(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 114187623 A
(43)申请公布日 2022.03.15

(21)申请号 202111304240.7

(22)申请日 2021.11.05

(71)申请人 中国科学院计算技术研究所
地址 100080 北京市海淀区中关村科学院南路6号

(72)发明人 李一帆 韩琥 山世光 陈熙霖

(74)专利代理机构 北京律诚同业知识产权代理有限公司 11006
代理人 祁建国

(51)Int.Cl.
G06V 40/16(2022.01)
G06V 40/20(2022.01)
G06V 10/774(2022.01)
G06V 10/764(2022.01)
G06V 10/82(2022.01)
G06K 9/62(2022.01)
G06N 3/04(2006.01)
G06N 3/08(2006.01)

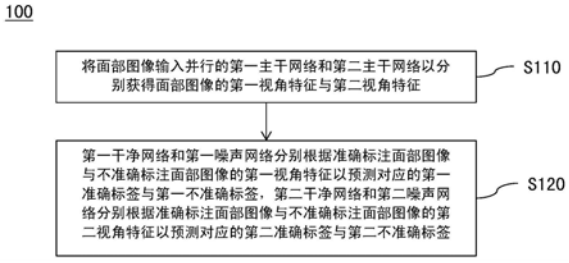
权利要求书3页 说明书10页 附图5页

(54)发明名称

面部动作单元识别方法与装置、模型训练方法与装置、存储介质及电子设备

(57)摘要

一种面部动作单元识别模型训练方法，包括：将面部图像输入并行的第一主干网络和第二主干网络以分别获得其第一视角特征与第二视角特征，面部图像包括准确标注面部图像、不准确标注面部图像或无标注面部图像；第一干净网络和第一噪声网络分别根据第一视角特征预测对应的第一准确标签与第一不准确标签，第二干净网络和第二噪声网络分别根据第二视角特征预测对应的第二准确标签与第二不准确标签；或者第一干净网络根据第一视角特征预测对应的第一准确标签与第一伪标签，第二干净网络根据第二视角特征预测对应的第二准确标签与第二伪标签。本发明的方法能利用准确标签数据集、不准确标签数据集以及无标注标签数据集，训练出准确率更高、泛化性能更强的面部动作单元识别模型。



CN 114187623 A

1.一种面部动作单元识别模型训练方法,其特征在于,包括:

将面部图像输入并行的第一主干网络和第二主干网络以分别获得所述面部图像的第一视角特征与第二视角特征,其中所述面部图像包括准确标注面部图像和不准确标注面部图像;

第一干净网络和第一噪声网络分别根据所述准确标注面部图像与所述不准确标注面部图像的第一视角特征以预测对应的第一准确标签与第一不准确标签,第二干净网络和第二噪声网络分别根据所述准确标注面部图像与所述不准确标注面部图像的第二视角特征以预测对应的第二准确标签与第二不准确标签。

2.如权利要求1所述的面部动作单元识别模型训练方法,其特征在于,还包括:

在网络收敛后,将所述第一噪声网络和所述第二噪声网络获得的所述第一不准确标签和所述第二不准确标签的预测均值与所述不准确标注面部图像的不准确标签进行线性组合获得新的不准确标签以对所述第一噪声网络和所述第二噪声网络进行重训练。

3.一种面部动作单元识别模型训练方法,其特征在于,包括:

将面部图像输入并行的第一主干网络和第二主干网络以分别获得所述面部图像的第一视角特征与第二视角特征,其中所述面部图像包括准确标注面部图像和无标注面部图像;

第一干净网络根据所述准确标注面部图像与所述无标注面部图像的第一视角特征以预测对应的第一准确标签与第一伪标签,第二干净网络根据所述准确标注面部图像与所述无标注面部图像的第二视角特征以预测对应的第二准确标签与第二伪标签。

4.如权利要求1-3任一所述的面部动作单元识别模型训练方法,其特征在于,使用准确标签损失对所述第一干净网络和所述第二干净网络进行约束,使用不准确标签损失对所述第一噪声网络和所述第二噪声网络进行约束;

所述准确标签损失为:

$$L_{clean}^{t}=\frac{1}{L}\sum_{k=1}^{L}\alpha_{clean}^{k}\left[y_{clean}^{k}\log\hat{p}_{clean}^{k}+(1-y_{clean}^{k})\log(1-\hat{p}_{clean}^{k})\right]$$

所述不准确标签损失为:

$$L_{noisy}^{t}=\frac{1}{L}\sum_{k=1}^{L}\alpha_{noisy}^{k}\left[y_{noisy}^{k}\log\hat{p}_{noisy}^{k}+(1-y_{noisy}^{k})\log(1-\hat{p}_{noisy}^{k})\right]$$

其中,t表示第t∈{1,2}个视角,y_{clean}^{k}表示第k个AU的准确标签,y_{noisy}^{k}表示第k个AU的不准确标签,$\hat{p}_{clean}^{k}$表示干净网络对于第k个AU的预测结果,$\hat{p}_{noisy}^{k}$表示噪声网络对于第k个AU的预测结果,α_{clean}^{k}和α_{noisy}^{k}表示用于数据平衡的参数。

5.如权利要求4所述的面部动作单元识别模型训练方法,其特征在于,使用正交损失对所述第一干净网络和所述第二干净网络的权重进行约束,使用一致性损失对所述第一干净网络和所述第二干净网络的预测结果进行约束;

所述正交损失为:

$$L_{mv}=\frac{1}{L}\sum_{k=1}^{L}\frac{\left(W_{clean}^{1,k}\right)^{T}W_{clean}^{2,k}}{\left\|W_{clean}^{1,k}\right\|\left\|W_{clean}^{2,k}\right\|}$$

其中，$W_{clean}^{t,k}=\left[w_{clean}^{t,k};b_{clean}^{t,k}\right]\in\mathbb{R}^{(D+1)\times 1},t=1,2,k=1,2,...,L$表示来自第t个视角的干净网络对于第k个AU的权重；

所述一致性损失为：

$$L_{cons}=\frac{1}{L}\sum_{k=1}^{L}\left[H\left(\frac{\hat{p}_{clean}^{1,2}+\hat{p}_{clean}^{2,k}}{2}\right)-\frac{H(\hat{p}_{clean}^{1,k})+H(\hat{p}_{clean}^{2,k})}{2}\right]$$

其中，H(p)＝-(plogp+(1-p)log(1-p))表示预测概率p的熵。

6.如权利要求1或3所述的面部动作单元识别模型训练方法，其特征在于，所述第一主干网络和所述第二主干网络采用ResNet34架构，所述第一干净网络、所述第一噪声网络、所述第二干净网络以及所述第二噪声网络采用全连接网络。

7.一种面部动作单元识别模型训练装置，其特征在于，包括：

特征采集单元，用于将面部图像输入并行的第一主干网络和第二主干网络以分别获得所述面部图像的第一视角特征与第二视角特征，其中所述面部图像包括准确标注面部图像和不准确标注面部图像；

标签预测单元，用于第一干净网络和第一噪声网络分别根据所述准确标注面部图像与所述不准确标注面部图像以预测对应的第一准确标签与第一不准确标签，第二干净网络和第二噪声网络分别根据所述准确标注面部图像与所述不准确标注面部图像的第二视角特征以预测对应的第二准确标签与第二不准确标签。

8.如权利要求7所述的面部动作单元识别模型训练装置，其特征在于，还包括：

重训练单元，用于将所述第一噪声网络和所述第二噪声网络获得的所述第一不准确标签和所述第二不准确标签的预测均值与所述不准确标注面部图像的不准确标签进行线性组合获得新的不准确标签以对所述第一噪声网络和所述第二噪声网络进行重训练。

9.一种面部动作单元识别模型训练装置，其特征在于，包括：

特征采集单元，用于将面部图像输入并行的第一主干网络和第二主干网络以分别获得所述面部图像的第一视角特征与第二视角特征，其中所述面部图像包括准确标注面部图像和无标注面部图像；

标签预测单元，用于第一干净网络根据所述准确标注面部图像与所述无标注面部图像的第一视角特征以预测对应的第一准确标签与第一伪标签，第二干净网络根据所述准确标注面部图像与所述无标注面部图像的第二视角特征以预测对应的第二准确标签与第二伪标签。

10.一种面部动作单元识别方法，其特征在于，包括：

将待识别面部图像输入并行的第一主干网络和第二主干网络以分别获得所述待识别面部图像的第一视角特征与第二视角特征；

将所述第一视角特征和所述第二视角特征分别输入通过如权利要求1-6任一所述的面部动作单元识别模型训练方法训练得到的所述第一干净网络与所述第二干净网络以预测

第一准确标签与第二准确标签；

通过sigmoid函数将所述第一准确标签与所述第二准确标签的均值变换为对应的概率值，判断所述概率值是否大于等于一阈值，如是则判别所述待识别面部图像对应的面部动作单元为激活，否则判别为抑制。

11.一种面部动作单元识别装置，其特征在于，包括：

特征采集单元，用于将待识别面部图像输入并行的第一主干网络和第二主干网络以分别获得所述待识别面部图像的第一视角特征与第二视角特征；

标签预测单元，用于将所述第一视角特征和所述第二视角特征分别输入通过如权利要求1-6任一所述的面部动作单元识别模型训练方法训练得到的所述第一干净网络与所述第二干净网络以预测第一准确标签与第二准确标签；

状态判别单元，用于通过sigmoid函数将所述第一准确标签与所述第二准确标签的均值变换为对应的概率值，判断所述概率值是否大于等于一阈值，如是则判别所述待识别面部图像对应的面部动作单元为激活，否则判别为抑制。

12.一种计算机可读存储介质，存储有计算机程序，其特征在于，所述计算机程序被处理器执行时，实现如权利要求1-6任一所述的面部动作单元识别模型训练方法和/或如权利要求10所述的面部动作单元识别方法。

13.一种电子设备，其特征在于，包括：处理器和存储器，所述存储器内存储有可在所述处理器运行的计算机程序，当所述计算机程序被所述处理器执行时，实现如权利要求1-6任一所述的面部动作单元识别模型训练方法和/或如权利要求10所述的面部动作单元识别方法。

面部动作单元识别方法与装置、模型训练方法与装置、存储介质及电子设备

技术领域

[0001] 本发明涉及计算机视觉领域，特别涉及一种面部动作单元识别方法与装置、模型训练方法与装置。

背景技术

[0002] 研究人员在1978年提出面部动作单元编码系统(Facial Action Coding System，FACS)，通过将面部动作编码成AU，可以更精确地描述面部动作的细微变化。FACS共包含近百个AU，不仅对于面部进行编码，而且对头部姿态也进行了编码。通过识别不同AU激活的程度，可以描述面部的动作，并进一步用于情感识别、谎言识别等领域。由于AU标注需要对标注人员进行大量训练，并且标注一幅图片往往需要花费大量时间和精力，因此想要获得大量AU标注数据需要消耗大量的人力、物力和财力，而通过AU识别算法可以很好地解决这些问题。

[0003] 现有阶段面部动作单元识别的算法根据训练数据的特点主要可以分为以下四类：监督学习算法(supervised learning)、弱监督学习算法(weakly supervised learning)、半监督学习算法(semi-supervised learning)以及自监督学习算法(self-supervised learning)。第一类监督学习算法主要利用面部图片以及AU标注信息，这类方法往往会引入面部局部区域特征(如面部感兴趣区域(region of interest，ROI))、AU之间关系、AU时序信息、表情、光流、特征点等信息进行建模，尽管这类方法具有较好的可解释性，但由于缺乏足够数据，对于小规模数据集容易出现过拟合的情况，模型泛化能力往往不是很好。第二类弱监督学习算法会利用与AU识别相关任务的标注或者带噪声的AU标签作为弱标注，而这类标注往往比较容易获得，如表情、特征点等等，从而可以利用到大规模的无标注数据。第三类半监督学习算法主要利用少量带标注数据以及大量无标注数据进行训练，这类算法的泛化性往往更好。第四类自监督学习算法主要利用大规模无标注数据集，通过挖掘无标注数据内部的信息，如时序信息、光流、面部之间的对比信息等等来得到具有较好表征能力的预训练模型，然后将该预训练模型用于下游任务进行微调可以获得较好的效果。

发明内容

[0004] 针对现有技术的不足，本发明的目的主要是解决现有AU数据集较小，如何同时利用准确标签数据集、带噪声标注的不准确标签数据集或无标注标签数据集，训练出准确率更高、泛化性能更强的面部动作单元(AU)识别模型的问题。

[0005] 为了实现上述目的，本发明提出一种面部动作单元识别模型训练方法，包括：

[0006] 将面部图像输入并行的第一主干网络和第二主干网络以分别获得所述面部图像的第一视角特征与第二视角特征，其中所述面部图像包括准确标注面部图像和不准确标注面部图像；

[0007] 第一干净网络和第一噪声网络分别根据所述准确标注面部图像与所述不准确标

注面部图像以预测对应的第一准确标签与第一不准确标签，第二干净网络和第二噪声网络分别根据所述准确标注面部图像与所述不准确标注面部图像的第二视角特征以预测对应的第二准确标签与第二不准确标签。

[0008] 上述的面部动作单元识别模型训练方法，还包括：

[0009] 将所述第一噪声网络和所述第二噪声网络获得的所述第一不准确标签和所述第二不准确标签的预测均值与所述不准确标注面部图像的不准确标签进行线性组合获得新的不准确标签以对所述第一噪声网络和所述第二噪声网络进行重训练。

[0010] 为了实现上述目的，本发明还提出一种面部动作单元识别模型训练方法，包括：

[0011] 将面部图像输入并行的第一主干网络和第二主干网络以分别获得所述面部图像的第一视角特征与第二视角特征，其中所述面部图像包括准确标注面部图像和无标注面部图像；

[0012] 第一干净网络根据所述准确标注面部图像与所述无标注面部图像的第一视角特征以预测对应的第一准确标签与第一伪标签，第二干净网络根据所述准确标注面部图像与所述无标注面部图像的第二视角特征以预测对应的第二准确标签与第二伪标签。

[0013] 上述的面部动作单元识别模型训练方法，其中，使用准确标签损失对所述第一干净网络和所述第二干净网络进行约束，使用不准确标签损失对所述第一噪声网络和所述第二噪声网络进行约束；

[0014] 所述准确标签损失为：

[0015] $$L_{clean}^{t}=\frac{1}{L}\sum_{k=1}^{L}\alpha_{clean}^{k}\left[y_{clean}^{k}\log\hat{p}_{clean}^{k}+(1-y_{clean}^{k})\log(1-\hat{p}_{clean}^{k})\right]$$

[0016] 所述不准确标签损失为：

[0017] $$L_{noisy}^{t}=\frac{1}{L}\sum_{k=1}^{L}\alpha_{noisy}^{k}\left[y_{noisy}^{k}\log\hat{p}_{noisy}^{k}+(1-y_{noisy}^{k})\log(1-\hat{p}_{noisy}^{k})\right]$$

[0018] 其中，t表示第$t\in\{1,2\}$个视角，y_{clean}^{k}表示第k个AU的准确标签，y_{noisy}^{k}表示第k个AU的不准确标签，$\hat{p}_{clean}^{k}$表示干净网络对于第k个AU的预测结果，$\hat{p}_{noisy}^{k}$表示噪声网络对于第k个AU的预测结果，α_{clean}^{k}和α_{noisy}^{k}表示用于数据平衡的参数。

[0019] 上述的面部动作单元识别模型训练方法，其中，使用正交损失对所述第一干净网络和所述第二干净网络的权重进行约束，使用一致性损失对所述第一干净网络和所述第二干净网络的预测结果进行约束；

[0020] 所述正交损失为：

[0021] $$L_{mv}=\frac{1}{L}\sum_{k=1}^{L}\frac{\left(W_{clean}^{1,k}\right)^{T}W_{clean}^{2,k}}{\left\|W_{clean}^{1,k}\right\|\left\|W_{clean}^{2,k}\right\|}$$

[0022] 其中，$W_{clean}^{t,k}=\left[w_{clean}^{t,k};b_{clean}^{t,k}\right]\in\mathbb{R}^{(D+1)\times1},t=1,2,k=1,2,...,L$表示来自第t个视角的干净网络对于第k个AU的权重；

[0023] 所述一致性损失为：

[0024] $$L_{cons}=\frac{1}{L}\sum_{k=1}^{L}\left[H\left(\frac{\hat{p}_{clean}^{1,2}+\hat{p}_{clean}^{2,k}}{2}\right)-\frac{H(\hat{p}_{clean}^{1,k})+H(\hat{p}_{clean}^{2,k})}{2}\right]$$

[0025] 其中,H(p)＝-(plogp+(1-p)log(1-p))表示预测概率p的熵。

[0026] 上述的面部动作单元识别模型训练方法,其中,所述第一主干网络和所述第二主干网络采用ResNet34架构,所述第一干净网络、所述第一噪声网络、所述第二干净网络以及所述第二噪声网络采用全连接网络。

[0027] 为了实现上述目的,本发明还提出一种面部动作单元识别模型训练装置,包括:

[0028] 特征采集单元,用于将面部图像输入并行的第一主干网络和第二主干网络以分别获得所述面部图像的第一视角特征与第二视角特征,其中所述面部图像包括准确标注面部图像和不准确标注面部图像;

[0029] 标签预测单元,用于第一干净网络和第一噪声网络分别根据所述准确标注面部图像与所述不准确标注面部图像以预测对应的第一准确标签与第一不准确标签,第二干净网络和第二噪声网络分别根据所述准确标注面部图像与所述不准确标注面部图像的第二视角特征以预测对应的第二准确标签与第二不准确标签。

[0030] 上述的面部动作单元识别模型训练装置,还包括:

[0031] 重训练单元,用于将所述第一噪声网络和所述第二噪声网络获得的所述第一不准确标签和所述第二不准确标签的预测均值与所述不准确标注面部图像的不准确标签进行线性组合获得新的不准确标签以对所述第一噪声网络和所述第二噪声网络进行重训练。

[0032] 为了实现上述目的,本发明还提出一种面部动作单元识别模型训练装置,包括:

[0033] 特征采集单元,用于将面部图像输入并行的第一主干网络和第二主干网络以分别获得所述面部图像的第一视角特征与第二视角特征,其中所述面部图像包括准确标注面部图像和无标注面部图像;

[0034] 标签预测单元,用于第一干净网络根据所述准确标注面部图像与所述无标注面部图像的第一视角特征以预测对应的第一准确标签与第一伪标签,第二干净网络根据所述准确标注面部图像与所述无标注面部图像的第二视角特征以预测对应的第二准确标签与第二伪标签。

[0035] 为了实现上述目的,本发明还提出一种面部动作单元识别方法,包括:

[0036] 将待识别面部图像输入并行的第一主干网络和第二主干网络以分别获得所述待识别面部图像的第一视角特征与第二视角特征;

[0037] 将所述第一视角特征和所述第二视角特征分别输入通过上述的面部动作单元识别模型训练方法训练得到的所述第一干净网络与所述第二干净网络以预测第一准确标签与第二准确标签;

[0038] 通过sigmoid函数将所述第一准确标签与所述第二准确标签的均值变换为对应的概率值,判断所述概率值是否大于等于一阈值,如是则判别所述待识别面部图像对应的面部动作单元为激活,否则判别为抑制。

[0039] 为了实现上述目的,本发明还提出一种面部动作单元识别装置,包括:

[0040] 特征采集单元,用于将待识别面部图像输入并行的第一主干网络和第二主干网络以分别获得所述待识别面部图像的第一视角特征与第二视角特征;

[0041] 标签预测单元，用于将所述第一视角特征和所述第二视角特征分别输入通过上述的面部动作单元识别模型训练方法训练得到的所述第一干净网络与所述第二干净网络以预测第一准确标签与第二准确标签；

[0042] 状态判别单元，用于通过sigmoid函数将所述第一准确标签与所述第二准确标签的均值变换为对应的概率值，判断所述概率值是否大于等于一阈值，如是则判别所述待识别面部图像对应的面部动作单元为激活，否则判别为抑制。

[0043] 为了实现上述目的，本发明还提出一种计算机可读存储介质，存储有计算机程序，所述计算机程序被处理器执行时，实现上述的面部动作单元识别模型训练方法和/或上述的面部动作单元识别方法。

[0044] 为了实现上述目的，本发明还提出一种电子设备，包括：处理器和存储器，所述存储器内存储有可在所述处理器运行的计算机程序，当所述计算机程序被所述处理器执行时，实现上述的面部动作单元识别模型训练方法和/或上述的面部动作单元识别方法。

[0045] 由以上方案可知，本发明的优点在于：本发明提出的技术方案可以利用准确标签数据、不准确标签数据以及无标注标签数据进行训练，实现更好的泛化性能和更高的识别精度。本发明技术方案的主要框架主要基于正则协同训练网络，两个视角用来学习相互独立的特征，通过视角一致性的约束可以利用到不准确标签数据以及无标注标签数据。另外，每个视角还用噪声网络作为正则化项对干净网络进行约束，防止干净网络出现过拟合，提高模型泛化性能。此外基于重训练的方法，利用模型收敛后的噪声网络的预测结果与不准确标签的线性组合对噪声网络进行重训练，在模型收敛后进一步提高面部动作单元识别精度。

附图说明

[0046] 图1A为本发明一实施例的面部动作单元识别模型训练方法的流程图。

[0047] 图1B为本发明另一实施例的面部动作单元识别模型训练方法的流程图

[0048] 图2A为对应图1A的面部动作单元识别模型的框架示意图。

[0049] 图2B为对应图1B的面部动作单元识别模型的框架示意图。

[0050] 图3A为本发明一实施例的面部动作单元识别模型训练装置的模块图。

[0051] 图3B为本发明另一实施例的面部动作单元识别模型训练装置的模块图。

[0052] 图4为本发明一实施例的面部动作单元识别方法的流程图。

[0053] 图5为本发明一实施例的面部动作单元识别模型的框架示意图。

[0054] 图6为本发明一实施例的面部动作单元识别模型的模块图。

[0055] 图7为本发明一实施例的电子设备的示意图。

具体实施方式

[0056] 为让本发明的上述特征和效果能阐述的更明确易懂，下文特举实施例，并配合说明书附图作详细说明如下。

[0057] 本发明提出了一种基于弱监督学习的面部动作单元识别模型训练方法，该方法可以利用少量准确的准确标签的数据、大量带不准确标签的数据以及无标注标签的数据作为训练集进行训练。尽管不准确标签数据带有噪声，但是仍然存在一些可供发掘的有用信息，

可以用来提高模型的识别准确度、加快模型收敛并且提高模型的泛化能力。与现有的基于深度学习的监督学习方法不同，本发明所提出的方法可以利用到大量噪声数据，从大量噪声数据中发掘有用信息，而现有的基于监督学习的方法主要利用表情、特征点这类与AU相关的标注作为额外信息，另外一些方法还会考虑到引入时序或光流信息，并结合面部局部信息进行训练。由于现有的AU数据集比较小，因此这类方法往往容易出现过拟合的现象，泛化性能并不出众。本发明所提出的方法主要利用大规模噪声数据中的有效信息，可以更好的提高网络的泛化性能。

[0058] 在本发明中，训练数据例如采用自然场景下的数据集EmotioNet，但不发明不以此为限，也可使用其他类型数据集；其中，训练数据集包括三个部分，一部分来自带手工标注的AU标签的面部图像数据，称之为准确标注面部图像，其具有准确标签；另一部分来自于无标注标签的面部图像，称之为无标注图像，其不具有标签；再一部分例如利用在准确标注面部图像数据集合上预训练的网络(例如ResNet34)对所述无标注图像进行标注后可以获得，称之为不准确标注面部图像，其具有不准确标签，这里对于不准确标注面部图像的获取方式并不做限制，可以是预训练模型获得，也可以来自于人工标注的噪声标注图像。优选的，还可以通过利用公开的面部识别引擎对面部RGB图像进行面部识别和5点(2个眼角、鼻尖和2个嘴角)定位，并裁剪出面部区域，将面部图像保存为例如256×256大小；另外，在训练过程中，对面部图像进行了图像增广，用来提升模型的泛化性能，包括将图像随机裁剪成例如224×224大小，并对图像进行随机水平翻转以及随机灰度化。

[0059] 参见图1A及图2A所示，本发明的实施例提出一种面部动作单元识别模型训练方法100，包括：步骤S110-S120。

[0060] 步骤S110，将面部图像输入并行的第一主干网络和第二主干网络以分别获得所述面部图像的第一视角特征与第二视角特征，其中所述面部图像包括准确标注面部图像和不准确标注面部图像。

[0061] 模型的网络整体结构包括两个视角，对于每一个视角，均包含一个主干网络、一个噪声网络以及一个干净网络；具体的，模型网络的第一视角包括第一主干网络、第一干净网络和第一噪声网络，模型网络的第二视角包括第二主干网络、第二干净网络和第二噪声网络。其中，第一主干网络和第二主干网络采用ResNet34架构，第一噪声网络、第二噪声网络、第一干净网络和第二干净网络均为全连接网络，优选的，网络的具体结构如下表1所示，但本发明并不以此为限。

[0062] 表1

[0063]

部件名称	层类型	卷积核	步长	输出大小
主干网络	conv1	7×7×64	2	64×128×128
	bn2d+relu			
	maxpool	3×3	2	64×64×64
	conv2_x	$\begin{bmatrix}3\times3\times64\\3\times3\times64\end{bmatrix}\times3$	1	64×64×64
	conv3_x	$\begin{bmatrix}3\times3\times128\\3\times3\times128\end{bmatrix}\times4$	1	128×32×32
	conv4_x	$\begin{bmatrix}3\times3\times256\\3\times3\times256\end{bmatrix}\times6$	1	256×16×16
	conv5_x	$\begin{bmatrix}3\times3\times512\\3\times3\times512\end{bmatrix}\times3$	2	512×8×8
	avgpool	-	-	512×1×1

[0064]

噪声网络/干净网络	fc	-	-	L

[0065] 将获得的上述训练数据(面部图像)输入并行的两个主干网络(第一主干网络和第二主干网络)以分别获得所述面部图像的两个视角的特征(第一视角特征与第二视角特征),其中面部图像包括准确标注面部图像和不准确标注面部图像。

[0066] 步骤S120,第一干净网络和第一噪声网络分别根据所述准确标注面部图像与所述不准确标注面部图像以预测对应的第一准确标签与第一不准确标签,第二干净网络和第二噪声网络分别根据所述准确标注面部图像与所述不准确标注面部图像的第二视角特征以预测对应的第二准确标签与第二不准确标签。

[0067] 在本实施例中,噪声网络用来学习图像特征到不准确标签的映射,并作为正则化项防止干净网络出现过拟合现象;干净网络用来学习图像特征到准确标签的映射,并且用于最终的图像分类。

[0068] 具体的,在第一视角下,将通过第一主干网络获得的准确标注面部图像的特征和不准确标注面部图像的特征分别输入第一干净网络与第一噪声网络中以预测对应的第一准确标签与第一不准确标签;在第二视角下,将通过第二主干网络获得的准确标注面部图像的特征和不准确标注面部图像的特征分别输入第二干净网络与第二噪声网络中以预测对应的第二准确标签与第二不准确标签。

[0069] 另外,在本发明的实施例中,在模型训练过程中,优选的,使用准确标签损失对干净网络进行约束,使用不准确标签损失对噪声网络进行约束;为了使两个视角学到相互独立的特征,采用正交损失对干净网络的权重参数进行约束;为了使两个视角的预测结果一致,采用一致性损失对两个视角中干净网络的预测结果进行约束。具体如下所示:

[0070] 使用准确标签损失对第一干净网络和第二干净网络进行约束,使用不准确标签损

失对第一噪声网络和第二噪声网络进行约束；

[0071] 准确标签损失为：

[0072] $$L_{clean}^{t}=\frac{1}{L}\sum_{k=1}^{L}\alpha_{clean}^{k}\left[y_{clean}^{k}\log\hat{p}_{clean}^{k}+(1-y_{clean}^{k})\log(1-\hat{p}_{clean}^{k})\right]$$

[0073] 不准确标签损失为：

[0074] $$L_{noisy}^{t}=\frac{1}{L}\sum_{k=1}^{L}\alpha_{noisy}^{k}\left[y_{noisy}^{k}\log\hat{p}_{noisy}^{k}+(1-y_{noisy}^{k})\log(1-\hat{p}_{noisy}^{k})\right]$$

[0075] 其中，t表示第t∈{1,2}个视角，y_{clean}^{k}表示第k个AU的准确标签，y_{noisy}^{k}表示第k个AU的不准确标签，$\hat{p}_{clean}^{k}$表示干净网络对于第k个AU的预测结果，$\hat{p}_{noisy}^{k}$表示噪声网络对于第k个AU的预测结果，α_{clean}^{k}和α_{noisy}^{k}表示用于数据平衡的参数。

[0076] 使用正交损失对第一干净网络和第二干净网络的权重进行约束，使用一致性损失对第一干净网络和第二干净网络的预测结果进行约束；

[0077] 正交损失为：

[0078] $$L_{mv}=\frac{1}{L}\sum_{k=1}^{L}\frac{\left(W_{clean}^{1,k}\right)^{T}W_{clean}^{2,k}}{\left\|W_{clean}^{1,k}\right\|\left\|W_{clean}^{2,k}\right\|}$$

[0079] 其中，$W_{clean}^{t,k}=\left[w_{clean}^{t,k};b_{clean}^{t,k}\right]\in\mathbb{R}^{(D+1)\times1},t=1,2,k=1,2,...,L$表示来自第t个视角的干净网络对于第k个AU的权重；

[0080] 一致性损失为：

[0081] $$L_{cons}=\frac{1}{L}\sum_{k=1}^{L}\left[H\left(\frac{\hat{p}_{clean}^{1,2}+\hat{p}_{clean}^{2,k}}{2}\right)-\frac{H(\hat{p}_{clean}^{1,k})+H(\hat{p}_{clean}^{2,k})}{2}\right]$$

[0082] 其中，H(p)＝-(plogp+(1-p)log(1-p))表示预测概率p的熵。

[0083] 另外，在本发明的实施例中，优选的，在网络收敛后，将不准确标签的预测值以及不准确标签进行线性融合，用新生成的融合标签对两个噪声网络进行后期的训练，换言之，用新的不准确标签替代原不准确标签，即将第一噪声网络和第二噪声网络获得的第一不准确标签和第二不准确标签的预测均值与不准确标注面部图像的不准确标签进行线性组合获得新的不准确标签以对第一噪声网络和第二噪声网络进行重训练。示例如下：

[0084] 对于一张来自于噪声数据集的不准确标注面部图像x_{noisy}^{j}，其不准确标签可以表示为$y_{noisy}^{j}\in\mathbb{Z}^{L\times1}$，用来训练噪声网络的新的不准确标签$\tilde{y}_{noisy}^{j}$，其可以表示为不准确标签与两个视角的噪声网络预测均值的线性组合，公式如下：

[0085] $$\tilde{y}_{noisy}^{j}=\gamma y_{noisy}^{j}+(1-\gamma)\bar{p}_{noisy}^{j}$$

[0086] 其中，$\overline{p}_{noisy}^{j}=\frac{p_{noisy}^{1,j}+p_{noisy}^{2,j}}{2}\in\mathbb{R}^{L\times 1}$表示两个视角噪声网络预测均值，γ表示平衡不准确标签以及$\overline{p}_{noisy}^{j}$的权重，通常情况下设置为0.5，然不以此为限，得到的新的不准确标签通过阈值0.5变成“硬标签”。

[0087] 此外，第t个视角下由噪声网络产生的AU预测概率可以表示为如下形式

[0088] $$\hat{p}_{noisy}^{t}=\sigma\left(f_{clean}^{t}\oplus f_{noisy}^{t}\right)$$

[0089] 其中，⊕表示逐元素相加，f_{clean}^{t}和f_{noisy}^{t}分别表示在第t个视角下干净网络和噪声网络产生的特征。

[0090] 参见图1B及图2B所示，基于同样的发明构思，本发明的另一实施例提出一种面部动作单元识别模型训练方法100’，包括：步骤S110’-S120’。

[0091] 步骤S110’，将面部图像输入并行的第一主干网络和第二主干网络以分别获得所述面部图像的第一视角特征与第二视角特征，其中所述面部图像包括准确标注面部图像和无标注面部图像；

[0092] 步骤S120’，第一干净网络根据所述准确标注面部图像与所述无标注面部图像的第一视角特征以预测对应的第一准确标签与第一伪标签，第二干净网络根据所述准确标注面部图像与所述无标注面部图像的第二视角特征以预测对应的第二准确标签与第二伪标签。

[0093] 模型的网络整体结构包括两个视角，对于每一个视角，均包含一个主干网络和一个干净网络；具体的，模型网络的第一视角包括第一主干网络和第一干净网络，模型网络的第二视角包括第二主干网络和第二干净网络。其中，第一主干网络和第二主干网络采用ResNet34架构，第一干净网络和第二干净网络均为全连接网络，网络的具体结构可参照上述表1所示。

[0094] 此外，在本实施例中，干净网络用来学习准确标注面部图像特征到准确标签以及无标注面部图像特征到伪标签的映射，并且准确标签用于最终的图像分类。

[0095] 另外，在本实施例中，在模型训练过程中，优选的，使用准确标签损失对干净网络进行约束；为了使两个视角学到相互独立的特征，采用正交损失对干净网络的权重参数进行约束；为了使两个视角的预测结果一致，采用一致性损失对两个视角中干净网络的预测结果进行约束。具体的损失函数可参照上述实施例所示。

[0096] 参见图3A所示，基于相同的发明构思，本发明的实施例提出一种面部动作单元识别模型训练装置200，包括：

[0097] 特征采集单元210，用于将面部图像输入并行的第一主干网络和第二主干网络以分别获得所述面部图像的第一视角特征与第二视角特征，其中所述面部图像包括准确标注面部图像和不准确标注面部图像；

[0098] 标签预测单元220，用于第一干净网络和第一噪声网络分别根据所述准确标注面部图像与所述不准确标注面部图像以预测对应的第一准确标签与第一不准确标签，第二干净网络和第二噪声网络分别根据所述准确标注面部图像与所述不准确标注面部图像的第二视角特征以预测对应的第二准确标签与第二不准确标签。

[0099] 在一实施方式中，面部动作单元识别模型训练装置200，还包括：

[0100] 重训练单元230，用于将所述第一噪声网络和所述第二噪声网络获得的所述第一不准确标签和所述第二不准确标签的预测均值与所述不准确标注面部图像的不准确标签进行线性组合获得新的不准确标签以对所述第一噪声网络和所述第二噪声网络进行重训练。

[0101] 参见图3B所示，基于相同的发明构思，本发明的实施例提出一种面部动作单元识别模型训练装置200'，包括：

[0102] 特征采集单元210'，用于将面部图像输入并行的第一主干网络和第二主干网络以分别获得所述面部图像的第一视角特征与第二视角特征，其中所述面部图像包括准确标注面部图像和无标注面部图像；

[0103] 标签预测单元220'，用于第一干净网络根据所述准确标注面部图像与所述无标注面部图像的第一视角特征以预测对应的第一准确标签与第一伪标签，第二干净网络根据所述准确标注面部图像与所述无标注面部图像的第二视角特征以预测对应的第二准确标签与第二伪标签。

[0104] 通过上述方式完成了对面部动作单元识别模型的训练，下面对所述模型的应用进行说明。需要说明的是，所属领域的技术人员可以清楚地了解，为描述的方便和简洁，上述描述的方法、装置和模块的具体工作过程，可以参考前述方法实施例中的对应过程，在此不再赘述。

[0105] 参见图4及图5所示，基于相同的发明构思，本发明的实施例提出一种面部动作单元识别方法300，包括：步骤S310-S330。

[0106] 步骤S310，将待识别面部图像输入并行的第一主干网络和第二主干网络以分别获得所述待识别面部图像的第一视角特征与第二视角特征。

[0107] 其中，所述待识别面部图像为一无标签标注的面部图像。

[0108] 步骤S320，将所述第一视角特征和所述第二视角特征分别输入通过上述实施例的面部动作单元识别模型训练方法100、100'训练得到的所述第一干净网络与所述第二干净网络以预测第一准确标签与第二准确标签。

[0109] 步骤S330，通过sigmoid函数将所述第一准确标签与所述第二准确标签的均值变换为对应的概率值，判断所述概率值是否大于等于一阈值，如是则判别所述待识别面部图像对应的面部动作单元为激活，否则判别为抑制。

[0110] 具体的，将两个视角的干净网络的预测结果(第一准确标签与第二准确标签)的均值通过sigmoid函数变换成为相应的概率值，当概率值大于等于0.5则判别为所述AU被激活，否则判别为抑制，但本发明并不限制阈值的具体数值，其可根据实际进行调整。

[0111] 参见图6所示，基于相同的发明构思，本发明的实施例提出一种面部动作单元识别装置400，包括：

[0112] 特征采集单元410，用于将待识别面部图像输入并行的第一主干网络和第二主干网络以分别获得所述待识别面部图像的第一视角特征与第二视角特征；

[0113] 标签预测单元420，用于将所述第一视角特征和所述第二视角特征分别输入通过上述实施例的面部动作单元识别模型训练方法100、100'训练得到的所述第一干净网络与所述第二干净网络以预测第一准确标签与第二准确标签；

[0114] 状态判别单元430,用于通过sigmoid函数将所述第一准确标签与所述第二准确标签的均值变换为对应的概率值,判断所述概率值是否大于等于一阈值,如是则判别所述待识别面部图像对应的面部动作单元为激活,否则判别为抑制。

[0115] 参见图7所示,基于相同的发明构思,本发明的实施例还提出一种电子设备500,电子设备500例如为,但不限于个人计算机(PC)、智能手机、平板电脑、个人数字助理(Personal DigitalAssistant,PDA)、移动上网设备(Mobile Internet Device,MID)等,其包括处理器510和存储器520,处理器510与存储器520为直接或间接地电性连接,以实现数据的传输或交互。面部动作单元识别模型训练装置200、200'和/或面部动作单元识别模型400包括至少一个可以软件或固件(Firmware)的形式存储在存储器520中或固化在电子设备500的操作系统(Operating System,OS)中的软件模块。处理器510用于执行存储器520中存储的可执行模块,例如,面部动作单元识别模型训练装置200、200'包括的软件功能模块及计算机程序等,以实现面部动作单元识别模型训练方法100、100',抑或是,面部动作单元识别模型400包括的软件功能模块及计算机程序等,以实现面部动作单元识别方法300。处理器510在接收到执行指令后,执行计算机程序。

[0116] 基于相同的发明构思,本发明的实施例还提出一种计算机可读存储介质,存储有计算机程序,所述计算机程序被处理器执行时,实现上述实施例的面部动作单元识别模型训练方法100、100'和/或上述实施例的面部动作单元识别方法300。所述存储介质可以是计算机能够存取的任何可用介质或者是包含一个或多个可用介质集成的服务器、数据中心等数据存储设备。所述可用介质可以是磁性介质,(例如,软盘、硬盘、磁带)、光介质(例如,DVD)、或者半导体介质(例如固态硬盘Solid State Disk(SSD))等。

[0117] 综上所述,针对本发明所提出的技术方案,对无标注面部图像数据进行标注,得到不准确标签面部图像数据,然后将无标注标签、不准确标签和准确标签的数据送入两个并行的网络进行协同训练,这两个网络从两个不同的视角学习数据中不同的特征,能够明显提高模型的泛化性能。为了保证两个视角学习到的特征是相互独立的,通过正交损失对全连接网络的权重进行约束。为了保证两个视角判别的结果一致,通过一致性损失对来自两个视角的全连接网络的预测结果进行了约束。而对于每一个视角,首先采用一个主干网络提取特征,然后将每个视角提取的特征通过噪声网络和干净网络,这两个网络分别用来学习不准确标签和准确标签,噪声网络可以作为正则化项对干净网络进行约束,从而防止干净网络过拟合。在网络收敛后,网络预测结果的可信度相对较高,此时的噪声网络预测结果与不准确标签结合可以更好地抵消掉不准确标签中的噪声,因此本发明将两个视角噪声网络预测结果的均值与不准确标签例如通过线性加权的方式进行融合,然后用融合后的新的不准确标签对噪声网络进行更新,实现了进一步提高精度的效果。

[0118] 当然,本发明还可有其它多种实施例,在不背离本发明精神及其实质的情况下,熟悉本领域的技术人员当可根据本发明作出各种相应的改变和变形,但这些相应的改变和变形都应属于本发明所附的权利要求的保护范围。

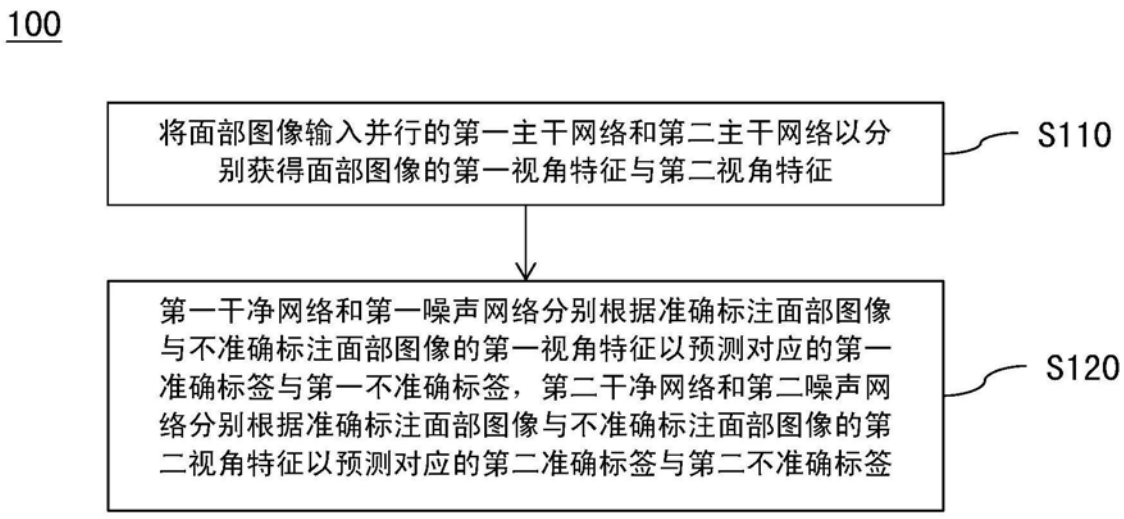

100
将面部图像输入并行的第一主干网络和第二主干网络以分别获得面部图像的第一视角特征与第二视角特征
S110
第一干净网络和第一噪声网络分别根据准确标注面部图像与不准确标注面部图像的第一视角特征以预测对应的第一准确标签与第一不准确标签，第二干净网络和第二噪声网络分别根据准确标注面部图像与不准确标注面部图像的第二视角特征以预测对应的第二准确标签与第二不准确标签
S120

图1A

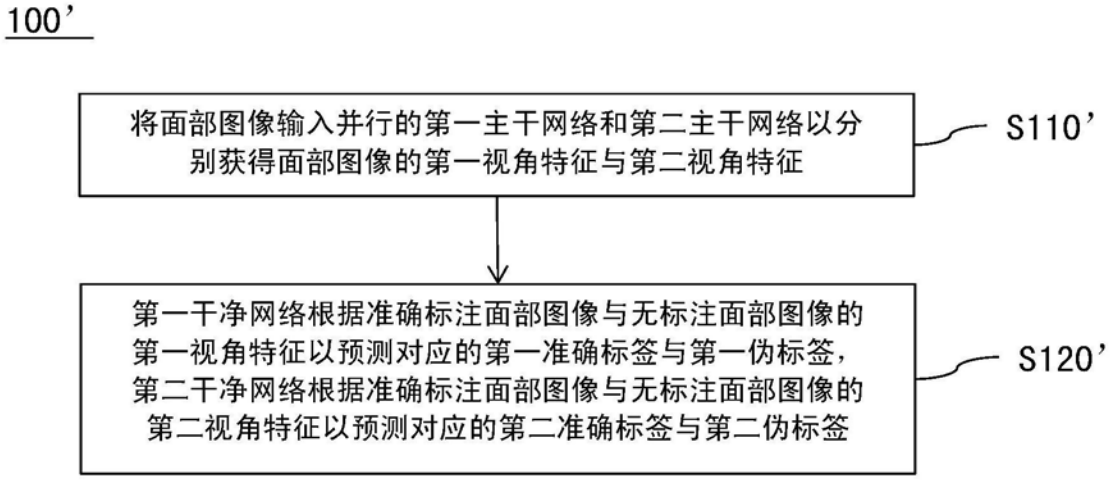

100'
将面部图像输入并行的第一主干网络和第二主干网络以分别获得面部图像的第一视角特征与第二视角特征
S110'
第一干净网络根据准确标注面部图像与无标注面部图像的第一视角特征以预测对应的第一准确标签与第一伪标签，第二干净网络根据准确标注面部图像与无标注面部图像的第二视角特征以预测对应的第二准确标签与第二伪标签
S120'

图1B

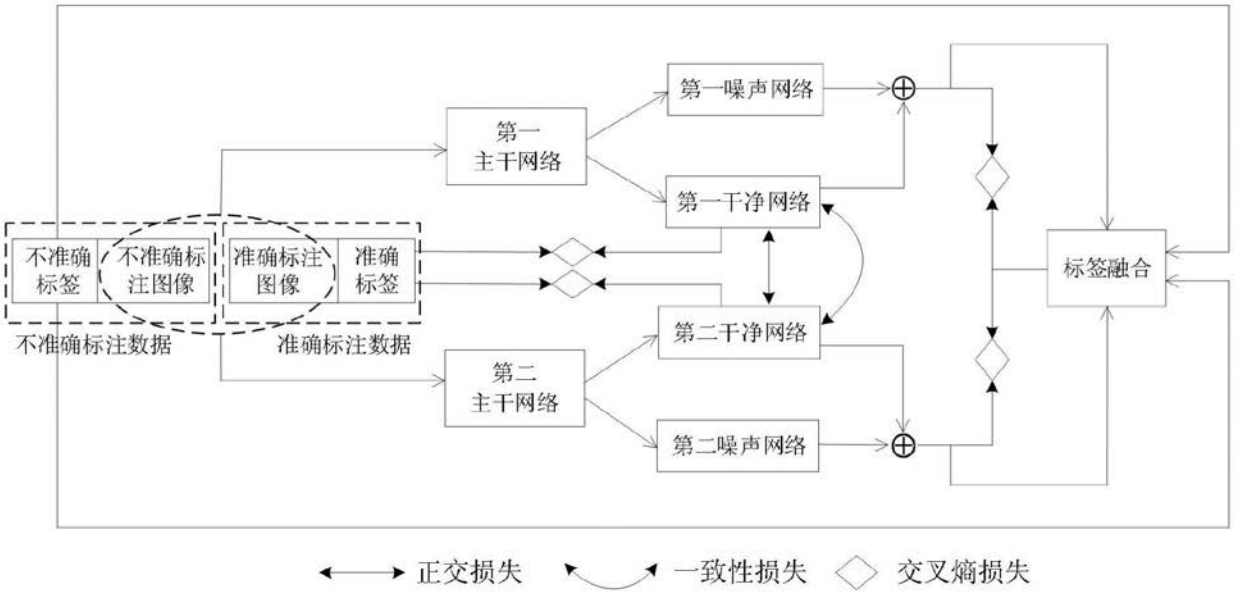

第一噪声网络
第一
主干网络
第一干净网络
不准确
标签
不准确标
注图像
准确标注
图像
准确
标签
标签融合
第二干净网络
第二
主干网络
第二噪声网络
不准确标注数据
准确标注数据
正交损失
一致性损失
交叉熵损失

图2A

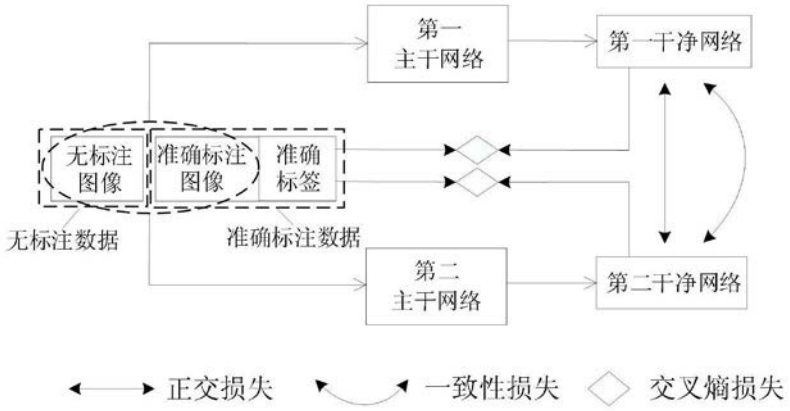

第一
主干网络
第一干净网络
无标注
图像
准确标注
图像
准确
标签
无标注数据
准确标注数据
第二
主干网络
第二干净网络
正交损失
一致性损失
交叉熵损失

图2B

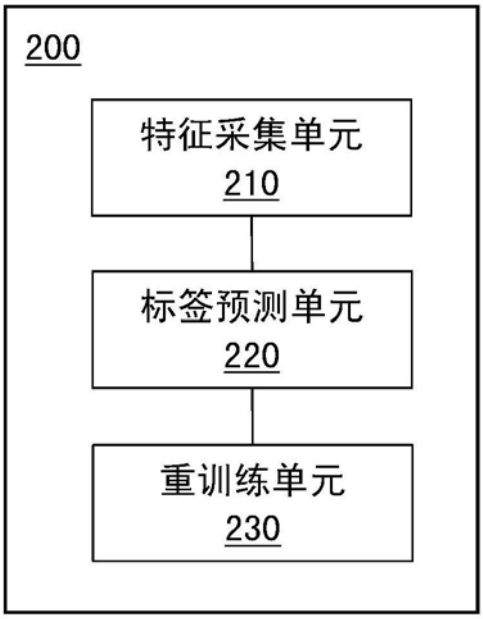

200
特征采集单元
210
标签预测单元
220
重训练单元
230

图3A

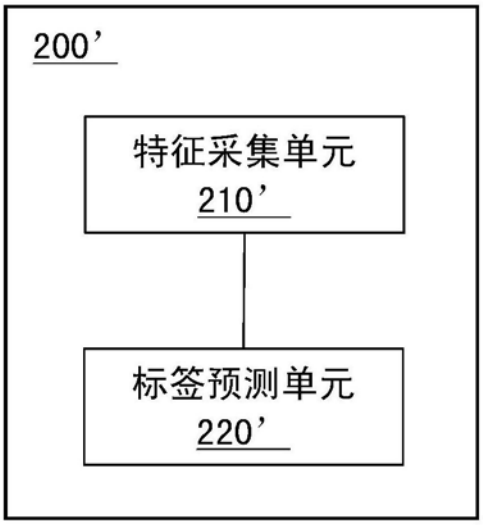

200'
特征采集单元
210'
标签预测单元
220'

图3B

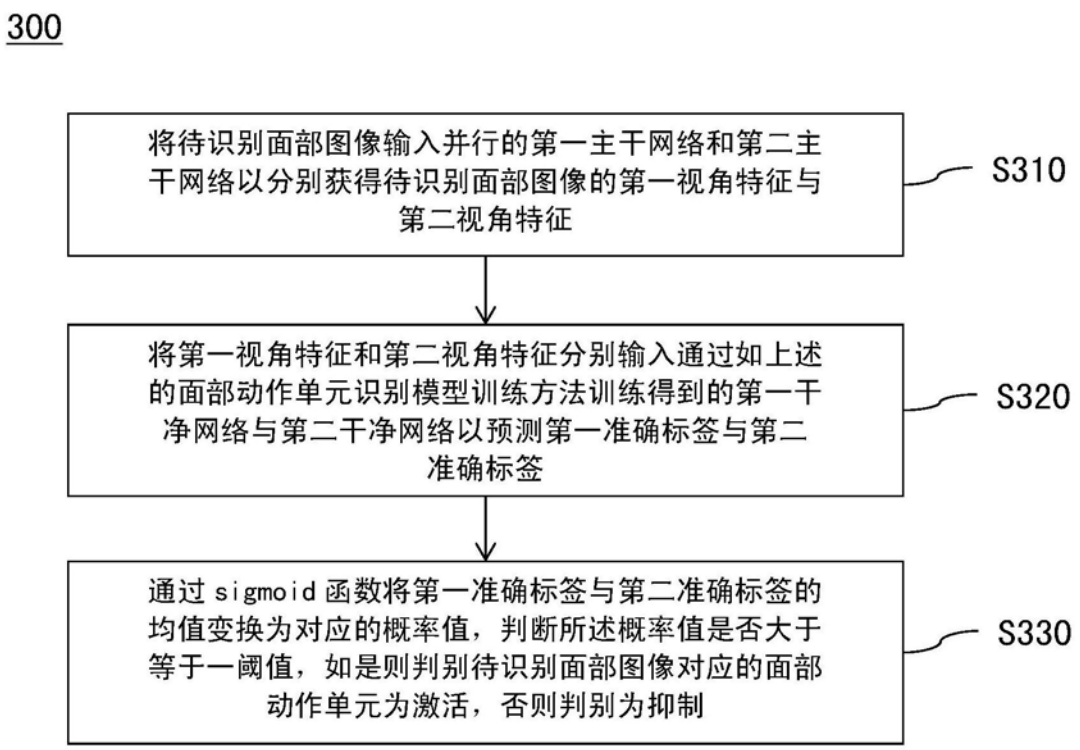

300
将待识别面部图像输入并行的第一主干网络和第二主干网络以分别获得待识别面部图像的第一视角特征与第二视角特征
S310
将第一视角特征和第二视角特征分别输入通过如上述的面部动作单元识别模型训练方法训练得到的第一干净网络与第二干净网络以预测第一准确标签与第二准确标签
S320
通过 sigmoid 函数将第一准确标签与第二准确标签的均值变换为对应的概率值，判断所述概率值是否大于等于一阈值，如是则判别待识别面部图像对应的面部动作单元为激活，否则判别为抑制
S330

图4

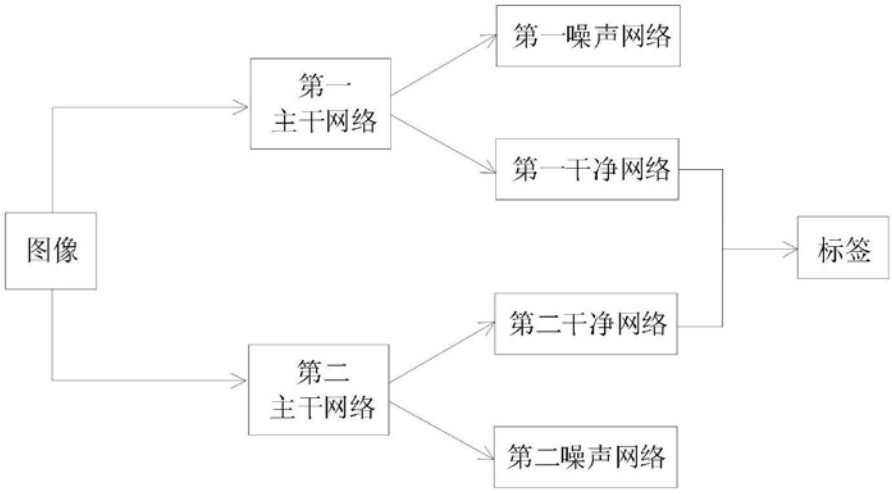

图像
第一
主干网络
第二
主干网络
第一噪声网络
第一干净网络
第二干净网络
第二噪声网络
标签

图5

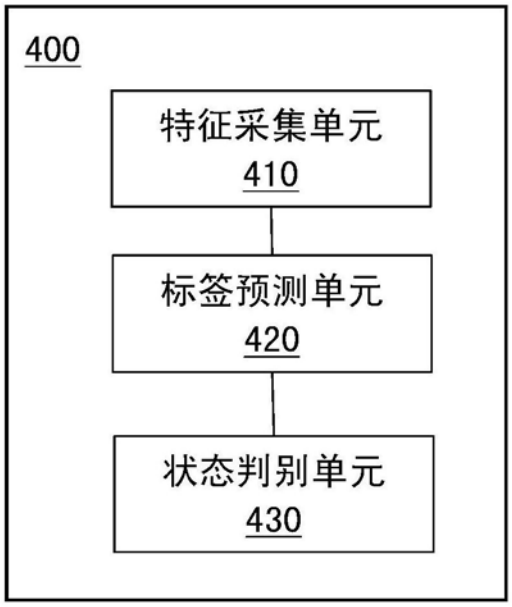


图6

图7