页 附图2页

(54) 发明名称

基于对比放大网络的微表情识别方法及装置

(57) 摘要

本发明公开了一种基于对比放大网络的微表情识别方法及装置,方法包括:(1)获取微表情数据库(2)将微表情视频转换为微表情帧序列,预处理后采样作为源样本;(3)对于每个源样本,计算剩余帧与该源样本在嵌入空间的距离,并映射成概率,得到距离概率分布;(4)根据概率分布从剩余视频帧中采样多个帧作为负样本;(5)构建对比放大网络;(6)每个源样本进行数据增强形成锚样本和正样本,将锚样本、正样本和对应的负样本作为训练样本输入对比放大网络进行训练,损失函数为视频内对比损失、类间对比损失、交叉熵损失之和;(7)将待识别的微表情视频预处理后输入训练好的对比放大网络,识别出微表情类别。本发明准确率更高,更方便。



1. 一种基于对比放大网络的微表情识别方法,其特征在於该方法包括:

(1) 获取微表情数据库,所述微表情数据库中包含有若干微表情视频和对应的微表情类别标签;

(2) 将数据库中的每个微表情视频转换为一个微表情帧序列,并对微表情帧序列进行预处理,从每个微表情帧序列中采样固定数量的帧作为源样本;

(3) 对于每个源样本,计算其所在帧序列中剩余帧与该源样本在嵌入空间的距离,并将距离映射成概率,从而得到帧序列中每个源样本与剩余帧的距离概率分布;

(4) 对于每个源样本,根据计算得到的概率分布从剩余的视频帧中采样多个帧作为对应的负样本;

(5) 构建对比放大网络,包括没有全连接层的Resnet-18、连接在其后面的多层感知机和连接在多层感知机后面的softmax层;

(6) 将微表情帧序列中每个源样本进行数据增强形成一对样本,即锚样本和正样本,将锚样本、正样本和对应的负样本作为训练样本输入对比放大网络进行训练,训练时的损失函数为视频内对比损失、类间对比损失、交叉熵损失之和,通过梯度下降优化训练;

(7) 将待识别的微表情视频按照步骤(2)预处理后输入训练好的对比放大网络,识别出微表情类别。

2. 根据权利要求1所述的基于对比放大网络的微表情识别方法,其特征在於:步骤(2)中所述预处理包括人脸配准和面部区域剪切。

3. 根据权利要求1所述的基于对比放大的微表情识别方法,其特征在於:步骤(2)中所述从每个微表情帧序列中采样固定数量的帧作为源样本具体包括:

对于一个预处理后的微表情帧序列,根据均匀分布采样N帧作为源样本,均匀分布的形式为:

$$f(n) = \begin{cases} \frac{1}{N}, & 1 \leq n \leq N \\ 0, & \text{其他} \end{cases}$$

式中, $f(n)$ 表示对源样本采样的概率密度函数, n 表示采样的源样本的索引号, N 表示采样的源样本的数量。

4. 根据权利要求1所述的基于对比放大网络的微表情识别方法,其特征在於:步骤(3)具体包括:

(3-1) 对于每个微表情帧序列,使用预训练好的神经网络提取源样本和剩余帧的特征向量 u :

$$u_i = g(x_i), i = 1, \dots, N$$

$$u_{i_j} = g(x_{i_j}), j = 1, \dots, n_i$$

式中, $g(\cdot)$ 为神经网络, N 表示源样本的数量, x_i 、 x_{i_j} 表示第*i*个源样本和 x_i 的第*j*个剩余帧, u_i 、 u_{n_i} 表示 x_i 、 x_{i_j} 的特征向量, n_i 表示 x_i 的剩余帧数量;

(3-2) 对于每个微表情帧序列,按照下式计算该帧序列中剩余帧与源样本的特征向量在嵌入空间的距离:

$$d_{ij} = u_i \cdot u_j, i = 1, \dots, N, j = 1, \dots, ni$$

式中, d_{ij} 表示第*i*个源样本与其第*j*个剩余帧在嵌入空间的距离;

(3-3) 对于每个微表情帧序列,按照下式使用softmax函数将不同的距离映射成概率,从而得到每个源样本与剩余帧的距离概率分布:

$$p_{ij} = \frac{e^{d_{ij}}}{\sum_l e^{d_{il}}}, i = 1, \dots, N, j = 1, \dots, ni$$

$$P(X = d_{ij}) = \{p_{ij} | i = 1, \dots, N, j = 1, \dots, ni\}$$

式中, p_{ij} 表示 d_{ij} 映射的概率, P 表示源样本与剩余帧的距离概率分布。

5. 根据权利要求1所述的基于对比放大网络的微表情识别方法,其特征在于:步骤(4)具体包括:

对于每个源样本,从其距离概率分布中进行采样,将采样得到的概率对应的视频帧作为负样本。

6. 根据权利要求1所述的基于对比放大网络的微表情识别方法,其特征在于:步骤(6)中所述损失函数具体为:

$$\mathcal{L} = \lambda_1 \mathcal{L}_{intra-video} + \lambda_2 \mathcal{L}_{inter-class} + \lambda_3 \mathcal{L}_{classifier}$$

其中, \mathcal{L} 为总损失, $\lambda_1, \lambda_2, \lambda_3$ 为三个损失函数的权重, $\mathcal{L}_{intra-video}$ 为视频内对比损失, $\mathcal{L}_{inter-class}$ 为类间对比损失, $\mathcal{L}_{classifier}$ 为交叉熵损失。

7. 根据权利要求6所述的基于对比放大网络的微表情识别方法,其特征在于:所述视频内对比损失具体函数如下:

$$\begin{aligned} \mathcal{L}_{intra-video} = & \sum_{k \in I} \frac{-1}{N} \sum_{i=1}^N \left[\log \frac{\text{sim}((\tilde{z}_1)_i^k, (\tilde{z}_2)_i^k)}{\text{sim}((\tilde{z}_1)_i^k, (\tilde{z}_2)_i^k) + \sum_{j=1}^S \exp\left((\tilde{z}_1)_i^k \cdot \frac{(z_j^-)_i^k}{\tau}\right)} \right. \\ & \left. + \log \frac{\text{sim}((\tilde{z}_1)_i^k, (\tilde{z}_2)_i^k)}{\text{sim}((\tilde{z}_1)_i^k, (\tilde{z}_2)_i^k) + \sum_{j=1}^S \exp\left((\tilde{z}_2)_i^k \cdot \frac{(z_j^-)_i^k}{\tau}\right)} \right] \end{aligned}$$

式中, I 为微表情数据库, $\text{sim}((\tilde{z}_1)_i^k, (\tilde{z}_2)_i^k) = \exp((\tilde{z}_1)_i^k \cdot (\tilde{z}_2)_i^k / \tau)$, $(\tilde{z}_1)_i^k, (\tilde{z}_2)_i^k, (z_j^-)_i^k$ 分别为 $(\tilde{x}_1)_i^k, (\tilde{x}_2)_i^k, (x_j^-)_i^k$ 输入对比放大网络后多层感知机输出的特征向量, $(\tilde{x}_1)_i^k, (\tilde{x}_2)_i^k, (x_j^-)_i^k$ 分别为 I 中第*k*个微表情视频的第*i*个源样本经过数据增强得到的正样本、锚样本、第*j*个负样本, τ 为温度系数, S 为负样本的个数, N 为源样本的个数。

8. 根据权利要求6所述的基于对比放大网络的微表情识别方法,其特征在于:所述类间对比损失具体函数如下:

$$\mathcal{L}_{inter-class} = \sum_{k \in I} \frac{-1}{|P(k)|} \sum_{p \in P(k)} \log \frac{\exp(v_k \cdot v_p / \tau)}{\sum_{a \in A(k)} \exp(v_k \cdot v_a / \tau)}$$

式中, I 为微表情数据库, $P(k) \equiv \{p \in A(k) : y_p = y_k\}$ 表示在 $A(k)$ 中与当前样本*k*有相同微

表情类别标签的样本集合, $A(k)$ 表示当前样本的正样本与所对应的负样本构成的集合, $v = \text{LSTM}(z_1, z_2, \dots, z_N)$, $\text{LSTM}(z_1, z_2, \dots, z_N)$ 表示用长短时记忆网络LSTM将 z_1, z_2, \dots, z_N 整合成该视频样本的特征向量, z_1, z_2, \dots, z_N 分别表示第1, 2, \dots , N个源样本输入对比放大网络后多层感知机输出的特征向量, τ 为温度系数, v_k, v_p, v_a 分别表示当前以下标为k的视频样本经LSTM整合后的特征向量、同一批样本中与 v_k 具有相同标签的样本的特征向量、同一批样本中除了 v_k 之外的其他样本的特征向量。

9. 根据权利要求6所述的基于对比放大网络的微表情识别方法, 其特征在于: 所述交叉熵损失具体函数如下:

$$\mathcal{L}_{\text{classifier}} = - \sum_{c=1}^C y_c \log(p_c)$$

式中, C 为微表情类别数, y_c 为类别 c 的标签值, p_c 为属于类别 c 的概率。

10. 一种基于对比放大网络的微表情识别装置, 其特征在于: 包括处理器及存储在存储器上并可在处理器上运行的计算机程序, 其特征在于: 所述处理器执行所述程序时实现权利要求1-9中任意一项所述的方法。

基于对比放大网络的微表情识别方法及装置

技术领域

[0001] 本发明涉及图像处理技术,尤其涉及一种基于对比放大网络的微表情识别方法及装置。

背景技术

[0002] 微表情是一种自发的、短暂的面部表情,它可以隐藏人类真实的情绪状态。在此基础上,微表情的识别对情感计算和心理治疗具有重要意义。然而,与宏观的面部表情相比,它由细微的面部运动诱发,并且出现的时间短得多,自动的微表情识别因此成为一项具有挑战性的任务。

[0003] 微表情是一种动态的面部肌肉运动的过程,对运动的捕捉对于准确识别微表情是必不可少的。然而,由于微表情的强度较低,运动特征的提取非常困难。针对这一问题,许多方法采用运动放大技术对微表情进行放大,使面部运动更加显著。一些传统的放大技术和深度学习滤波器,如欧拉运动放大(EMM),全局拉格朗日运动放大(GLMM),基于学习的运动放大(LMM),已经证明通过放大微表情强度可以进一步提升微表情识别的性能。尽管有这种优势存在,但现有的放大方法在适应性上还并不完善。不同的被试在呈现不同的表情状态时,面部肌肉的变化趋势是不一致的,一个统一的放大水平无法适配所有微表情样本。例如,相同的放大系数对于一个微表情样本可能不够充分,放大结果不足以显示该表情的情感类别,而在另一个微表情样本上可能会放大过度,从而引入噪声。因此,现有技术中基于对比放大的微表情识别的准确性仍然有待提高。

发明内容

[0004] 发明目的:本发明针对现有技术存在的问题,提供一种准确性更高的基于对比放大网络的微表情识别方法及装置。

[0005] 技术方案:本发明所述的基于对比放大网络的微表情识别方法包括:

[0006] (1) 获取微表情数据库,所述微表情数据库中包含有若干微表情视频和对应的微表情类别标签;

[0007] (2) 将数据库中的每个微表情视频转换为一个微表情帧序列,并对微表情帧序列进行预处理,从每个微表情帧序列中采样固定数量的帧作为源样本;

[0008] (3) 对于每个源样本,计算其所在帧序列中剩余帧与该源样本在嵌入空间的距离,并将距离映射成概率,从而得到帧序列中每个源样本与剩余帧的距离概率分布;

[0009] (4) 对于每个源样本,根据计算得到的概率分布从剩余的视频帧中采样多个帧作为对应的负样本;

[0010] (5) 构建对比放大网络,包括没有全连接层的Resnet-18、连接在其后面的多层感知机和连接在多层感知机后面的softmax层;

[0011] (6) 将微表情帧序列中每个源样本进行数据增强形成一对样本,即锚样本和正样本,将锚样本、正样本和对应的负样本作为训练样本输入对比放大网络进行训练,训练时的

损失函数为视频内对比损失、类间对比损失、交叉熵损失之和,通过梯度下降优化训练;

[0012] (7) 将待识别的微表情视频按照步骤(2)预处理后输入训练好的对比放大网络,识别出微表情类别。

[0013] 进一步的,步骤(2)中所述预处理包括人脸配准和面部区域剪切。

[0014] 进一步的,步骤(2)中所述从每个微表情帧序列中采样固定数量的帧作为源样本具体包括:

[0015] 对于一个预处理后的微表情帧序列,根据均匀分布采样N帧作为源样本,均匀分布的形式为:

$$[0016] \quad f(n) = \begin{cases} \frac{1}{N}, & 1 \leq n \leq N \\ 0, & \text{其他} \end{cases}$$

[0017] 式中, $f(n)$ 表示对源样本采样的概率密度函数, n 表示采样的源样本的索引号, N 表示采样的源样本的数量。

[0018] 进一步的,步骤(3)具体包括:

[0019] (3-1) 对于每个微表情帧序列,使用已经在宏表情数据库上预训练好的网络模型:Resnet-18提取源样本和剩余帧的特征向量 u ,数据库的名称为FER+;

$$[0020] \quad u_i = g(x_i), i = 1, \dots, N$$

$$[0021] \quad u_{ij} = g(x_{ij}), j = 1, \dots, ni$$

[0022] 式中, $g(\cdot)$ 为神经网络, N 表示源样本的数量, x_i 、 x_{ij} 表示第 i 个源样本和 x_i 的第 j 个剩余帧, u_i 、 u_{ni} 表示 x_i 、 x_{ij} 的特征向量, ni 表示 x_i 的剩余帧数量;

[0023] (3-2) 对于每个微表情帧序列,按照下式计算该帧序列中剩余帧与源样本的特征向量在嵌入空间的距离:

$$[0024] \quad d_{ij} = u_i \cdot u_{ij}, i = 1, \dots, N, j = 1, \dots, ni$$

[0025] 式中, d_{ij} 表示第 i 个源样本与其第 j 个剩余帧在嵌入空间的距离;

[0026] (3-3) 对于每个微表情帧序列,按照下式使用softmax函数将不同的距离映射成概率,从而得到每个源样本与剩余帧的距离概率分布:

$$[0027] \quad p_{ij} = \frac{e^{d_{ij}}}{\sum_l e^{d_{il}}}, i = 1, \dots, N, j = 1, \dots, ni$$

$$[0028] \quad P = \{p_{ij} | i = 1, \dots, N, j = 1, \dots, ni\}$$

[0029] 式中, p_{ij} 表示 d_{ij} 映射的概率, P 表示源样本与剩余帧的距离概率分布。

[0030] 进一步的,步骤(4)具体包括:对于每个源样本,从其距离概率分布中进行采样,将采样得到的概率对应的视频帧作为负样本。。

[0031] 进一步的,步骤(6)中所述损失函数具体为:

$$[0032] \quad \mathcal{L} = \lambda_1 \mathcal{L}_{intra-video} + \lambda_2 \mathcal{L}_{inter-clas} + \lambda_3 \mathcal{L}_{classifier}$$

[0033] 其中, \mathcal{L} 为总损失, $\lambda_1, \lambda_2, \lambda_3$ 为三个损失函数的权重, $\mathcal{L}_{intra-video}$ 为视频内对比损

失, $\mathcal{L}_{inter-clas}$ 为类间对比损失, $\mathcal{L}_{classifier}$ 为交叉熵损失。

[0034] 进一步的, 所述视频内对比损失具体函数如下:

$$\mathcal{L}_{intra-video} = \sum_{k \in I} \frac{-1}{N} \sum_{i=1}^N \left[\log \frac{\text{sim}((\tilde{z}_1)_i^k, (\tilde{z}_2)_i^k)}{\text{sim}((\tilde{z}_1)_i^k, (\tilde{z}_2)_i^k) + \sum_{j=1}^S \exp\left((\tilde{z}_1)_i^k \cdot \frac{(\tilde{z}_j^-)_i^k}{\tau}\right)} \right]$$

[0035]

$$+ \log \frac{\text{sim}((\tilde{z}_1)_i^k, (\tilde{z}_2)_i^k)}{\text{sim}((\tilde{z}_1)_i^k, (\tilde{z}_2)_i^k) + \sum_{j=1}^S \exp\left((\tilde{z}_2)_i^k \cdot \frac{(\tilde{z}_j^-)_i^k}{\tau}\right)}$$

[0036] 式中, I 为微表情数据库, $\text{sim}((\tilde{z}_1)_i^k, (\tilde{z}_2)_i^k) = \exp((\tilde{z}_1)_i^k \cdot (\tilde{z}_2)_i^k / \tau)$, $(\tilde{z}_1)_i^k, (\tilde{z}_2)_i^k, (\tilde{z}_j^-)_i^k$ 分别为 $(\tilde{x}_1)_i^k, (\tilde{x}_2)_i^k, (x_j^-)_i^k$ 输入对比放大网络后多层感知机输出的特征向量, $(\tilde{x}_1)_i^k, (\tilde{x}_2)_i^k, (x_j^-)_i^k$ 分别为 I 中第 k 个微表情视频的第 i 个源样本经过数据增强得到的正样本、锚样本、第 j 个负样本, τ 为温度系数, S 为负样本的个数, N 为源样本的个数。

[0037] 进一步的, 所述类间对比损失具体函数如下:

$$\mathcal{L}_{inter-clas} = \sum_{k \in I} \frac{-1}{|P(k)|} \sum_{p \in P(k)} \log \frac{\exp(v_k \cdot v_p / \tau)}{\sum_{a \in A(k)} \exp(v_k \cdot v_a / \tau)}$$

[0039] 式中, I 为微表情数据库, $P(k) \equiv \{p \in A(k) : y_p = y_k\}$ 表示在 $A(k)$ 中与当前样本 k 有相同微表情类别标签的样本集合, $A(k)$ 表示当前样本的正样本与所对应的负样本构成的集合, $v = \text{LSTM}(z_1, z_2, \dots, z_N)$, $\text{LSTM}(z_1, z_2, \dots, z_N)$ 表示用长短时记忆网络 LSTM 将 z_1, z_2, \dots, z_N 整合成该视频样本的特征向量, z_1, z_2, \dots, z_N 分别表示第 1, 2, \dots , N 个源样本输入对比放大网络后多层感知机输出的特征向量, τ 为温度系数, v_k, v_p, v_a 分别表示当前以下标为 k 的视频样本经 LSTM 整合后的特征向量、同一批样本中与 v_k 具有相同标签的样本的特征向量、同一批样本中除了 v_k 之外的其他样本的特征向量。

[0040] 进一步的, 所述交叉熵损失具体函数如下:

$$\mathcal{L}_{classifier} = - \sum_{c=1}^C y_c \log(p_c)$$

[0042] 式中, C 为微表情类别数, y_c 为类别 c 的标签值, p_c 为属于类别 c 的概率。

[0043] 本发明所述的基于对比放大网络的微表情识别装置包括处理器及存储在存储器上并可在处理器上运行的计算机程序, 所述处理器执行所述程序时实现上述方法。

[0044] 有益效果: 本发明与现有技术相比, 其显著优点是: 本发明采用对比学习的思想, 通过构造正负样本, 可以使网络优化过程中更集中于区分正负样本的不同, 扩大它们在嵌入空间中的距离, 使得强度对比和类别对比更加明显, 就能直接产生更直接的判别特征, 从而使网络能够更顺畅地感知运动变化, 提高识别准确率, 且减少人为设定的超参数, 更方便。

附图说明

[0045] 图1是本发明提供的基于对比放大网络的微表情识别方法的一个实施例的流程示意图；

[0046] 图2是本发明设计的对比放大网络的示意图；

[0047] 图3是负样本生成过程的示意图。

具体实施方式

[0048] 本实施例提供了一种基于对比放大网络的微表情识别方法,如图1所示,包括:

[0049] (1) 获取微表情数据库,所述微表情数据库中包含有若干微表情视频和对应的微表情类别标签。为了提高识别准确率,可以采用多个常用数据库。

[0050] (2) 将数据库中的每个微表情视频转换为一个微表情帧序列,并对微表情帧序列进行预处理,从每个微表情帧序列中采样固定数量的帧作为源样本。

[0051] 其中,所述预处理包括人脸配准和面部区域剪切。所述从每个微表情帧序列中采样固定数量的帧作为源样本具体包括:

[0052] 对于一个预处理后的微表情帧序列,根据均匀分布采样N帧作为源样本,均匀分布的形式为:

$$[0053] \quad f(n) = \begin{cases} \frac{1}{N}, & 1 \leq n \leq N \\ 0, & \text{其他} \end{cases}$$

[0054] 式中, $f(n)$ 表示对源样本采样的概率密度函数, n 表示采样的源样本的索引号, N 表示采样的源样本的数量。

[0055] (3) 对于每个源样本,计算其所在帧序列中剩余帧与该源样本在嵌入空间的距离,并将距离映射成概率,从而得到帧序列中每个源样本与剩余帧的距离概率分布。

[0056] 该步骤具体包括:

[0057] (3-1) 对于每个微表情帧序列,使用预训练好的神经网络提取源样本和剩余帧的特征向量 u :

$$[0058] \quad u_i = g(x_i), i = 1, \dots, N$$

$$[0059] \quad u_{ij} = g(x_{ij}), j = 1, \dots, ni$$

[0060] 式中, $g(\cdot)$ 为神经网络,例如可以是已经在宏表情数据库上预训练好的网络Resnet-18,预训练使用数据集FER+; N 表示源样本的数量, x_i 、 x_{ij} 表示第*i*个源样本和 x_i 的第*j*个剩余帧, u_i 、 u_{ni} 表示 x_i 、 x_{ij} 的特征向量, ni 表示 x_i 的剩余帧数量;

[0061] (3-2) 对于每个微表情帧序列,按照下式计算该帧序列中剩余帧与源样本的特征向量在嵌入空间的距离:

$$[0062] \quad d_{ij} = u_i \cdot u_{ij}, i = 1, \dots, N, j = 1, \dots, ni$$

[0063] 式中, d_{ij} 表示第*i*个源样本与其第*j*个剩余帧在嵌入空间的距离;

[0064] (3-3) 对于每个微表情帧序列,按照下式使用softmax函数将不同的距离映射成概率,从而得到每个源样本与剩余帧的距离概率分布:

$$[0065] \quad p_{ij} = \frac{e^{d_{ij}}}{\sum_l e^{d_{il}}}, \quad i = 1, \dots, N, j = 1, \dots, ni$$

$$[0066] \quad P = \{p_{ij} | i = 1, \dots, N, j = 1, \dots, ni\}$$

[0067] 式中, p_{ij} 表示 d_{ij} 映射的概率, P 表示源样本与剩余帧的距离概率分布。

[0068] 这样就得到了一个源样本与剩余所有视频帧的以距离为离散变量的概率分布, 即: 与源样本在特征层面上距离越远的, 被选择作为负样本的概率也越大。在同一个微表情序列中, 不同帧之间的差异实际上只反映在强度信息上, 帧之间的距离越大意味着强度差异就越明显, 通过计算源样本与剩余的视频帧的距离, 在采样负样本的时候可以保证距离较远的, 或者说强度差异较大的视频帧有更大的概率被采样到。

[0069] (4) 对于每个源样本, 根据计算得到的概率分布从剩余的视频帧中采样多个帧作为对应的负样本, 如图3所示。

[0070] 步骤(4)具体包括: 对于每个源样本, 从其距离概率分布中进行采样, 将采样得到的概率对应的视频帧作为负样本。

[0071] (5) 构建对比放大网络, 如图2所示, 包括没有全连接层的Resnet-18、连接在其后面的多层感知机和连接在多层感知机后面的softmax层。

[0072] (6) 将微表情帧序列中每个源样本进行数据增强形成一对样本, 即锚样本和正样本, 将锚样本、正样本和对应的负样本作为训练样本输入对比放大网络进行训练, 训练时的损失函数为视频内对比损失、类间对比损失、交叉熵损失之和, 通过梯度下降优化训练。

[0073] 其中, 所述损失函数具体为:

$$[0074] \quad \mathcal{L} = \lambda_1 \mathcal{L}_{intra-video} + \lambda_2 \mathcal{L}_{inter-class} + \lambda_3 \mathcal{L}_{classifier}$$

[0075] 其中, \mathcal{L} 为总损失, $\lambda_1, \lambda_2, \lambda_3$ 为三个损失函数的权重, $\mathcal{L}_{intra-video}$ 为视频内对比损失, $\mathcal{L}_{inter-class}$ 为类间对比损失, $\mathcal{L}_{classifier}$ 为交叉熵损失。

[0076] 微表情帧序列采样得到的每个源样本, 记为 x , 通过数据增强得到一对样本 \tilde{x}_1, \tilde{x}_2 , 在实际计算中, 称 \tilde{x}_1, \tilde{x}_2 互为锚样本和正样本, 即当 \tilde{x}_1 作为锚样本的时候, \tilde{x}_2 作为正样本, 反之亦然, 锚样本与正样本应该含有相同的强度信息, 同时可以在颜色, 风格等信息上不同, 所以在数据增强时选择高斯噪声和随机灰度变换。对比放大网络中没有全连接层的Resnet-18记为 $Enc(\cdot)$, 多层感知机记为 $Proj(\cdot)$, 将数据增强后的样本以及该源样本对应的所有负样本, 记为 x_j^- , 送入对比放大网络, 在特征层面上计算的损失函数为视频内对比损失, 具体函数如下:

$$[0077] \quad \mathcal{L}_{intra-video} = \sum_{k \in I} \frac{-1}{N} \sum_{i=1}^N \left[\log \frac{\text{sim}((\tilde{z}_1)_i^k, (\tilde{z}_2)_i^k)}{\text{sim}((\tilde{z}_1)_i^k, (\tilde{z}_2)_i^k) + \sum_{j=1}^S \exp\left((\tilde{z}_1)_i^k \cdot \frac{(z_j^-)_i^k}{\tau}\right)} \right. \\ \left. + \log \frac{\text{sim}((\tilde{z}_1)_i^k, (\tilde{z}_2)_i^k)}{\text{sim}((\tilde{z}_1)_i^k, (\tilde{z}_2)_i^k) + \sum_{j=1}^S \exp\left((\tilde{z}_2)_i^k \cdot \frac{(z_j^-)_i^k}{\tau}\right)} \right]$$

[0078] 式中，I 为微表情数据库， $\text{sim}((\tilde{z}_1)_i^k, (\tilde{z}_2)_i^k) = \exp((\tilde{z}_1)_i^k \cdot (\tilde{z}_2)_i^k / \tau)$ ， $(\tilde{z}_1)_i^k, (\tilde{z}_2)_i^k, (z_j^-)_i^k$ 分别为 $(\tilde{x}_1)_i^k, (\tilde{x}_2)_i^k, (x_j^-)_i^k$ 输入对比放大网络后多层感知机输出的特征向量， $(\tilde{x}_1)_i^k, (\tilde{x}_2)_i^k, (x_j^-)_i^k$ 分别为 I 中第 k 个微表情视频的第 i 个源样本经过数据增强得到的正样本、锚样本、第 j 个负样本， τ 为温度系数，S 为负样本的个数，N 为源样本的个数。

[0079] 所述类间对比损失具体函数如下：

$$[0080] \quad \mathcal{L}_{inter-clas} = \sum_{k \in I} \frac{-1}{|P(k)|} \sum_{p \in P(k)} \log \frac{\exp(v_k \cdot v_p / \tau)}{\sum_{a \in A(k)} \exp(v_k \cdot v_a / \tau)}$$

[0081] 式中，I 为微表情数据库， $P(k) = \{p \in A(k) : y_p = y_k\}$ 表示在 A(k) 中与当前样本 k 有相同微表情类别标签的样本集合，A(k) 表示当前样本的正样本与所对应的负样本构成的集合， $v = \text{LSTM}(z_1, z_2, \dots, z_N)$ ， $\text{LSTM}(z_1, z_2, \dots, z_N)$ 表示用长短时记忆网络 LSTM 将 z_1, z_2, \dots, z_N 整合成该视频样本的特征向量， z_1, z_2, \dots, z_N 分别表示第 1, 2, \dots , N 个源样本输入对比放大网络后多层感知机输出的特征向量， τ 为温度系数， v_k, v_p, v_a 分别表示当前以下标为 k 的视频样本经 LSTM 整合后的特征向量、同一批样本中与 v_k 具有相同标签的样本的特征向量、同一批样本中除了 v_k 之外的其他样本的特征向量。

[0082] 只有类间损失并不能使网络知道样本的具体类别，所以需要有一个 softmax 函数指导分类，为交叉熵损失，具体函数如下：

$$[0083] \quad \mathcal{L}_{classifier} = - \sum_{c=1}^C y_c \log(p_c)$$

[0084] 式中，C 为微表情类别数， y_c 为类别 c 的标签值， p_c 为属于类别 c 的概率。

[0085] (7) 将待识别的微表情视频按照步骤 (2) 预处理后输入训练好的对比放大网络，识别出微表情类别。

[0086] 本实施例还提供了一种基于对比放大网络的微表情识别装置，包括处理器及存储在存储器上并可在处理器上运行的计算机程序，所述处理器执行所述程序时实现上述方法。

[0087] 为验证本发明的有效性，在 CAME2 微表情数据库、SAMM 微表情数据库和 SMIC 数据库的 HS 子数据库之间做了微表情识别，验证结果及与其他最新方法的比较如表 1 所示：

[0088] 表 1

[0089]

方法	年份	CASME2	SAMM	SMIC-HS
LBP-SIP	2014	45.36	36.76	42.12
MagGA	2018	63.30	N/A	N/A
DSSN	2019	70.78	57.35	63.41
TSCNN-I	2020	74.05	63.53	72.74
LBPAccP ^{u2}	2021	69.03	N/A	76.59
AU-GCN	2021	74.27	74.26	N/A
本方法	2022	79.03	77.21	77.91

[0090] 表 1 中，N/A 表示没有相关记录。

[0091] 对CASME2数据库的表情做如下处理:将样本个数少于10个的类别省略,避免样本严重不平衡问题。完成5种类别的识别任务,分别是happy、regression、disgust、fear和surprise对SAMM数据库的表情做如下处理:将样本个数少于10个的类别省略,避免样本严重不平衡问题,完成5种类别的识别任务,分别是、happiness、angry、disgust、fear和surprise。SMIC数据库本身类别为positive、negative、surprise,完成三分类的识别任务。

[0092] 实验结果表明,基于本发明提出的微表情识别方法取得了较高的微表情识别率。相较于传统的微表情放大方式,本发明可以免去部分超参数人为设置的繁琐,对个体的适应性更强,更方便。

[0093] 以上所揭露的仅为本发明一种较佳实施例而已,不能以此来限定本发明之权利范围,因此依本发明权利要求所作的等同变化,仍属本发明所涵盖的范围。

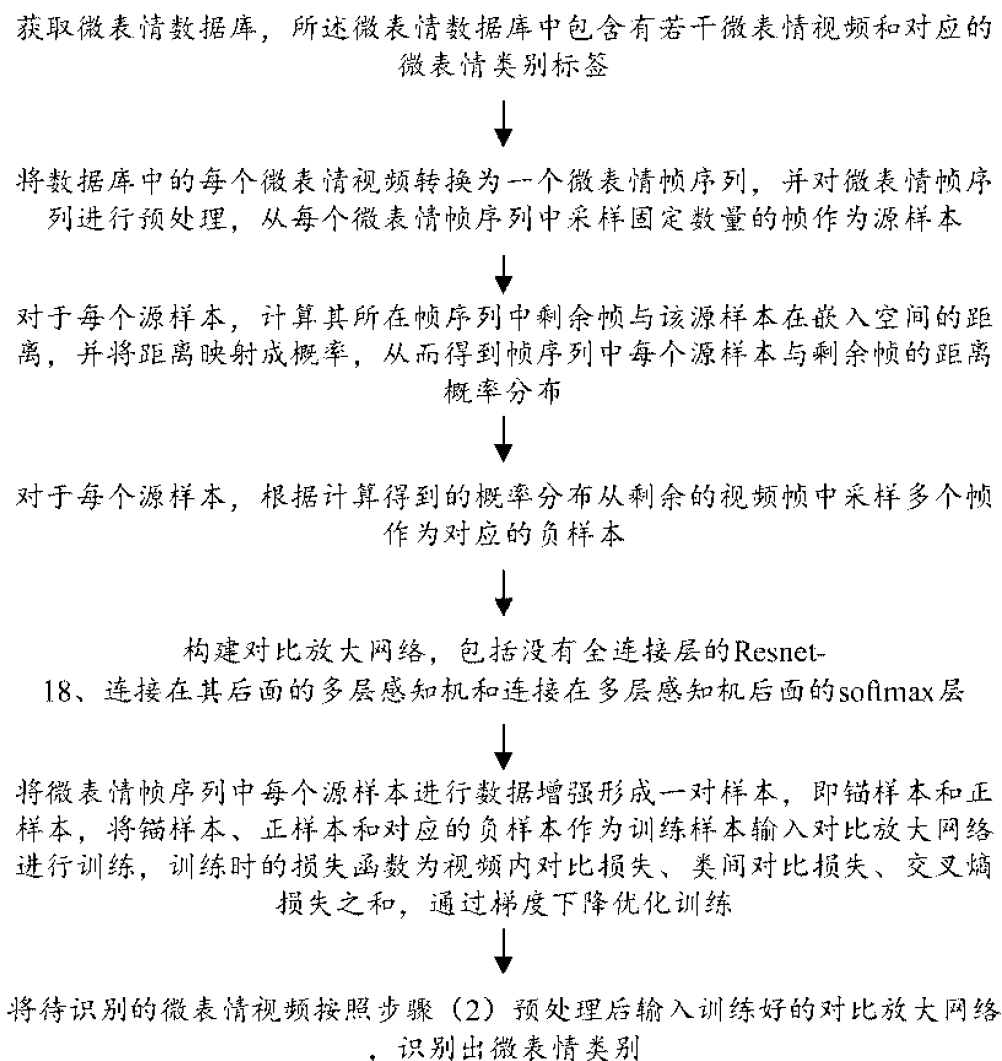


图1

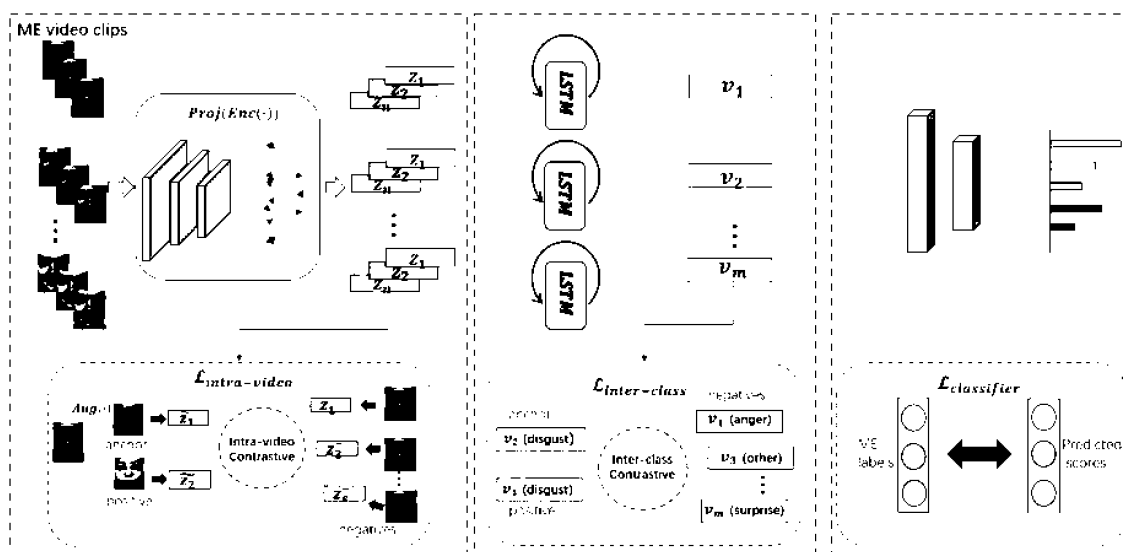


图2

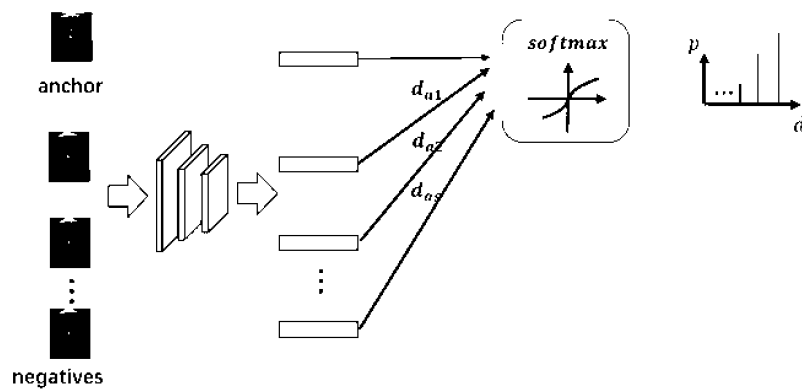


图3