



(12)发明专利申请

(10)申请公布号 CN 107169446 A

(43)申请公布日 2017.09.15

(21)申请号 201710334543.0

(22)申请日 2017.05.12

(71)申请人 深圳市唯特视科技有限公司

地址 518057 广东省深圳市高新技术产业园区高新南一道009号中科研发园新产业孵化中心楼610室

(72)发明人 夏春秋

(51)Int.Cl.

G06K 9/00(2006.01)

G06N 3/08(2006.01)

G06K 9/40(2006.01)

G06N 99/00(2010.01)

权利要求书2页 说明书5页 附图1页

(54)发明名称

一种基于注意定势度量学习的视频面部识别方法

(57)摘要

本发明中提出的一种基于注意定势度量学习的视频面部识别方法，其主要内容包括：注意定势的度量学习(ASML)、记忆注意权重、将ASML自然融入到卷积神经网络(CNN)中，其过程为，先定义图像集上的有效距离度量，显著地最小化集合内距离，并同时最大化集中距离，再把权重作为神经图灵机，其中人脸特征集作为记忆，权重作为地址读写内存，最后将ASML自然融入到卷积神经网络中，从而形成端到端的学习方案。本发明缩小了相同集合的概率分布之间的差距，同时扩大了不同集合之间的差距；减少了视频或图像集中的样本偏差和噪声，有效地利用视频或图像集中的信息，从而提高了识别的性能。

注意定势的度量学习(ASML)

记忆注意权重

将ASML自然融入到卷积神经网络(CNN)中

1. 一种基于注意定势度量学习的视频面部识别方法,其特征在于,主要包括注意定势的度量学习(ASML)(一);记忆注意权重(二);将ASML自然融入到卷积神经网络(CNN)中(三)。

2. 基于权利要求书1所述的注意定势的度量学习(ASML)(一),其特征在于,它纠正了样本偏差,并测量面部图像组之间的相关性,缩小了相同集合的概率分布之间的差距,同时扩大了不同集合之间的差距;与ASML组合的端对端可训练的深卷积神经网络(CNN),可以学习更多的歧视性深层次表示,获得可靠的人脸识别;ASML是最大平均偏差与记忆注意权重的广义扩展。

3. 基于权利要求书1所述的最大平均偏差(MMD),其特征在于,设 \mathcal{F} 是一类函数 $f: x \rightarrow R$,同上定义 $p, q, X, Y; MMD$ 及其经验估计为:

$$MMD[\mathcal{F}, p, q] := \sup_{f \in \mathcal{F}} (E_{x \sim p}[f(x)] - E_{y \sim q}[f(y)]) \quad (1)$$

$$MMD[\mathcal{F}, X, Y] := \sup_{f \in \mathcal{F}} \left(\frac{1}{n} \sum_{i=1}^n f(x_i) - \frac{1}{m} \sum_{i=1}^m f(y_i) \right) \quad (2)$$

当 \mathcal{F} 是紧凑度量空间 x 上定义的一个单位球,方程 $MMD[\mathcal{F}, X, Y] = 0$ 成立,当且仅当 $p = q$;直观地说,MMD越小,X和Y之间的分布越相关;因此,MMD越大,分布之间的差异越大,特征完全不同的。

4. 基于权利要求书3所述的基于集合的度量学习(MSML),其特征在于,根据最大平均偏差,基于集合的度量学习(MSML)表示为:

$$MSML(X, Y, Z) = \|E_{x \sim p}[f(x)] - E_{y \sim q}[f(y)]\|_2 + [\alpha - \|E_{x \sim p}[f(x)] - E_{z \sim q}[f(z)]\|_2]_+ \quad (3)$$

其中, $[\cdot]_+$ 表示最大值, $(\cdot, 0)$ 和 α 是一个恒定的余量; X, Y, Z 表示图像集;其中, X 和 Y 来自同一类,但又不同于 Z ;减小MSML的数量,可以增加相同与不同组之间的相关性和差异分别。

5. 基于权利要求书3所述的不考虑样本偏差和异常值的集合,其特征在于,最大平均偏差不考虑样本偏差和异常值的集合:

$$\omega(x) = p(x)/(\hat{p}(x)) :$$

$$\underset{\omega(x) \geq 0}{\text{minimize}} \|E_p[f(x)] - E_{\hat{p}}[\omega(x)f(x)]\|_2 \quad s.t. E_{\hat{p}}[\omega(x)] = 1 \quad (4)$$

其中, \hat{p} 是一个分布,其对应于 p ; $\omega(x)$ 可以纠正样本偏差和消除异常值(或减少噪音)。

6. 基于权利要求书5所述的整流平均差异(RMD),其特征在于, $\omega(x)$ 作为样本偏差为MMD修正项,整流平均差异(RMD)如下:

$$RMD(X, Y) = \|E_{\omega(x)}[f(x)] - E_{\omega(x)}[f(y)]\|_2$$

$$\left\| \sum_{i=1}^n \omega(x_i) f(x_i) - \sum_{j=1}^m \omega(y_j) f(y_j) \right\|_2 \quad s.t. \sum_{i=1}^n \omega(x_i) = 1, \sum_{j=1}^m \omega(y_j) = 1 \quad (5)$$

结合公式(3)和公式(5),我们联结基于整流项的度量学习的平均集,并获得高级的度量学习方法,命名为注意定势度量学习的视频人脸识别:

$$ASML(X, Y, Z) = \|E_{\omega(x)}[f(x)] - E_{\omega(x)}[f(y)]\|_2 + [\alpha - \|E_{\omega(x)}[f(x)] - E_{\omega(x)}[f(z)]\|_2]_+$$

$$\text{s.t. } \sum \omega(x) = 1, \sum \omega(y) = 1, \sum \omega(z) = 1 \quad (6)$$

如上式所示。

7. 基于权利要求书1所述的记忆注意权重(二),其特征在于,权重应满足以下三点:

(1) 加权方法容易集成在CNN的框架,其参数是端到端可训练的监督方式;

(2) 权重基于全局内容和设置感知,因为构建了一个训练实例作为特殊的图像,能更好地利用信息;

(3) 学习权重图像顺序必须不变,每个图像的权重和图像在训练实例中是无序。

8. 基于权利要求书7所述的权重,其特征在于,把权重作为神经图灵机(NTM),其中人脸特征集作为记忆,权重视为地址读写内存;NTM涉及三个基本组件:读取向量,擦除向量和添加向量;令 $\{f_i\}$ 成为一个面部特征集,在 s_i 上应用softmax操作以形成归一化权重 ω_i ;公式(6)中操作和重新配置的 $E_\omega[f(x)]$ 如下:

$$\omega_i = \frac{\exp(s_i)}{\sum_j \exp(s_j)} \quad (7)$$

$$E_\omega[f(x)] = \sum_i \left[\frac{\exp(s_i)}{\sum_j \exp(s_j)} \cdot f_i \right] \quad (8)$$

权重 $\{\omega_i\}$ 是无序的, $\{s_i\}$ 是可微的。

9. 基于权利要求书1所述的将ASML自然融入到卷积神经网络(CNN)中(三),其特征在于,CNN包含29层残块卷积和最大特征映射操作;基于网络,ASML和记忆注意权重耦合,实现视频人脸识别;Softmax是重要的监督信号方法,得到以下目标函数:

$$L = \lambda_1 \text{Softmax} + \lambda_2 \text{ASML} \quad (9)$$

其中, λ_1 和 λ_2 是这两个项之间的权重;Softmax功能用于标准的人脸识别任务,ASML惩罚项增加了相关性、类似的分布和不同脸部集合之间的差异。

10. 基于权利要求书9所述的卷积神经网络(CNN),其特征在于,网络中的所有参数是可微的,梯度可以计算:

$$\frac{\partial L}{\partial \theta} = \lambda_1 \frac{\partial \text{Softmax}}{\partial \theta} + \lambda_2 \frac{\partial \text{ASML}}{\partial \theta} \quad (10)$$

$$\frac{\partial L}{\partial s} = \lambda_2 \frac{\partial \text{ASML}}{\partial \omega} \cdot \frac{\partial \omega}{\partial s} \quad (11)$$

其中, θ 代表网络参数。

一种基于注意定势度量学习的视频面部识别方法

技术领域

[0001] 本发明涉及面部识别领域,尤其是涉及了一种基于注意定势度量学习的视频面部识别方法。

背景技术

[0002] 由于视频监控正在快速普及,众多的视频监控应用迫切需要一种远距离、用户非配合状态下的快速身份识别技术,以求远距离快速确认人员身份,实现智能预警。人脸识别技术可以从监控视频图像中实时查找人脸,并与人脸数据库进行实时比对,从而实现快速身份识别。因此,视频人脸识别广泛应用在协助公安刑侦破案、门禁系统、摄像监视系统、身份辨识和支付系统等。然而,由于照明变化、分辨率低、姿态变化和运动产生的模糊等,给视频脸部识别技术的研究带来了一定的难度。

[0003] 本发明提出了一种基于注意定势度量学习的视频面部识别方法,先定义图像集上的有效距离度量,显著地最小化集合内距离,并同时最大化集中距离,再把权重作为神经图灵机,其中人脸特征集作为记忆,权重作为地址读写内存,最后将注意定势的度量学习(ASML)自然融入到卷积神经网络中,从而形成端到端的学习方案。本发明缩小了相同集合的概率分布之间的差距,同时扩大了不同集合之间的差距;减少了视频或图像集中的样本偏差和噪声,有效地利用视频或图像集中的信息,从而提高了识别的性能。

发明内容

[0004] 针对由于照明变化、分辨率低、姿态变化和运动产生的模糊等问题,本发明的目的在于提供一种基于注意定势度量学习的视频面部识别方法,先定义图像集上的有效距离度量,显著地最小化集合内距离,并同时最大化集中距离,再把权重作为神经图灵机,其中人脸特征集作为记忆,权重作为地址读写内存,最后将ASML自然融入到卷积神经网络中,从而形成端到端的学习方案。

[0005] 为解决上述问题,本发明提供一种基于注意定势度量学习的视频面部识别方法,其主要内容包括:

[0006] (一) 注意定势的度量学习(ASML);

[0007] (二) 记忆注意权重;

[0008] (三) 将ASML自然融入到卷积神经网络(CNN)中。

[0009] 其中,所述的注意定势的度量学习(ASML),它纠正了样本偏差,并测量面部图像组之间的相关性,缩小了相同集合的概率分布之间的差距,同时扩大了不同集合之间的差距;与ASML组合的端对端可训练的深卷积神经网络(CNN),可以学习更多的歧视性深层次表示,获得可靠的人脸识别;ASML是最大平均偏差与记忆注意权重的广义扩展。

[0010] 进一步地,所述的最大平均偏差(MMD),设 \mathcal{F} 是一类函数 $f: \mathcal{X} \rightarrow \mathbb{R}$,同上定义 $p, q, X, Y; MMD$ 及其经验估计为:

$$[0011] \quad \text{MMD}[\mathcal{F}, p, q] := \sup_{f \in \mathcal{F}} (E_{x \sim p}[f(x)] - E_{y \sim q}[f(y)]) \quad (1)$$

$$[0012] \quad \text{MMD}[\mathcal{F}, X, Y] := \sup_{f \in \mathcal{F}} \left(\frac{1}{n} \sum_{i=1}^n f(x_i) - \frac{1}{m} \sum_{i=1}^m f(y_i) \right) \quad (2)$$

[0013] 当 \mathcal{F} 是紧湊度量空间 \mathcal{X} 上定义的一个单位球,方程 $\text{MMD}[\mathcal{F}, X, Y] = 0$ 成立,当且仅当 $p=q$;直观地说,MMD越小,X和Y之间的分布越相关;因此,MMD越大,分布之间的差异越大,特征完全不同的。

[0014] 进一步地,所述的基于集合的度量学习(MSML),根据最大平均偏差,基于集合的度量学习(MSML)表示为:

$$[0015] \quad \text{MSML}(X, Y, Z) = \|E_{x \sim p}[f(x)] - E_{y \sim q}[f(y)]\|_2 + [\alpha - \|E_{x \sim p}[f(x)] - E_{z \sim q}[f(z)]\|_2]_+ \quad (3)$$

[0016] 其中, $[\cdot]_+$ 表示最大值, $(\cdot, 0)$ 和 α 是一个恒定的余量; X, Y, Z 表示图像集;其中, X 和 Y 来自同一类,但又不同于 Z ;减小MSML的数量,可以增加相同与不同组之间的相关性和差异分别。

[0017] 进一步地,所述的不考虑样本偏差和异常值的集合,最大平均偏差不考虑样本偏差和异常值的集合:

$$[0018] \quad \omega(x) = p(x)/(\hat{p}(x)) :$$

$$[0019] \quad \underset{\omega(x) \geq 0}{\text{minimize}} \|E_p[f(x)] - E_{\hat{p}}[\omega(x)f(x)]\|_2 \quad s.t. E_{\hat{p}}[\omega(x)] = 1 \quad (4)$$

[0020] 其中, \hat{p} 是一个分布,其对应于 p ; $\omega(x)$ 可以纠正样本偏差和消除异常值(或减少噪音)。

[0021] 进一步地,所述的整流平均差异(RMD), $\omega(x)$ 作为样本偏差为MMD修正项,整流平均差异(RMD)如下:

$$[0022] \quad \text{RMD}(X, Y) = \|E_{\omega(x)}[f(x)] - E_{\omega(x)}[f(y)]\|_2$$

[0023]

$$\left\| \sum_{i=1}^n \omega(x_i) f(x_i) - \sum_{j=1}^m \omega(y_j) f(y_j) \right\|_2 \quad s.t. \sum_{i=1}^n \omega(x_i) = 1, \sum_{j=1}^m \omega(y_j) = 1 \quad (5)$$

[0024] 结合公式(3)和公式(5),我们联结基于整流项的度量学习的平均集,并获得高级的度量学习方法,命名为注意定势度量学习的视频人脸识别:

$$[0025] \quad \text{ASML}(X, Y, Z) = \|E_{\omega(x)}[f(x)] - E_{\omega(x)}[f(y)]\|_2 + [\alpha - \|E_{\omega(x)}[f(x)] - E_{\omega(x)}[f(z)]\|_2]_+$$

$$[0026] \quad s.t. \sum \omega(x) = 1, \sum \omega(y) = 1, \sum \omega(z) = 1 \quad (6)$$

[0027] 如上式所示。

[0028] 其中,所述的记忆注意权重,权重应满足以下三点:

[0029] (1) 加权方法容易集成在CNN的框架,其参数是端到端可训练的监督方式;

[0030] (2) 权重基于全局内容和设置感知,因为构建了一个训练实例作为特殊的图像,能更好地利用信息;

[0031] (3) 学习权重图像顺序必须不变,每个图像的权重和图像在训练实例中是无序。

[0032] 进一步地,所述的权重,把权重作为神经图灵机(NTM),其中人脸特征集作为记忆,权重视为地址读写内存;NTM涉及三个基本组件:读取向量,擦除向量和添加向量;令 $\{f_i\}$ 成

为一个面部特征集,在 s_i 上应用softmax操作以形成归一化权重 ω_i ;公式(6)中操作和重新配置的 $E_\omega[f(x)]$ 如下:

$$[0033] \quad \omega_i = \frac{\exp(s_i)}{\sum_j \exp(s_j)} \quad (7)$$

$$[0034] \quad E_\omega[f(x)] = \sum_i \left[\frac{\exp(s_i)}{\sum_j \exp(s_j)} \cdot f_i \right] \quad (8)$$

[0035] 权重 $\{\omega_i\}$ 是无序的, $\{s_i\}$ 是可微的。

[0036] 进一步地,所述的将ASML自然融入到卷积神经网络(CNN)中,CNN包含29层残块卷积和最大特征映射操作;基于网络,ASML和记忆注意权重耦合,实现视频人脸识别;Softmax是重要的监督信号方法,得到以下目标函数:

$$[0037] \quad L = \lambda_1 \text{Softmax} + \lambda_2 \text{ASML} \quad (9)$$

[0038] 其中, λ_1 和 λ_2 是这两个项之间的权重;Softmax功能用于标准的人脸识别任务,ASML惩罚项增加了相关性、类似的分布和不同脸部集合之间的差异。

[0039] 进一步地,所述的卷积神经网络(CNN),网络中的所有参数是可微的,梯度可以计算:

$$[0040] \quad \frac{\partial L}{\partial \theta} = \lambda_1 \frac{\partial \text{Softmax}}{\partial \theta} + \lambda_2 \frac{\partial \text{ASML}}{\partial \theta} \quad (10)$$

$$[0041] \quad \frac{\partial L}{\partial s} = \lambda_2 \frac{\partial \text{ASML}}{\partial \omega} \cdot \frac{\partial \omega}{\partial s} \quad (11)$$

[0042] 其中, θ 代表网络参数。

附图说明

[0043] 图1是本发明一种基于注意定势度量学习的视频面部识别方法的系统流程图。

[0044] 图2是本发明一种基于注意定势度量学习的视频面部识别方法的基本框架图。

具体实施方式

[0045] 需要说明的是,在不冲突的情况下,本申请中的实施例及实施例中的特征可以相互结合,下面结合附图和具体实施例对本发明作进一步详细说明。

[0046] 图1是本发明一种基于注意定势度量学习的视频面部识别方法的系统流程图。主要包括注意定势的度量学习(ASML),记忆注意权重,将ASML自然融入到卷积神经网络(CNN)中。

[0047] 注意定势的度量学习(ASML),它纠正了样本偏差,并测量面部图像组之间的相关性,缩小了相同集合的概率分布之间的差距,同时扩大了不同集合之间的差距;与ASML组合的端对端可训练的深卷积神经网络(CNN),可以学习更多的歧视性深层次表示,获得可靠的人脸识别;ASML是最大平均偏差与记忆注意权重的广义扩展。

[0048] 最大平均偏差(MMD),设 \mathcal{F} 是一类函数 $f: \mathcal{X} \rightarrow \mathbb{R}$,同上定义 $p, q, X, Y; MMD$ 及其经验估计为:

$$[0049] \quad MMD[\mathcal{F}, p, q] := \sup_{f \in \mathcal{F}} (E_{x \sim p}[f(x)] - E_{y \sim q}[f(y)]) \quad (1)$$

$$[0050] \quad \text{MMD}[\mathcal{F}, X, Y] := \sup_{f \in \mathcal{F}} \left(\frac{1}{n} \sum_{i=1}^n f(x_i) - \frac{1}{m} \sum_{i=1}^m f(y_i) \right) \quad (2)$$

[0051] 当 \mathcal{F} 是紧凑度量空间 \mathcal{X} 上定义的一个单位球,方程 $\text{MMD}[\mathcal{F}, X, Y] = 0$ 成立,当且仅当 $p=q$;直观地说,MMD越小,X和Y之间的分布越相关;因此,MMD越大,分布之间的差异越大,特征完全不同的。

[0052] 基于集合的度量学习(MSML),根据最大平均偏差,基于集合的度量学习(MSML)表示为:

$$[0053] \quad \text{MSML}(X, Y, Z) = \|E_{x \sim p}[f(x)] - E_{y \sim q}[f(y)]\|_2 + [\alpha - \|E_{x \sim p}[f(x)] - E_{z \sim q}[f(z)]\|_2]_+ \quad (3)$$

[0054] 其中, $[\cdot]_+$ 表示最大值, $(\cdot, 0)$ 和 α 是一个恒定的余量; X, Y, Z 表示图像集;其中, X 和 Y 来自同一类,但又不同于 Z ;减小MSML的数量,可以增加相同与不同组之间的相关性和差异分别。

[0055] 不考虑样本偏差和异常值的集合,最大平均偏差不考虑样本偏差和异常值的集合:

$$[0056] \quad \omega(x) = p(x)/(\hat{p}(x)) :$$

$$[0057] \quad \underset{\omega(x) \geq 0}{\text{minimize}} \|E_p[f(x)] - E_{\hat{p}}[\omega(x)f(x)]\|_2 \quad s.t. E_{\hat{p}}[\omega(x)] = 1 \quad (4)$$

[0058] 其中, \hat{p} 是一个分布,其对应于 p ; $\omega(x)$ 可以纠正样本偏差和消除异常值(或减少噪音)。

[0059] 整流平均差异(RMD), $\omega(x)$ 作为样本偏差为MMD修正项,整流平均差异(RMD)如下:

$$[0060] \quad \text{RMD}(X, Y) = \|E_{\omega(x)}[f(x)] - E_{\omega(x)}[f(y)]\|_2$$

[0061]

$$\left\| \sum_{i=1}^n \omega(x_i)f(x_i) - \sum_{j=1}^m \omega(y_j)f(y_j) \right\|_2 \quad s.t. \sum_{i=1}^n \omega(x_i) = 1, \sum_{j=1}^m \omega(y_j) = 1 \quad (5)$$

[0062] 结合公式(3)和公式(5),我们联结基于整流项的度量学习的平均集,并获得高级的度量学习方法,命名为注意定势度量学习的视频人脸识别:

$$[0063] \quad \text{ASML}(X, Y, Z) = \|E_{\omega(x)}[f(x)] - E_{\omega(x)}[f(y)]\|_2 + [\alpha - \|E_{\omega(x)}[f(x)] - E_{\omega(x)}[f(z)]\|_2]_+$$

$$[0064] \quad s.t. \sum \omega(x) = 1, \sum \omega(y) = 1, \sum \omega(z) = 1 \quad (6)$$

[0065] 如上式所示。

[0066] 记忆注意权重,权重应满足以下三点:

[0067] (1) 加权方法容易集成在CNN的框架,其参数是端到端可训练的监督方式;

[0068] (2) 权重基于全局内容和设置感知,因为构建了一个训练实例作为特殊的图像,能更好地利用信息;

[0069] (3) 学习权重图像顺序必须不变,每个图像的权重和图像在训练实例中是无序。

[0070] 把权重作为神经图灵机(NTM),其中人脸特征集作为记忆,权重视为地址读写内存;NTM涉及三个基本组件:读取向量,擦除向量和添加向量;令 $\{f_i\}$ 成为一个面部特征集,在 s_i 上应用softmax操作以形成归一化权重 ω_i ;公式(6)中操作和重新配置的 $E_{\omega}[f(x)]$ 如下:

$$[0071] \quad \omega_i = \frac{\exp(s_i)}{\sum_j \exp(s_j)} \quad (7)$$

$$[0072] \quad E_{\omega}[f(x)] = \sum_i \left[\frac{\exp(s_i)}{\sum_j \exp(s_j)} \cdot f_i \right] \quad (8)$$

[0073] 权重 $\{\omega_i\}$ 是无序的, $\{s_i\}$ 是可微的。

[0074] 将ASML自然融入到卷积神经网络(CNN)中,CNN包含29层残块卷积和最大特征映射操作;基于网络,ASML和记忆注意权重耦合,实现视频人脸识别;Softmax是重要的监督信号方法,得到以下目标函数:

$$[0075] \quad L = \lambda_1 \text{Softmax} + \lambda_2 \text{ASML} \quad (9)$$

[0076] 其中, λ_1 和 λ_2 是这两个项之间的权重;Softmax功能用于标准的人脸识别任务,ASML惩罚项增加了相关性、类似的分布和不同脸部集合之间的差异。

[0077] 网络中的所有参数是可微的,梯度可以计算:

$$[0078] \quad \frac{\partial L}{\partial \theta} = \lambda_1 \frac{\partial \text{Softmax}}{\partial \theta} + \lambda_2 \frac{\partial \text{ASML}}{\partial \theta} \quad (10)$$

$$[0079] \quad \frac{\partial L}{\partial s} = \lambda_2 \frac{\partial \text{ASML}}{\partial \omega} \cdot \frac{\partial \omega}{\partial s} \quad (11)$$

[0080] 其中, θ 代表网络参数。

[0081] 图2是本发明一种基于注意定势度量学习的视频面部识别方法的基本框架图。该框架的每个训练实例由三个不同脸部视频的子集组成。与ASML组合的端对端可训练的深卷积神经网络(CNN),可以学习更多的歧视性深层次表示,获得可靠的人脸识别。过程为:先定义图像集上的有效距离度量,显著地最小化集合内距离,并同时最大化集中距离,再把权重作为神经图灵机,其中人脸特征集作为记忆,权重作为地址读写内存,最后将ASML自然融入到卷积神经网络中,从而形成端到端的学习方案。

[0082] 对于本领域技术人员,本发明不限制于上述实施例的细节,在不背离本发明的精神和范围的情况下,能够以其他具体形式实现本发明。此外,本领域的技术人员可以对本发明进行各种改动和变型而不脱离本发明的精神和范围,这些改进和变型也应视为本发明的保护范围。因此,所附权利要求意欲解释为包括优选实施例以及落入本发明范围的所有变更和修改。

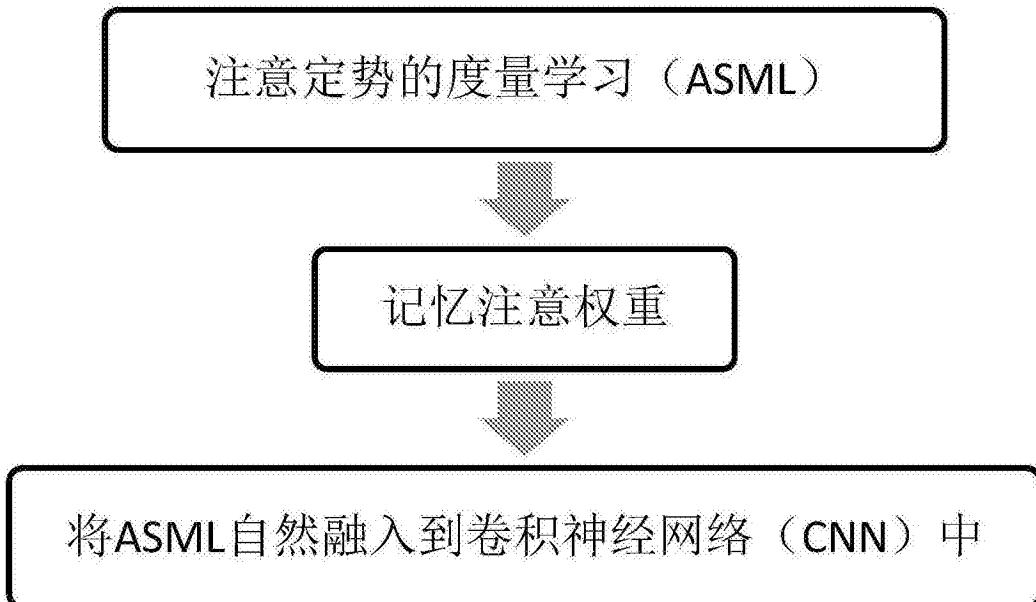


图1

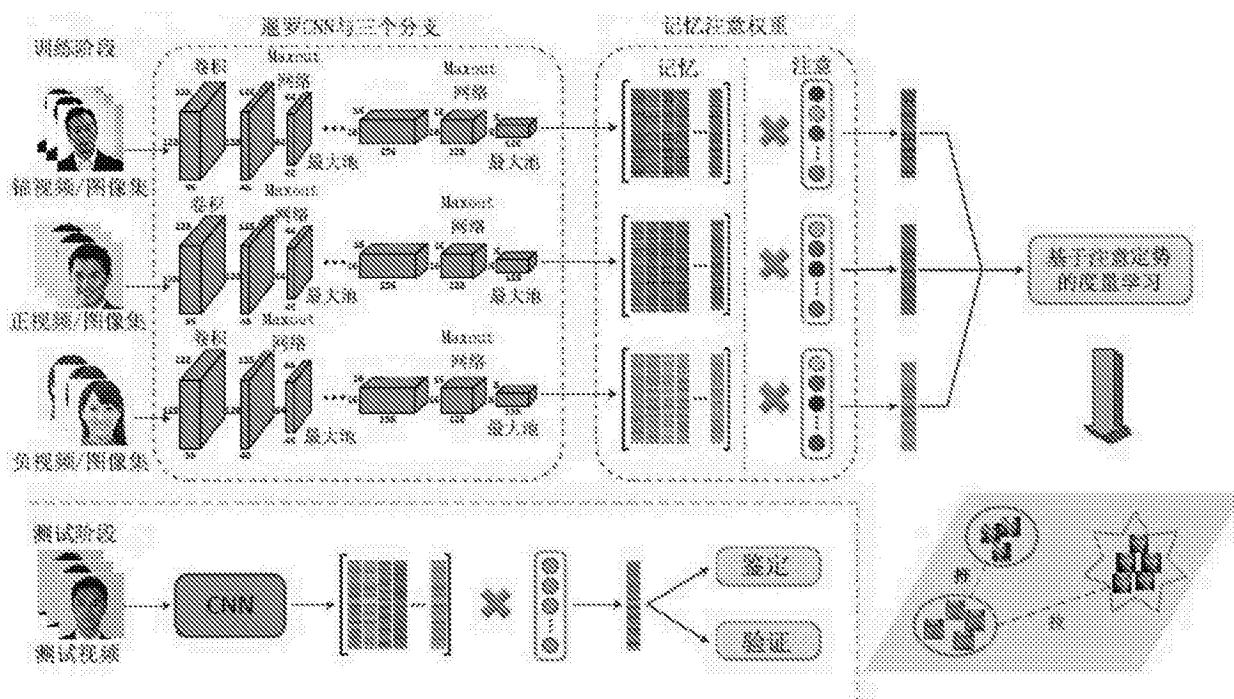


图2