



(12) 发明专利申请

(10) 申请公布号 CN 112101119 A

(43) 申请公布日 2020. 12. 18

(21) 申请号 202010831485.4

(22) 申请日 2020.08.18

(71) 申请人 东南大学

地址 210096 江苏省南京市玄武区四牌楼2号

(72) 发明人 郑文明 江星洵 宗源 夏万闯

(74) 专利代理机构 南京众联专利代理有限公司  
32206

代理人 李雪萍

(51) Int. Cl.

G06K 9/00 (2006.01)

G06K 9/62 (2006.01)

G06N 3/04 (2006.01)

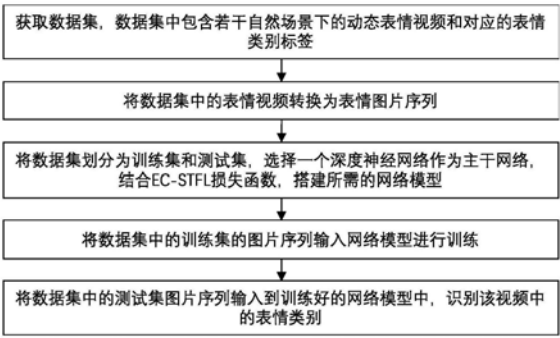
权利要求书2页 说明书4页 附图1页

(54) 发明名称

基于EC-STFL损失函数的自然场景动态表情识别方法及装置

(57) 摘要

本发明公开了一种基于EC-STFL损失函数的自然场景动态表情识别方法及装置,包括:(1)获取数据集,数据集中包含若干自然场景下的动态表情视频和对应的表情类别标签;(2)将数据集中的表情视频转换为表情图片序列;(3)将数据集划分为训练集和测试集,选择一个深度神经网络作为主干网络,结合EC-STFL损失函数,搭建网络模型;(4)将数据集中的训练集的图片序列输入网络模型进行训练;(5)将数据集中的测试集图片序列输入到训练好的网络模型中,识别该视频中的表情类别。本发明结合了基于表情聚集-时空特征学习(EC-STFL)的损失函数和交叉熵损失函数的优势,在批处理阶段同时聚集同类表情、分离不同类的表情,有效地提高了自然场景下的动态表情识别率。



1. 一种基于EC-STFL损失函数的自然场景动态表情识别方法,其特征在于,该方法包括:

(1) 获取数据集,数据集中包含若干自然场景下的动态表情视频和对应的表情类别标签;

(2) 将数据集中的表情视频转换为表情图片序列;

(3) 将数据集划分为训练集和测试集,选择一个深度神经网络作为主干网络,结合EC-STFL损失函数,搭建所需的网络模型;

(4) 将数据集中的测试集图片序列输入到训练好的网络模型中,识别该视频中的表情类别;

(5) 将数据集中的测试集图片序列输入到训练好的网络模型中,识别该视频中的表情类别。

2. 根据权利要求1所述的基于EC-STFL损失函数的自然场景动态表情识别方法,其特征在于:步骤(2)具体包括:

(2-1) 将数据库中的每个自然场景下的表情视频转换为表情图像序列;

(2-2) 检测所述表情图像序列中各帧的人脸,丢弃不包含人脸的帧,将包含人脸帧中的人脸部分切割出来;

(2-3) 结合(2-2)检测出的人脸关键点,对人脸表情图片进行仿射变换,将切割出的各帧人脸表情图片规范化;

(2-4) 利用插值和关键帧挑选算法对将由(2-3)获取的人脸图像进行处理,得到每个表情视频对应的相同帧数人脸图像。

3. 根据权利要求1所述的基于EC-STFL损失函数的自然场景动态表情识别方法,其特征在于:

(3-1) 所述的神经网络可为任一可进行视频分类的深度时空神经网络;

(3-2) 所述的EC-STFL损失函数应接在深度神经时空网络中最后几层全连接中的任一层之后;

(3-3) 所述的EC-STFL损失函数计算方法为:

$$L_{EC-STFL} = \sum_{1 \leq i \leq n} \frac{\sum_{1 \leq j \leq n, x_j \in N\{x_i\}} \frac{\|x_i - x_j\|}{|N\{x_i\}|}}{\sum_{1 \leq j \leq n, x_j \notin N\{x_i\}} \frac{\|x_i - x_j\|}{|N\{x_j\}|}}$$

式中 $L_{EC-STFL}$ 的计算方法为神经网络训练时一次批处理的损失函数计算,其中, $x_i \in \mathbb{R}^{N_F \times 1}$ 表示批处理中的任一样本,形如 $\mathbb{R}^{* \times *}$ 表示\*行\*列的矩阵; $n$ 表示批处理时包含的样本个数, $N\{x_i\}$ 表示批处理中与 $x_i$ 具有相同类别标签的样本集合, $|N\{x_i\}|$ 代表集合 $N\{x_i\}$ 的数目。

4. 根据权利要求3所述的基于EC-STFL损失函数的自然场景动态表情识别方法,其特征在于:步骤(3-3)中的 $L_{EC-STFL}$ 损失函数与softmax损失函数一起使用,对主干神经网络进行联合训练:

$$L=L_S+\lambda L_{EC-STFL}$$

其中, $L_S$ 表示softmax损失函数, $\lambda$ 表示损失函数 $L$ 中用来平衡损失函数 $L_{EC-STFL}$ 和损失函数 $L_S$ 的超参数。

5.一种基于EC-STFL损失函数的自然场景动态表情识别装置,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1-4中任意一项所述的方法。

## 基于EC-STFL损失函数的自然场景动态表情识别方法及装置

### 技术领域

[0001] 本发明涉及视频处理、深度学习技术领域,尤其涉及一种基于EC-STFL损失函数的自然场景动态表情识别方法及装置。

### 背景技术

[0002] 面部表情作为人类情感交流中的一种重要方式,一直是情感计算、人工智能研究中的热点话题。在过去二十年中,人们对面部表情图片和视频进行了大量研究。为了实现更加完美的、符合应用需求的表情识别系统,近年来,人们将研究目光从实验室场景下的表情转向了自然场景下的表情识别。与实验室场景相比,自然场景下的人脸表情识别受到来自不同光照、不同姿态、不同种族和其他不可预测、高挑战性场景因素的影响。因此,自然场景下的表情识别是一个挑战性极高的话题。自动表情识别在诸如心理疾病检测、教育评估、人机交互等领域有着巨大的潜在价值。目前许多研究者已经完成了实验场景下的表情识别分析。近来来,人们更加关注自然场景下的表情识别研究,因为自然场景下的表情识别包含了更多极端光照、姿态变化的情况。这些极具挑战性的使在实验室场景中训练出的模型效果变差。另一方面,与静态表情不同,动态表情包含了人脸肌肉运动的信息,也就是说,动态表情较之静态表情,更能反映出人的具体情感。

[0003] 深度学习作为推动近十年来人工智能浪潮的一类方法,已经在计算机视觉、自然语言处理、语音识别等领域取得了巨大成功,它也为真实场景下的动态表情识别指明了道路。以物体识别为例,一些基于卷积神经网络的方法,在诸如ImageNet等数据库上,已经达到了超过人类的识别能力的水平。然而,因为包含更多极端光照、姿态变化等情况,基于深度学习的自然场景动态表情识别仍然十分困难。

### 发明内容

[0004] 发明目的:本发明针对现有技术存在的问题,提供一种基于EC-STFL损失函数的自然场景动态表情识别方法及装置,识别准确率更高。

[0005] 技术方案:本发明所述的基于EC-STFL损失函数的自然场景动态表情识别方法包括:

[0006] (1) 获取数据集,数据集中包含若干自然场景下的动态表情视频和对应的表情类别标签;

[0007] (2) 将数据集中的表情视频转换为表情图片序列;

[0008] (3) 将数据集划分为训练集和测试集,选择一个深度神经网络作为主干网络,结合EC-STFL损失函数,搭建所需的网络模型;

[0009] (4) 将数据集中的测试集图片序列输入到训练好的网络模型中,识别该视频中的表情类别;

[0010] (5) 将数据集中的测试集图片序列输入到训练好的网络模型中,识别该视频中的表情类别。

[0011] 进一步的,步骤(2)具体包括:

[0012] (2-1)将数据库中的每个自然场景下的表情视频转换为表情图像序列;

[0013] (2-2)检测所述表情图像序列中各帧的人脸,丢弃不包含人脸的帧,将包含人脸帧中的人脸部分切割出来;

[0014] (2-3)结合(2-2)检测出的人脸关键点,对人脸表情图片进行仿射变换,将切割出的各帧人脸表情图片规范化;

[0015] (2-4)利用插值和关键帧挑选算法对将由(2-3)获取的人脸图像进行处理,得到每个表情视频对应的相同帧数人脸图像。

[0016] 进一步的,步骤(3)具体包括:

[0017] (3-1)所述的神经网络可为任一可进行视频分类的深度时空神经网络;

[0018] (3-2)所述的EC-STFL损失函数应接在深度神经时空网络中最后几层全连接中的任一层之后;

[0019] (3-3)所述的EC-STFL损失函数计算方法为:

$$[0020] \quad L_{EC-STFL} = \sum_{1 \leq i \leq n} \frac{\sum_{1 \leq j \leq n, x_j \in N\{x_i\}} \frac{\|x_i - x_j\|}{|N\{x_i\}|}}{\sum_{1 \leq j \leq n, x_j \notin N\{x_i\}} \frac{\|x_i - x_j\|}{|N\{x_j\}|}}$$

[0021] 式中 $L_{EC-STFL}$ 的计算方法为神经网络训练时一次批处理的损失函数计算,其中, $x_i \in \mathbb{R}^{N_F \times 1}$ 表示批处理中的任一样本,形如 $\mathbb{R}^{* \times *}$ 表示\*行\*列的矩阵。 $n$ 表示批处理时包含的样本个数, $N\{x_i\}$ 表示批处理中与 $x_i$ 具有相同类别标签的样本集合, $|N\{x_i\}|$ 代表集合 $N\{x_i\}$ 的数目。

[0022] 进一步的,步骤(3-3)中的 $L_{EC-STFL}$ 损失函数与softmax损失函数一起使用,对主干神经网络进行联合训练:

$$[0023] \quad L = L_S + \lambda L_{EC-STFL}$$

[0024] 其中, $L_S$ 表示softmax损失函数, $\lambda$ 表示损失函数 $L$ 中用来平衡损失函数 $L_{EC-STFL}$ 和损失函数 $L_S$ 的超参数。

[0025] 本发明所述的基于EC-STFL损失函数的自然场景动态表情识别装置,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。

[0026] 有益效果:本发明识别准确率更高。本发明结合了基于表情聚集-时空特征学习(EC-STFL)的损失函数和交叉熵损失函数的优势,在批处理阶段同时聚集同类表情、分离不同类的表情,有效地提高了自然场景下的动态表情识别率。

## 附图说明

[0027] 图1是本发明提供的基于EC-STFL损失函数的自然场景动态表情识别方法的一个实施例的流程示意图;

[0028] 图2是EC-STFL损失函数的示意图。

## 具体实施方式

[0029] 本实施例提供了一种基于EC-STFL损失函数的自然场景动态表情识别方法,如图1、图2所示,包括以下步骤:

[0030] (1) 获取数据集,数据集中包含若干自然场景下的动态表情视频和对应的表情类别标签。

[0031] (2) 将数据集中的表情视频转换为表情图片序列。

[0032] 该步骤具体包括:

[0033] (2-1) 将数据库中的每个自然场景下的表情视频转换为表情图像序列;

[0034] (2-2) 检测所述表情图像序列中各帧的人脸,丢弃不包含人脸的帧,将包含人脸帧中的人脸部分切割出来;人脸检测采用face++提供的人脸检测接口,其中可获人脸关键点信息;

[0035] (2-3) 结合(2-2)检测出的人脸关键点,对人脸表情图片进行仿射变换,将切割出的各帧人脸表情图片规范化;规范化采用Seetaface工具包完成;

[0036] (2-4) 利用插值和关键帧挑选算法对将由(2-3)获取的人脸图像进行处理,得到每个表情视频对应的相同帧数人脸图像。

[0037] (3) 将数据集划分为训练集和测试集,选择一个深度神经网络作为主干网络,结合EC-STFL损失函数,搭建所需的网络模型;

[0038] 该步骤具体包括:

[0039] (3-1) 选择C3D神经网络作为主干神经网络;

[0040] (3-2) 将EC-STFL损失函数接在C3D神经网络中倒数第二个全连接层(用于产生分类概率的全连接层之前的全连接层)之后;

[0041] (3-3) EC-STFL损失函数计算方法为:

$$[0042] \quad L_{EC-STFL} = \sum_{1 \leq i \leq n} \frac{\sum_{1 \leq j \leq n, x_j \in N\{x_i\}} \frac{\|x_i - x_j\|}{|N\{x_i\}|}}{\sum_{1 \leq j \leq n, x_j \notin N\{x_i\}} \frac{\|x_i - x_j\|}{|N\{x_j\}|}}$$

[0043] 式中 $L_{EC-STFL}$ 的计算方法为神经网络训练时一次批处理的损失函数计算,其中, $x_i \in \mathbb{R}^{N_f \times 1}$ 表示批处理中的任一样本,形如 $\mathbb{R}^{* \times *}$ 表示\*行\*列的矩阵。 $n$ 表示批处理时包含的样本个数, $N\{x_i\}$ 表示批处理中与 $x_i$ 具有相同类别标签的样本集合, $|N\{x_i\}|$ 代表集合 $N\{x_i\}$ 的数目。

[0044] (4) 将数据集中的测试集图片序列输入到训练好的网络模型中,识别该视频中的表情类别。

[0045] 步骤(3-3)中的 $L_{EC-STFL}$ 损失函数与softmax损失函数一起使用,对主干神经网络C3D进行联合训练:

$$[0046] \quad L = L_s + \lambda L_{EC-STFL}$$

[0047] 其中, $L_s$ 表示softmax损失函数, $\lambda$ 表示损失函数 $L$ 中用来平衡损失函数 $L_{EC-STFL}$ 和损失函数 $L_s$ 的超参数。

[0048] 本实施例还提供了一种基于EC-STFL损失函数的自然场景动态表情识别装置,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。

[0049] 为验证本发明的有效性,在DFEW自然场景动态表情数据库上进行测试,验证结果如表1所示:

[0050] 表1

[0051]	Method	Happy	Sad	Neutral	Angry	Surprise	Disgust	Fear	UAR	WAR
[0052]	C3D	75.17	39.49	<b>55.11</b>	<b>62.49</b>	45.00	1.38	20.51	42.74	53.54
	C3D,EC-STFL	<b>75.87</b>	<b>49.26</b>	54.81	61.53	<b>45.95</b>	<b>3.45</b>	24.83	<b>45.10</b>	<b>55.50</b>

[0053] 实验结果表明,基于本发明提出的自然场景动态表情识别方法,取得了较高的表情识别率。

[0054] 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。

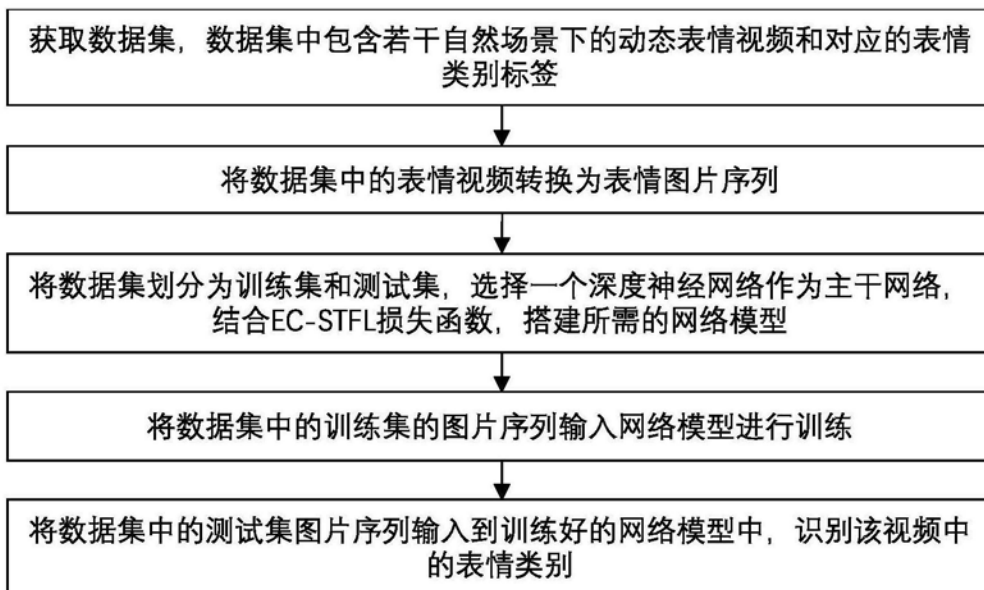


图1

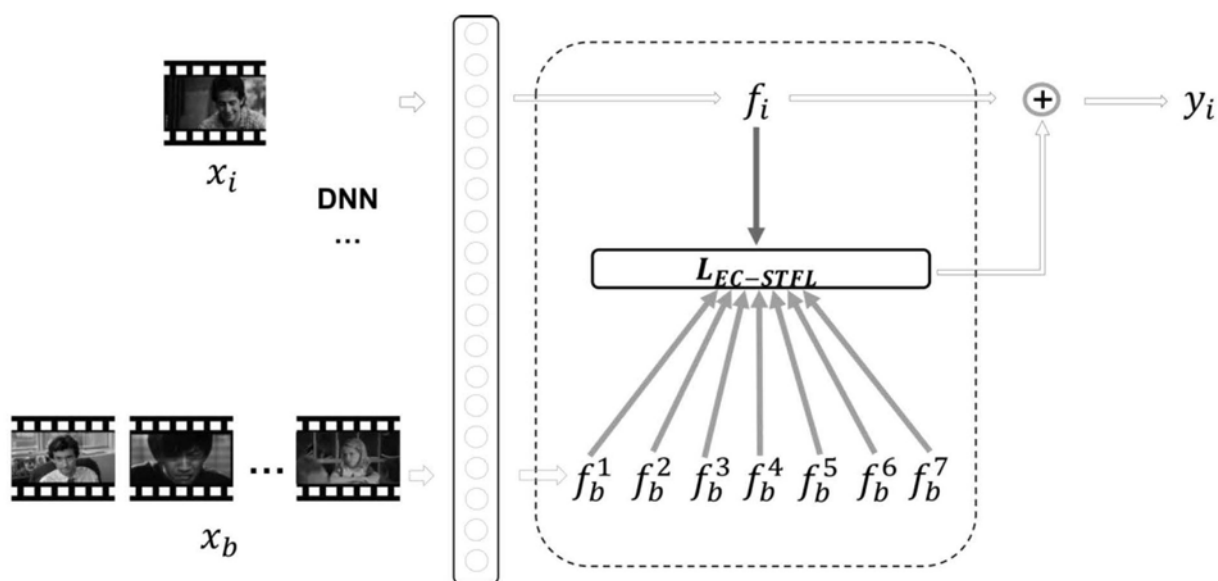


图2