



(12)发明专利申请

(10)申请公布号 CN 106250840 A

(43)申请公布日 2016.12.21

(21)申请号 201610603175.0

(22)申请日 2016.07.27

(71)申请人 中国科学院自动化研究所

地址 100190 北京市海淀区中关村东路95号

(72)发明人 孙哲南 赫然 李海青 曹杰

(74)专利代理机构 中科专利商标代理有限责任公司 11021

代理人 钟文芳

(51) Int. Cl.

G06K 9/00(2006.01)

G06N 3/08(2006.01)

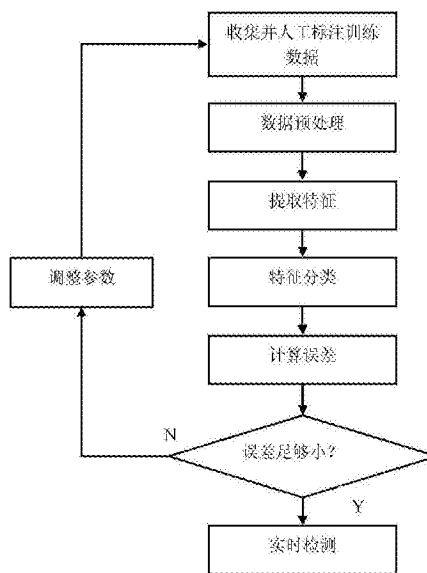
权利要求书1页 说明书6页 附图2页

(54)发明名称

基于深度学习的嘴巴张闭状态检测方法

(57)摘要

本发明公开了一种基于深度学习的嘴巴张闭状态检测方法,其主要包括数据预处理部分、特征提取部分、特征分类部分和误差计算部分。本发明最大的特点是充分利用深度卷积神经网络提取高层次特征的能力,提取出可以应对实际应用场景中经常存在的无规律噪声、较大光照变化以及通过遮挡嘴部关键部位进行恶意攻击等情况的鲁棒特征,并且利用全连接层来对提取到的特征分类,并通过计算误差并采用随机梯度下降法调整参数从而减小误差的方法来使检测方法自动习得分辨嘴巴张闭状态的能力。此外,本发明能够保证所需的计算资源和存储空间都不会因待检测图像分辨率的变化而产生大幅度波动。本发明操作方便,简单易用,精度更高、更加安全可靠。



1. 一种基于深度学习的嘴巴张闭状态检测方法,其特征在于,所述方法包括:
步骤S1,对输入的图像进行预处理;
步骤S2,对于预处理后的图像提取特征;
步骤S3,利用神经网络对于提取到的特征进行分类,通过对于分类结果进行归一化得到某一图像属于张嘴和闭嘴的预测结果;
步骤S4,计算神经网络输出结果的误差并将其反传,以调整神经网络的参数;
步骤S5,对输入的待检测图像进行测试,得到检测结果。
2. 根据权利要求1所述的方法,其特征在于,所述步骤S1包括以下步骤:
步骤S11,将输入的图像归一化;
步骤S12,将归一化后的图像转换为灰度图像。
3. 根据权利要求1所述的方法,其特征在于,所述步骤S2中,采用神经网络中的卷积层和最大池化层来进行特征的提取。

4. 根据权利要求3所述的方法,其特征在于,所述卷积层的数量为四个,最大池化层的数量为三个。

5. 根据权利要求3所述的方法,其特征在于,所述步骤S2中,卷积计算按照下式进行:

$$y^{j(r)} = |\tanh(b^{j(r)} + \sum_{i=1} w^{ij(r)} \times x^{i(r)})|$$

其中, x^i 代表第*i*层神经元的输入, y^j 代表第*j*层神经元的输出,上标*r*代表该层神经元的序号, w^{ij} 代表从第*i*层到第*j*层的权重矩阵, b^j 是第*j*层神经元的偏置项, \tanh 代表双曲正切函数。

6. 根据权利要求5所述的方法,其特征在于,所述步骤S2中,池化层按照下式进行操作:

$$y^{j(r)} = \max\left(0, b^{j(r)} + \sum_{i=1} w^{ij(r)} \times x^{i(r)}\right)$$

其中,max代表取大值。

7. 根据权利要求1所述的方法,其特征在于,所述步骤S3通过神经网络中的全连接层以及归一化操作来进行特征分类。

8. 根据权利要求1所述的方法,其特征在于,所述步骤S4中,误差表示为:

$$J = -\frac{1}{m} \sum_{i=1}^m y^i \log x^i + (1 - y^i) \log(1 - x^i)$$

其中,*m*为样本数量, x^i 、 y^i 分别代表第*i*个样本判断的张闭状态和第*i*个样本的实际张闭状态,且 $x \in \{0, 1\}$, $y \in \{0, 1\}$,1代表张嘴,0代表闭嘴。

9. 根据权利要求1所述的方法,其特征在于,所述步骤S5中,对于输入的待检测图像先重复所述步骤S1、S2、S3,当输出的张嘴概率大于闭嘴概率时便认为预测结果为张嘴,反之则认为是闭嘴。

10. 根据权利要求1所述的方法,其特征在于,所述步骤S5还包括根据对检测结果和输入图像人工标注标签的比较得到检测准确率的步骤。

基于深度学习的嘴巴张闭状态检测方法

技术领域

[0001] 本发明属于图像处理、模式识别和软件工程等技术领域,尤其涉及一种基于深度学习的嘴巴张闭状态检测方法,并且对现实中常见的存在光照变化干扰和人脸关键部位部分遮挡的场景具有很强的适应性。

背景技术

[0002] 人脸是一个人最重要的外貌特征,人脸识别技术在身份认证与安全防护等当前热门应用领域具有广泛的应用,具有非常广阔的前景。然而人脸极易用照片、视频等方式进行复制,因此对合法用户人脸的假冒是人脸识别与认证系统安全的重要威胁。因此现有的人脸识别系统中一个必不可少的环节就是活体检测部分,用于判断提交的生物特征是否来自有生命的个体。当前高精度的活体检测系统往往采取交互式的检测手段,例如,要求用户按照指示张、闭嘴巴。目前主流的检测算法的实施步骤先是从输入的图片中提取人工设计的特征,再对提取到的特征做分类从而判断输入是否来自活体。

[0003] 为了保证检测的准确率,人脸识别系统需要对现实生活场景中存在的各种干扰都具有较强的鲁棒性,然而在现有阶段由于检测算法的不成熟,活体检测系统对具有较大光照变化、姿态变化、部分遮挡以及随机噪声的场景的鲁棒性仍然需要进一步提高才能满足实际的应用。

[0004] 由于传统的活体检测方法不能满足实际应用中对高精度实时活体检测的需求,为此本发明提出了一种可以应用于交互式活体检测的高精度嘴巴张闭状态检测方法。

发明内容

[0005] 本发明的目的是准确地对人嘴部图片的张闭状态进行检测,进一步增强现有算法对光照变化和部分遮挡场景的鲁棒性。

[0006] 为实现人嘴部图片的张闭状态进行高精度检测,本发明提供如下解决方案:通过神经网络的卷积层实现鲁棒特征提取并通过全连接层来实现特征分类,通过随机梯度下降法来根据误差对卷积层和全连接层的参数进行调整,使得对嘴巴张闭状态检测的准确度更高,而且不依赖人工设计的特征提取器。

[0007] 目前的高精度活体检测方法采用的是交互式活体方法,传统的方法需要经验丰富的专业人士根据具体的活体检测任务来设计特定的特征提取方法。虽然在过去的相当长时间内,传统方法在干扰较小的环境中的活体检测取得了较好的效果,但是面对实际应用中存在大量光照变化、姿态变化以及各种无规律噪声的环境,始终不能达到令人满意的泛化性能。考虑到实际存在的噪声是极其丰富的,本发明提出了使用深度卷积神经网络层来提取传统方法难以取得的具有很强鲁棒性的高层次的特征。

[0008] 本发明基于深度学习的嘴巴张闭状态检测方法包括以下步骤:

[0009] 步骤S1,对输入的图像进行预处理;

[0010] 步骤S2,对于预处理后的图像提取特征;

[0011] 步骤S3,利用神经网络对于提取到的特征进行分类,通过对于分类结果进行归一化得到某一图像属于张嘴和闭嘴的预测结果;

[0012] 步骤S4,计算神经网络输出结果的误差并将其反传,以调整神经网络的参数;

[0013] 步骤S5,对输入的待检测图像进行测试,得到检测结果。

[0014] 可选地,所述步骤S1包括以下步骤:

[0015] 步骤S11,将输入的图像归一化;

[0016] 步骤S12,将归一化后的图像转换为灰度图像。

[0017] 可选地,所述步骤S2中,采用神经系统中的卷积层和最大池化层来进行特征的提取。

[0018] 可选地,所述卷积层的数量为四个,最大池化层的数量为三个。

[0019] 可选地,所述步骤S2中,卷积计算按照下式进行:

$$[0020] \quad y^{j(r)} = |\tanh(b^{j(r)} + \sum_{i=1}^m w^{ij(r)} \times x^{i(r)})|$$

[0021] 其中, x^i 代表第*i*层神经元的输入, y^j 代表第*j*层神经元的输出,上标*r*代表该层神经元的序数, w^{ij} 代表从第*i*层到第*j*层的权重矩阵, b^j 是第*j*层神经元的偏置项,tanh代表双曲正切函数。

[0022] 可选地,所述步骤S2中,池化层按照下式进行操作:

$$[0023] \quad y^{j(r)} = \max\left(0, b^{j(r)} + \sum_{i=1}^m w^{ij(r)} \times x^{i(r)}\right)$$

[0024] 其中,max代表取大值。

[0025] 可选地,所述步骤S3通过神经网络中的全连接层以及归一化操作来进行特征分类。

[0026] 可选地,所述步骤S4中,误差表示为:

$$[0027] \quad J = -\frac{1}{m} \sum_{i=1}^m y^i \log x^i + (1 - y^i) \log(1 - x^i)$$

[0028] 其中,*m*为样本数量, x^i 、 y^i 分别代表第*i*个样本判断的张闭状态和第*i*个样本的实际张闭状态,且 $x \in \{0, 1\}$, $y \in \{0, 1\}$,1代表张嘴,0代表闭嘴。

[0029] 可选地,所述步骤S5中,对于输入的待检测图像先重复所述步骤S1、S2、S3,当输出的张嘴概率大于闭嘴概率时便认为预测结果为张嘴,反之则认为闭嘴。

[0030] 可选地,所述步骤S5还包括根据对检测结果和输入图像人工标注标签的比较得到检测准确率的步骤。

[0031] 与现有技术相比,本的有益效果是:在实际应用场景中的较大光照变化以及关键部位部分遮挡的情况下的检测仍具有较强的鲁棒性,可以更好地应用于充满噪声的活体检测系统中。本发明充分利用了深度卷积神经网络提取高层次特征的能力,并根据全连接层对提取到的特征进行分类,并且在一个独立的系统中实现了图像输入、特征提取、特征分类以及结果预测四个功能。神经网络模型可以根据误差计算自动地对网络中的参数进行调整来达到理想的预测结果,调整参数时除了人工标注每一张输入嘴部图像的张闭状态无需其他的监督方法,简单易行,且避免了传统的人工设计特征提取方法由于特征设计不合理等原因导致的系统准确率降低或者泛化性能不足等情况。

附图说明

[0032] 图1是根据本发明一实施例的基于深度学习的嘴巴张闭状态检测方法的流程图；

[0033] 图2为本发明采用的神经网络结构示意图。

具体实施方式

[0034] 为使本发明的目的、技术方案和优点更加清楚明白，以下结合具体实施例，并参照附图，对本发明进一步详细说明。

[0035] 本发明提供一种技术方案：一种高精度嘴巴张闭状态检测方法，目前实际应用的活体检测系统中，如何提高在充满噪声以及存在恶意攻击的场景下的检测准确率是一个急需解决的问题。本发明针对此种问题采用了卷积神经网络层来提取特征，同时使用全连接层和softmax函数来进行特征分类，采用神经网络的反向传播和随机梯度下降法来对网络参数进行调整，使提取到的特征更加鲁棒，对噪声的适应性更强。另外，本发明由于对输入图像进行归一化的预处理方法，可以使系统的运算开销和内存占用都保持在一个合理范围内。

[0036] 图1是根据本发明一实施例的基于深度学习的嘴巴张闭状态检测方法的流程图，如图1所示，所述基于深度学习的嘴巴张闭状态检测方法包括以下几个步骤：

[0037] 步骤S1，对输入的图像进行预处理；

[0038] 本方法要求输入的图像是人嘴部的对齐图像，并且人工标记嘴巴的张闭状态作为监督信息，在输入训练集图像时要求样本具有一定的代表性，且数量尽量不小于四到五千张。实际应用中输入的图像可能是各种大小不同类型的图像，因此该步骤首先使用线性插值法将输入的图像归一化到大小为 60×60 像素图像，然后将归一化后的输入图像转换为灰度图像，即只使用图像的灰度特征。归一化输入图像的规格可以减少噪声，还可以使得后续的训练以及检测过程所需要的计算资源以及硬盘存储资源都维持在一个相对稳定的范围，既可以避免输入图像分辨率过高时产生大量不必要的运算开销也可以在输入图像分辨率过低时适当地放大图像以便尽可能地提升检测精度。

[0039] 步骤S2，对于预处理后的图像提取特征；

[0040] 嘴巴张闭状态检测在实际应用中会面临各种各样的干扰，另外考虑到在极端的情况下某些不法分子可能会对检测系统进行恶意攻击，现有的活体检测系统都需要足够高的精度才能够满足实际应用的需要。为了解决此问题，在本发明一实施例中，该步骤采用神经网络中的卷积层和最大池化层来进行鲁棒特征的提取，相比较以往的方法可以更有效地应对输入图像的光照变化以及嘴部关键部位部分遮挡的情况。在该实施例中，利用四个卷积层和三个最大池化层进行特征提取：对于预处理后的图像，先由20个尺寸为 $5 \times 5 \times 1$ 的步长为1的卷积核进行卷积计算并经过最大池化层降采样，再由40个尺寸为 $3 \times 3 \times 20$ 的卷积核进行卷积计算并经过最大池化层降采样，第三次卷积计算的卷积核尺寸为 $2 \times 2 \times 40$ ，数量为60，经过第三个最大池化层后进行第四次卷积计算，卷积核的尺寸为 $2 \times 2 \times 60$ ，数量为80，最大池化层的卷积核的尺寸均为 2×2 。由于深层神经网络模型的强大学习能力，该步骤可以提取出能非常有效地分辨嘴巴张闭状态的高层次特征用于后续分类任务。

[0041] 在所述步骤S2中，卷积计算按照下式进行：

$$[0042] \quad y^{j(r)} = |\tanh(b^{j(r)} + \sum_{i=1} w^{ij(r)} \times x^{i(r)})|$$

[0043] 其中, x^i 代表第 i 层神经元的输入, y^j 代表第 j 层神经元的输出, 上标 r 代表该层神经元的序数, w^{ij} 代表从第 i 层到第 j 层的权重矩阵, b^j 是第 j 层神经元的偏置项, \tanh 代表双曲正切函数。

[0044] 池化层按照下式进行操作:

$$[0045] \quad y^{j(r)} = \max\left(0, b^{j(r)} + \sum_{i=1} w^{ij(r)} \times x^{i(r)}\right)$$

[0046] 其中, x^i 、 y^j 、 w^{ij} 、 b^j 代表的含义和上述卷积计算中的相同, \max 代表取大值。

[0047] 步骤 S3, 利用神经网络对于提取到的特征进行分类, 通过对于分类结果进行归一化得到某一图像属于张嘴和闭嘴的预测结果;

[0048] 该步骤的对于提取到的特征通过神经网络中的全连接层以及归一化来实现特征分类, 进一步输出嘴部图像属于张嘴和闭嘴的概率。具体地, 该步骤对于提取到的特征进行二分类, 即把根据输入图像的特征判断其为张嘴或者闭嘴。该步骤采用两个相连的神经网络中经典的全连接层来实现特征分类, 输入为所述步骤 S2 提取到的特征序列, 输出为嘴部图像属于张嘴和闭嘴的概率, 其中, 第一个全连接层由 120 个神经元构成, 第二个全连接层由 2 个神经元构成, 输入的特征序列经过两个全连接层后再经过归一化, 就得到了该输入图像属于张嘴和闭嘴的概率。

[0049] 3. 事实上全连接层也可是看作是卷积层采用尺寸为 1×1 的卷积核并且步长为 1 的一种特殊情况, 所以全连接层的计算方法也可以用上述的卷积计算式来表达。全连接层的计算复杂度相比于同样规模的卷积层要复杂很多, 而且在实际应用中如果直接使用全连接层往往会产生严重的过拟合现象。所以本发明采用了先用卷积层从输入的图像中提取不随输入图片光照、遮挡等干扰情况而变化, 即可以应对较大光照变化、关键位置遮挡情况的鲁棒特征, 然后再使用全连接层做分类。然后对第二个全连接层的两个神经元的输出进行归一化, 即把它们的输出等比例缩放, 使它们的和等于 1, 此时的输出即为图片属于张嘴和闭嘴的概率。其中, 用于提取特征的神经网络中的卷积层的参数可以自动根据其误差进行优化, 不需要更多的人工修正方法。

[0050] 图 2 为本发明采用的神经网络结构示意图。

[0051] 步骤 S4, 计算神经网络输出结果的误差并将其反传, 以调整神经网络的参数;

[0052] 该步骤中各个神经网络层的参数由均值为 0, 方差为 0.01 的高斯函数进行随机初始化, 然后根据神经网络输出结果的误差, 通过随机梯度下降法来调整神经网络的参数, 经过多次的迭代来不断地降低误差, 直到误差下降到某个预先设定的值。经过该步骤后, 神经网络学习得到了判断嘴巴张闭状态的方法, 以便用于后续检测。

[0053] 所述步骤 S4 具体为:

[0054] 本发明处理的问题是一个二分类问题, 采用下式来表示检测方法预测结果的误差:

$$[0055] \quad J = -\frac{1}{m} \sum_{i=1}^m y^i \log x^i + (1 - y^i) \log(1 - x^i)$$

[0056] 其中, m 为样本数量, x^i 、 y^i 分别代表第 i 个样本判断的张闭状态和第 i 个样本的实际张闭状态, 且 $x \in \{0, 1\}$, $y \in \{0, 1\}$, 1 代表张嘴, 0 代表闭嘴。

[0057] 该步骤中神经网络的权重采用批次样本数量为 128 个的随机梯度下降法来更新, 权重衰减系数设为 0.005, 动量系数设为 0.9。具体权重更新表达式如下所示:

$$[0058] \quad v_{i+1} = 0.9 \cdot v_i - 0.0005 \cdot \varepsilon \cdot w_i - \varepsilon \cdot \left\langle \frac{\partial L}{\partial w} \Big|_{w_i} \right\rangle_{D_i}$$

$$[0059] \quad w_{i+1} = w_j + v_{i+1}$$

[0060] 其中 i 代表迭代的次数, v 代表权重的增量, ε 代表学习率, $\left\langle \frac{\partial L}{\partial w} \Big|_{w_i} \right\rangle_{D_i}$ 代表第 i 批样本中目标函数对权重 w 的偏导数在 w_i 处的值, 本发明采用处理大数据样本常用的批次训练法, D_i 代表训练时的第 i 个批次的训练数据。

[0061] 本方法中初始学习率设置为 0.05 且按照下式更新:

$$[0062] \quad \varepsilon_i = \varepsilon_0 \cdot (1 + \gamma)^{-p}$$

[0063] ε_0 和 ε_i 分别表示初始学习率和迭代的到第 i 次的学习率, 参数 $\gamma = 0.0001$ 且参数 $p = 0.75$ 。

[0064] 步骤 S5, 对输入的待检测图像进行测试, 得到检测结果。

[0065] 在根据训练样本学习得到理想的网络参数后, 就可以进行高精度的嘴巴张闭状态检测了, 输入待检测嘴部图片后可自动预测嘴巴张闭状态。在本步骤中, 对于输入的待检测图像先重复步骤 S1、S2、S3, 当输出的张嘴概率大于闭嘴概率时便认为预测结果为张嘴, 反之则认为是闭嘴。

[0066] 如果有人工标记的输入实际状态, 那么所述步骤 S5 还可以根据得到的检测结果统计检测准确率, 从而评估本方法在各种应用场景下的表现。

[0067] 具体地, 根据对检测结果和输入图片人工标注的标签的比较可以得到检测的准确率, 通过准确率可以评估系统在各种环境下的表现。

[0068] 所述检测准确率用下式表示:

$$[0069] \quad \text{准确率} = \frac{TP + TN}{N} \times 100\%$$

[0070] 其中, N 代表待检测图像的总数, TP 代表被正确检测的张嘴图像, TN 代表被正确检测的闭嘴图像。在具有较大光照变化以及关键部件遮挡的环境下进行检测, 较高的准确率可以表明本发明对这些干扰环境的鲁棒性。

[0071] 综上, 本发明公开了一种基于深度学习的嘴巴张闭状态检测方法。本发明最大的特点是充分利用深度卷积神经网络提取高层次特征的能力, 提取出可以应对实际应用场景中经常存在的无规律噪声、较大光照变化以及通过遮挡嘴部关键部位进行恶意攻击等情况的鲁棒特征, 并且利用全连接层来对提取到的特征分类, 并通过计算误差并采用随机梯度下降法调整参数从而减小误差的方法来使检测方法自动习得分辨嘴巴张闭状态的能力。此外, 本发明通过将输入的图像通过预处理归一化为固定大小灰度图像, 保证所需的计算资源和存储空间都不会因待检测图像分辨率的变化而产生大幅度波动。本发明在一个独立的系统中实现了图像输入、特征提取、特征分类以及结果预测四个功能, 输入待检测图像, 系统经过计算即可直接输出嘴巴的张闭状态。软件操作方便, 简单易用。本发明在训练阶段只

需要人工标注足量的嘴部图片的张闭状态作为监督信息即可自动对参数进行调整,无需其他人工干预手段,与以往的检测手段相比,精度更高、更加安全可靠。

[0072] 本发明的应用场景举例如下:

[0073] 应用场景1:基于深度学习的嘴巴张闭状态检测方法在应用软件刷脸登陆活体检测环节中的应用。

[0074] 本发明可广泛应用于刷脸认证登陆系统中的活体检测环节。假设某应用程序的手机移动客户端支持刷脸登陆,在进行身份认证之前需要进行活体检测,要求登陆客户按照要求在听到提示音后张嘴,只有系统检测到用户做出了合格的动作才会把输入的图片进行后续的身份认证等步骤,否则系统会重复活体检测的需求直到通过为止。如果有不法分子通过盗取了合法用户的照片、视频等认证信息,或者伪造合成照片的方法,企图在手机移动客户端前用假照片或者回放视频来假冒合法用户登陆,本发明中嘴巴张闭状态检测方法可以发现登陆的“用户”不能按照要求完成张闭嘴的动作,就会拒绝登陆要求并记录登陆失败次数,如果登陆失败次数超过一定数量,还可以配合其他的安全防护系统通过冻结账户、要求输入手机验证码等方式来保证账户的安全。

[0075] 以上所述的具体实施例,对本发明的目的、技术方案和有益效果进行了进一步详细说明,所应理解的是,以上所述仅为本发明的具体实施例而已,并不用于限制本发明,凡在本发明的精神和原则之内,所做的任何修改、等同替换、改进等,均应包含在本发明的保护范围之内。

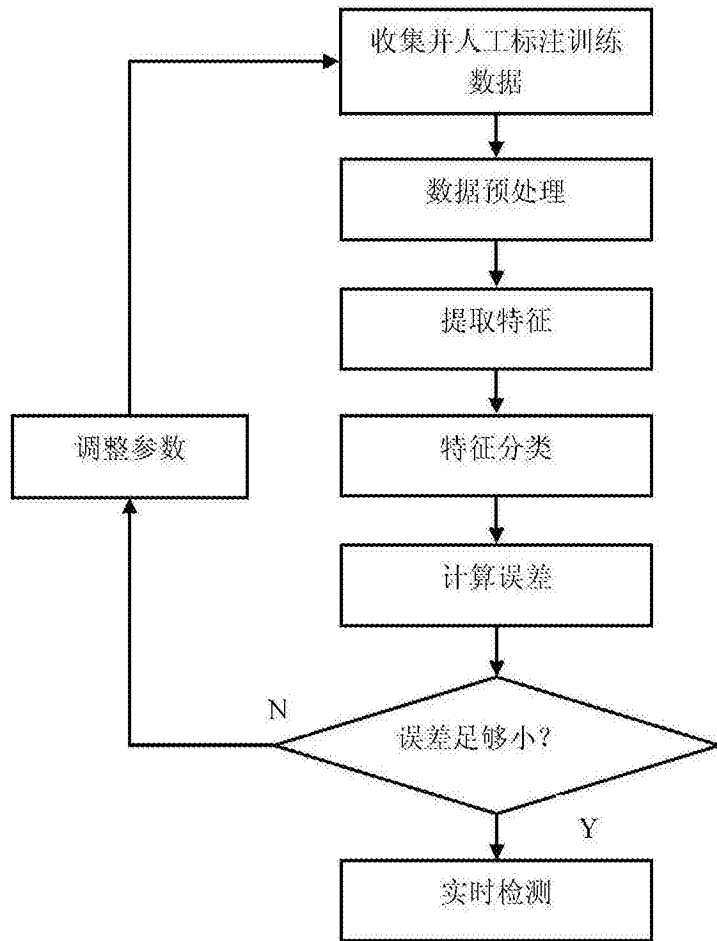


图1

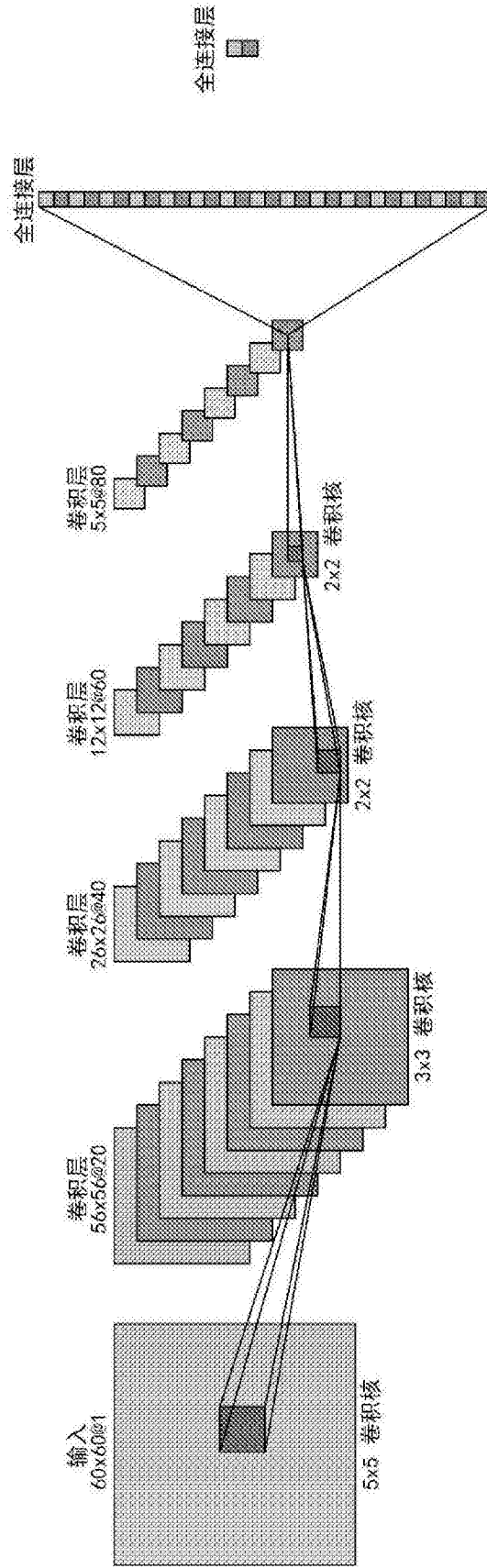


图2