(19)国家知识产权局

(12)发明专利

(10)授权公告号 CN 112200075 B
(45)授权公告日 2024.06.04

(21)申请号 202011075186.9

(22)申请日 2020.10.09

(65)同一申请的已公布的文献号
申请公布号 CN 112200075 A

(43)申请公布日 2021.01.08

(73)专利权人 西安西图之光智能科技有限公司
地址 710075 陕西省西安市沣东新城红光大道协同创新港2号楼301室-123号工位

(72)发明人 郝坤坤 魏丹丹 李慧斌

(74)专利代理机构 深圳泛航知识产权代理事务所(普通合伙) 44867
专利代理师 邓爱军

(51)Int.Cl.
G06V 40/16(2022.01)
G06V 10/82(2022.01)
G06N 3/0464(2023.01)
G06N 3/08(2023.01)
G06V 10/774(2022.01)

(56)对比文件
AU 2019100806 A4,2019.08.29
CA 2625795 A1,2009.10.25
CN 103034874 A,2013.04.10
CN 107862299 A,2018.03.30
CN 110516616 A,2019.11.29
CN 110956681 A,2020.04.03
CN 109858368 A,2019.06.07
CN 111160313 A,2020.05.15
WO 2020199475 A1,2020.10.08
CN 111695432 A,2020.09.22
CN 109753864 A,2019.05.14
CN 107247916 A,2017.10.13
CN 107992842 A,2018.05.04
CN 111639589 A,2020.09.08
CN 104463137 A,2015.03.25
WO 2019214557 A1,2019.11.14
CN 110443203 A,2019.11.12
宛根训;田青;朱红徽;葛利军.人脸识别应用活体检测技术研究.中国安全防范技术与应用.2019,(06),第59-63页.
汪亚航;宋晓宁;吴小俊.结合混合池化的双流人脸活体检测网络.中国图象图形学报.2020,(07),第130-142页.

审查员 肖明月

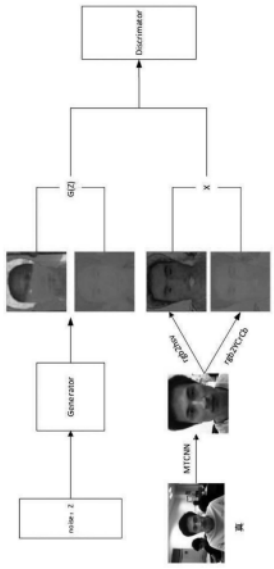
权利要求书2页 说明书5页 附图4页

(54)发明名称

一种基于异常检测的人脸防伪方法

(57)摘要

本发明公开了一种基于异常检测的人脸防伪方法,包括使用真实人脸数据集对其进行裁剪对齐,将其变换到HSV、YCbCr颜色空间,然后进行拼接生成HSV-YCbCr特征;使用GAN网络来学习真实人脸HSV-YCbCr特征的分布,在训练集上对GAN网络进行无监督训练;测试时,使用训练后的模型将预处理后的人脸HSV-YCbCr特征用梯度下降方法映射回隐空间,计算loss值,然后与阈值进行比较判断测试图片的真假。



CN 112200075 B

1.一种基于异常检测的人脸防伪方法,其特征在于:包括以下步骤:

步骤1:使用RGB人脸数据集,预处理后,将其转换到HSV空间和YCbCr颜色空间,然后拼接成6通道的HSV-YCbCr特征;

步骤2:从异常检测角度出发,使用真实人脸数据集的HSV-YCbCr特征,对生成网络模型进行无监督训练;

步骤3:根据训练好的生成器G和判别器D,将测试集图片的HSV-YCbCr特征x映射回隐空间中,即$x \mapsto z$,得到z;

步骤4:使用与测试集图片的HSV-YCbCr特征对应的隐空间的值z,将其带入loss计算公式$L(z)=\lambda L_G(z)+(1-\lambda)L_D(z)$中计算,与阈值η进行比较,若计算出的结果大于阈值,即若$L(z)>\eta$则将HSV-YCbCr特征x对应的裁剪前的RGB原人脸图片判断为攻击,否则判断其为真实人脸图片;

步骤1包括如对预处理的图片将其转换到HSV下步骤:

步骤1.1:对预处理的图片将其转换到HSV空间,转换方式为:

$$H=\begin{cases}0^\circ, & if\ MAX=MIN\\ 60^\circ\times\dfrac{G-B}{\Delta}+0^\circ, & if\ MAX=R\ and\ G\geq B\\ 60^\circ\times\dfrac{G-B}{\Delta}+360^\circ, & if\ MAX=R\ and\ G<B\\ 60^\circ\times\dfrac{B-R}{\Delta}+120^\circ, & if\ MAX=G\\ 60^\circ\times\dfrac{R-G}{\Delta}+240^\circ, & if\ MAX=B\end{cases};$$

$$S=\begin{cases}0, & if\ MAX=0\\ 1-\dfrac{MIN}{MAX}, & if\ otherwise\end{cases};$$

V=MAX;

其中MAX=max{R/255,G/255,B/255};

MIN=min{R/255,G/255,B/255},△=MAX-MIN;

步骤1.2:对预处理的图片将其转换到YCbCr空间,转换方式为:

$$\begin{pmatrix}Y\\C_b\\C_r\end{pmatrix}=\begin{pmatrix}0.299 & 0.587 & 0.114\\-0.169 & -0.331 & 0.500\\0.500 & -0.419 & -0.081\end{pmatrix}\cdot\begin{pmatrix}R\\G\\B\end{pmatrix};$$

步骤1.3:将上述两步骤得到的两个3通道特征进行拼接,拼接成一个6通道的HSV-YCbCr特征。

2.根据权利要求1所述的一种基于异常检测的人脸防伪方法,其特征在于:步骤2包括如下步骤:

步骤2.1:建立一个残差网络作为生成器,用于学习真实人脸HSV-YCbCr特征的分布,实现输入一个服从随机分布的向量z,能够生成一个跟真实样本具有同样分布的足以欺骗判

别器的特征;

步骤2.2建立一个一分类的卷积神经网络作为判别器,用于判断输入的样本是服从真实人脸分布的样本还是生成器生成的样本;

步骤2.3:对无监督训练所设计的网络的损失函数为GAN的生成对抗损失L_{G+D}和残差损失L_G,G代表生成器,D代表判别器,其中GAN的生成对抗损失如下:

$$L_{G+D}=\min_G\max_D V(G,D)=E_{x\ p_{data}(x)}\left[\log D(x)\right]+E_{z\ p_z(z)}\left[\log\left(1-D(G(z))\right)\right]$$残差损失L_G是通过对生成器生成的人脸特征和真实人脸的HSV-YCbCr特征做差运算,然后逐通道计算L1范数并求和所计算得到的,即$L_G=\sum_{channel=1}^{6}\left\|x-G(z)\right\|_1$,使得生成的人脸HSV-YCbCr特征更加真实。

3.根据权利要求1所述的一种基于异常检测的人脸防伪方法,其特征在于:步骤3中求解测试集图片的HSV-YCbCr特征到隐空间的映射,其损失函数包括残差损失L_G和基于特征匹配的判别损失L_D,其中残差损失$L_G(z)=\sum_{channel=1}^{6}\left\|x-G(z)\right\|_1$,表示测试集图片的HSV-YCbCr特征与生成器生成的特征的差异;基于特征匹配的判别损失$L_D(z)=||f(x)-f(G(z))||_1$,表示测试集图片HSV-YCbCr特征的特征与生成器所生成的特征的特征差异,通过最小化这两个损失函数来找到使得测试集图片的HSV-YCbCr特征x与生成器G所生成的6通道特征差异最小的隐变量z。

4.根据权利要求3所述的一种基于异常检测的人脸防伪方法,其特征在于:具体为:给定一个HSV-YCbCr特征x,在隐空间中找到最优的z对应到G(z),HSV-YCbCr特征x和G(z)的相似度在于特征x服从用于训练生成器的G的训练集的分布p_{data}的程度,通过最小化总损失函数$L=\lambda L_G(z)+(1-\lambda)L_D(z)$,以梯度下降的形式来更新参数z,即$z\leftarrow z-\alpha\frac{\partial L}{\partial z}$,直到z的变化量$\nabla z$几乎不再改变时停止迭代更新,最终找到最优的z,α为超参数;

其中f为判别器结构中至全局池化层GAP部分,λ为超参数。

5.根据权利要求4所述的一种基于异常检测的人脸防伪方法,其特征在于:所有超参数的经验值设为$\alpha=0.001$,$\beta=0.4$,$\lambda=0.9$,$\eta=0.1$。

一种基于异常检测的人脸防伪方法

技术领域

[0001] 本发明涉及一种基于异常检测的人脸防伪方法,属于人脸识别技术领域。

背景技术

[0002] 随着科技的迅猛发展和人脸识别技术本身的优越性如非侵入性、安全性等特点,人脸识别技术得到了越来越广泛的应用,如手机解锁,刷脸支付等。然而人脸识别系统容易受到非法用户的恶意攻击,如冒充者打印出别人的照片想以此来骗过人脸识别系统。因此人脸防伪技术成为人脸识别过程中必不可少的一个环节。

[0003] 人脸防伪是指人脸识别系统可以有效的区分出真脸和假脸。假脸也称为攻击,一般分为打印攻击、重放攻击和面具攻击。打印攻击是指冒充者打印出合法用户的照片,试图以此攻破人脸识别系统。重放攻击是指攻击者试图以显示在电子屏幕上合法用户的照片或者视频来攻击人脸识别系统。面具攻击是指攻击者通过佩戴面具的行为来攻击人脸识别系统。

[0004] 人脸防伪方法一般是将人脸防伪看成是一个二分类问题来处理。一般传统方法由LBP、SIFT、LPQ、IMQ等特征提取器和SVM、SRC、LDA等分类器组成,基于深度学习的方法则采用了神经网络来对图片进行特征提取及最终分类。这种二分类的策略,需要同时收集真脸样本和假脸样本训练,在需要增加训练样本量的时候不仅需要真脸样本还需要攻击样本,很难做到均衡;此外总会有新的攻击方式出现,这种基于分类的方法仅能对训练时用到的样本类型进行判断,泛化性能差。异常检测则为人脸防伪提供了另外一种思路。异常检测是指在一个模式中找到异常的特性或行为。异常检测可以只对正常样本进行训练,为了提高算法性能可以很容易地增加训练样本的规模,无需考虑正负样本数据量之间的均衡问题,针对先前未见的攻击具有泛化性能。

[0005] 因此,本发明提出的一种基于异常检测的人脸防伪方法,能够解决泛化性能差及正负样本数据量之间不平衡的问题,具有重要的实际应用价值。

发明内容

[0006] 本发明的目的在于提供一种基于异常检测的人脸防伪方法,以解决上述背景技术中提出的问题。

[0007] 一种基于异常检测的人脸防伪方法,包括以下步骤:

[0008] 步骤1:使用RGB人脸数据集(全部真脸,正样本),预处理后,将其转换到HSV空间和YCbCr颜色空间,然后拼接成6通道的HSV-YCbCr特征;

[0009] 步骤2:从异常检测角度出发,使用真实人脸数据集的HSV-YCbCr特征,对生成网络模型(GAN,一个生成器,一个判别器)进行无监督训练;

[0010] 步骤3:使用训练完成的模型将测试集图片的HSV-YCbCr特征映射回隐空间;即根据训练好的生成器G和判别器D,将HSV-YCbCr特征x映射回隐空间中,即$x \mapsto z$,得到z;

[0011] 步骤4:使用与测试集图片的HSV-YCbCr特征对应的隐空间的值z,将其带入loss计

算公式$L(z)=\lambda L_G(z)+(1-\lambda)L_D(z)$中进行计算,与阈值进行比较,若计算出的结果大于阈值,即若$L(z)>\eta$则判断测试图像为攻击,否则判断其为真实人脸图片;

[0012] 作为本发明进一步的方案,步骤1包括如下步骤:

[0013] 步骤1.1:对预处理的图片将其转换到HSV空间,转换方式为:

[0014] $$H=\begin{cases}0^\circ, & if\ MAX=MIN\\ 60^\circ\times\frac{G-B}{\Delta}+0^\circ, & if\ MAX=R\ and\ G\geq B\\ 60^\circ\times\frac{G-B}{\Delta}+360^\circ, & if\ MAX=R\ and\ G<B\\ 60^\circ\times\frac{B-R}{\Delta}+120^\circ, & if\ MAX=G\\ 60^\circ\times\frac{R-G}{\Delta}+240^\circ, & if\ MAX=B\end{cases};$$

[0015] $$S=\begin{cases}0, & if\ MAX=0\\ 1-\frac{MIN}{MAX}, & if\ otherwise\end{cases};$$

[0016] V=MAX;

[0017] 其中MAX=max{R/255,G/255,B/255};

[0018] MIN=min{R/255,G/255,B/255},$\nabla=MAX-MIN$;

[0019] 步骤1.2:对预处理的图片将其转换到YCbCr空间,转换方式为:

[0020] $$\begin{pmatrix}Y\\C_b\\C_r\end{pmatrix}=\begin{pmatrix}0.299 & 0.587 & 0.114\\-0.169 & -0.331 & 0.500\\0.500 & -0.419 & -0.081\end{pmatrix}\cdot\begin{pmatrix}R\\G\\B\end{pmatrix};$$

[0021] 步骤1.3:将上述两步骤得到的两个3通道特征进行拼接,拼接成一个6通道的HSV-YCbCr特征。

[0022] 作为本发明进一步的方案,步骤2包括如下步骤:

[0023] 步骤2.1:建立一个残差网络作为生成器,用于学习真实人脸HSV-YCbCr特征的分布,实现输入一个服从随机分布的向量z,能够生成一个跟真实样本具有同样分布的足以欺骗判别器的特征;

[0024] 步骤2.2建立一个一分类的卷积神经网络作为判别器,用于判断输入的样本是服从真实人脸分布的样本还是生成器生成的样本;

[0025] 步骤2.3:对无监督训练所设计的网络的损失函数为GAN的生成对抗损失L_{G+D}和残差损失L_G,其中GAN的生成对抗损失如下:

$L_{G+D}=\min_G\max_D V(G,D)=E_{x\sim p_{data}(x)}\left[\log D(x)\right]+E_{z\sim p_z(z)}\left[\log\left(1-D\left(G(z)\right)\right)\right]$残差损失$L_G$是通过对生成器生成的人脸特征和真实人脸的HSV-YCbCr特征做差运算,然后逐通道计算1范数并求和所计算得到的,即$L_G=\sum_{channel=1}^{6}\left\|x-G(z)\right\|_1$,使得生成的人脸HSV-YCbCr

特征更加真实。

[0026] 作为本发明进一步的方案,步骤3中求解测试集图片的HSV-YCbCr特征到隐空间的映射,其损失函数包括残差损失L_G和基于特征匹配的判别损失L_D,其中残差损失$L_G=\sum_{channel=1}^{6}\|x-G(z)\|_1$,表示测试集图片的HSV-YCbCr特征与生成器生成的特征的差异;基于特征匹配的判别损失$L_D(z)=||f(x)-f(G(z))||_1$,表示测试集图片HSV-YCbCr特征的特征与生成器所生成的特征的特征差异,通过最小化这两个损失函数来找到使得测试集图片的HSV-YCbCr特征x与生成器G所生成的6通道特征差异最小的隐变量z。

[0027] 作为本发明进一步的方案,步骤3具体为:给定一个HSV-YCbCr特征x,在隐空间中找到最优的z对应到G(z),HSV-YCbCr特征x和G(z)的相似度在于特征x服从用于训练生成器的G的训练集的分布p_{data}的程度,通过最小化总损失函数$L=\lambda L_G(z)+(1-\lambda)L_D(z)$,以梯度下降的形式来更新参数z,即$z\leftarrow z-\alpha\frac{\partial L}{\partial z}$(α为超参数),直到z的变化量$\nabla z$几乎不再改变时停止迭代更新,最终找到最优的z;

[0028] 其中$L_G(z)$为残差损失,$L_G(z)=\sum_{channel=1}^{6}\|x-G(z)\|_1$;$L_D(z)$为基于特征匹配的判别损失,$L_D(z)=||f(x)-f(G(z))||_1$,f为判别器结构中至全局池化层GAP部分,λ为超参数。

[0029] 作为本发明进一步的方案,所有超参数的经验值设为α=0.001,β=0.4,λ=0.9,η=0.1。

[0030] 与现有技术相比,本发明的有益效果是:本发明专利提出一种基于异常检测的人脸防伪方法,基于HSV,YCbCr的图像转换及HSV-YCbCr特征,避免了RGB三个通道的高相关性及其对亮度和色度的不完美分离;从异常检测角度出发,对真人脸图片的进行建模,建模时只需要正样本,不需要负样本,很容易通过增加正样本的数量来增加训练集的规模,不存在正负样本间数据不平衡的问题,对未见过的攻击具有更好的泛化性能。

附图说明

[0031] 图1是本申请实施例所述一种基于异常检测的人脸防伪方法的训练流程图;

[0032] 图2是本申请实施例所述生成器模型图;

[0033] 图3是本申请实施例所述判别器模型图;

[0034] 图4是本申请实施例所述测试流程图。

具体实施方式

[0035] 下面将结合本发明实施例中的附图,对本发明实施例中的技术方案进行清楚、完整的阐述。

[0036] 参照图4,其示出了本申请实施例所述一种基于异常检测的人脸防伪的流程图,包括:

[0037] 步骤1:使用RGB真人脸数据集(正样本,全部为真脸),并用MTCNN来对人脸照片进行对齐裁剪,变成256*256*3的图片,将裁剪后的RGB人脸数据集转换到HSV空间和YCbCr颜

色空间,然后拼接成256*256*6的HSV-YCbCr特征。

[0038] 步骤1.1:RGB转换到HSV方法为:

[0039] 首先将R,G,B值置于(0,1)间,即令R=R/255,G=G/255,B=B/255

[0040] 计算MAX=max{R,G,B},MIN=min{R,G,B},Δ=MAX-MIN则H(色调)、S(饱和度)、V(明度)分别为:

[0041] $$H=\begin{cases}0^\circ, & if\ MAX=MIN\\ 60^\circ\times\dfrac{G-B}{\Delta}+0^\circ, & if\ MAX=R\ and\ G\geq B\\ 60^\circ\times\dfrac{G-B}{\Delta}+360^\circ, & if\ MAX=R\ and\ G<B\\ 60^\circ\times\dfrac{B-R}{\Delta}+120^\circ, & if\ MAX=G\\ 60^\circ\times\dfrac{R-G}{\Delta}+240^\circ, & if\ MAX=B\end{cases};$$

[0042] $$S=\begin{cases}0, & if\ MAX=0\\ 1-\dfrac{MIN}{MAX}, & if\ otherwise\end{cases};$$

[0043] V=MAX

[0044] 步骤1.2:RGB转换到YCbCr方法为:

[0045] $$\begin{pmatrix}Y\\ C_b\\ C_r\end{pmatrix}=\begin{pmatrix}0.299 & 0.587 & 0.114\\ -0.169 & -0.331 & 0.500\\ 0.500 & -0.419 & -0.081\end{pmatrix}\cdot\begin{pmatrix}R\\ G\\ B\end{pmatrix};$$

[0046] 步骤1.3:将上述两步骤得到的两个3通道特征进行拼接,拼接成一个256*256*6的6通道的HSV-YCbCr特征。

[0047] 步骤2:建立一个基于GAN的网络架构。GAN由一个生成器和一个判别器构成。

[0048] 步骤2.1:所使用生成网络G为一个残差网络。输入是一个服从多元高斯分布的向量,输出是256*256*6的特征,具体结构如参照图2所示。生成器的目的是为了学习真实样本HSV-YCbCr特征的分布,实现给定一个输入向量z,生成一个真实的足以欺骗判别器的人脸特征。

[0049] 步骤2.2:建立一个一分类的卷积神经网络,作为判别器D,具体结构如参照图3所示。判别器的目的是能够准确的判断输入的样本是服从真实人脸分布的样本还是生成器生成的样本。

[0050] 步骤2.3:损失函数包括GAN的生成对抗损失L_{G+D}和残差损失L_G,总的损失函数$L=L_{G+D}+\beta L_G$,其中

[0051] $$L_{G+D}=\min_G\max_D V(G,D)=E_{x\sim p_{data}(x)}\left[\log D(x)\right]+E_{z\sim p_z(z)}\left[\log\left(1-D(G(z))\right)\right]$$

[0052] $L_G=\sum_{channel=1}^{6}\left\|x-G(z)\right\|_1$为正则项,β为超参数。

[0053] 生成器G将隐空间Z中服从100维标准高斯分布p_z的向量z映射到真人脸HSV-YCbCr特征的分布p_{data}的空间,即$z \sim p_z \rightarrow G(z) \sim p_{data}$。判别器D用来区别真实样本和生成器G生成的样本。G本来生成的是假样本,经过判别器D使得G生成的样本足以骗过判别器D,即生成的是真实样本而不是G生成的样本。同样地由于G生成的样本越来越真,同样使得判别器D的判别能力越来越强。二者彼此博弈,使得自身性能越来越高。对生成器G和判别器D采用对抗训练方式,首先固定生成器G来训练判别器D,之后固定判别器D来训练生成器G。

[0054] 步骤3:在训练数据集上对生成网络模型进行无监督训练,在验证集上进行超参数的选择,得到训练好的生成器G和判别器D。

[0055] 步骤4:对于测试图片进行预处理。首先使用MTCNN进行裁剪对齐,然后进行HSV,YCbCr颜色空间转换,拼接成大小为256*256*6的HSV-YCbCr特征。

[0056] 步骤5:根据训练好的生成器G和判别器D,将HSV-YCbCr特征x映射回隐空间中,即$x \mapsto z$,得到z。

[0057] 步骤5.1:给定一个HSV-YCbCr特征x,我们想在隐空间中找到最优的z对应到G(z)。HSV-YCbCr特征x和G(z)的相似度在于特征x服从用于训练生成器的G的训练集的分布p_{data}的程度。我们通过最小化总损失函数$L=\lambda L_G(z)+(1-\lambda)L_D(z)$,以梯度下降的形式来更新参数z,即$z \leftarrow z-\alpha\frac{\partial L}{\partial z}$(α为超参数),直到z的变化量$\nabla z$几乎不再改变时停止迭代更新,最终找到最优的z。

[0058] 其中$L_G(z)$为残差损失,$L_G(z)=\sum_{channel=1}^{6}\left\|x-G(z)\right\|_1$;$L_D(z)$为基于特征匹配的判别损失,$L_D(z)=||f(x)-f(G(z))||_1$,f为判别器结构中至全局池化层GAP部分,可参照图3,λ为超参数。

[0059] 步骤6:将得到的z值带入到公式$L(z)=\lambda L_G(z)+(1-\lambda)L_D(z)$计算与阈值η进行比较。若$L(z)>\eta$,将HSV-YCbCr特征x对应的裁剪前的RGB原人脸图片判断为攻击;否则,将其判断为真人照片。

[0060] 本实施例中所有超参数的经验值设为$\alpha=0.001$,$\beta=0.4$,$\lambda=0.9$,$\eta=0.1$。

[0061] 测试一张图片是攻击还是真人的流程,参照图4所示。

[0062] 综上所述,本实施例公开了一种基于异常检测的人脸防伪方法,基于HSV,YCbCr的图像转换及HSV-YCbCr特征,避免了RGB三个通道的高相关性及其对亮度和色度的不完美分离,基于异常检测的角度对真人脸的HSV-YCbCr特征进行建模,可对未见过的攻击具有更好的泛化性能。

[0063] 以上所述为本发明较佳实施例,对于本领域的普通技术人员而言,根据本发明的教导,在不脱离本发明的原理与精神的情况下,对实施方式所进行的改变、修改、替换和变型仍落入本发明的保护范围之内。

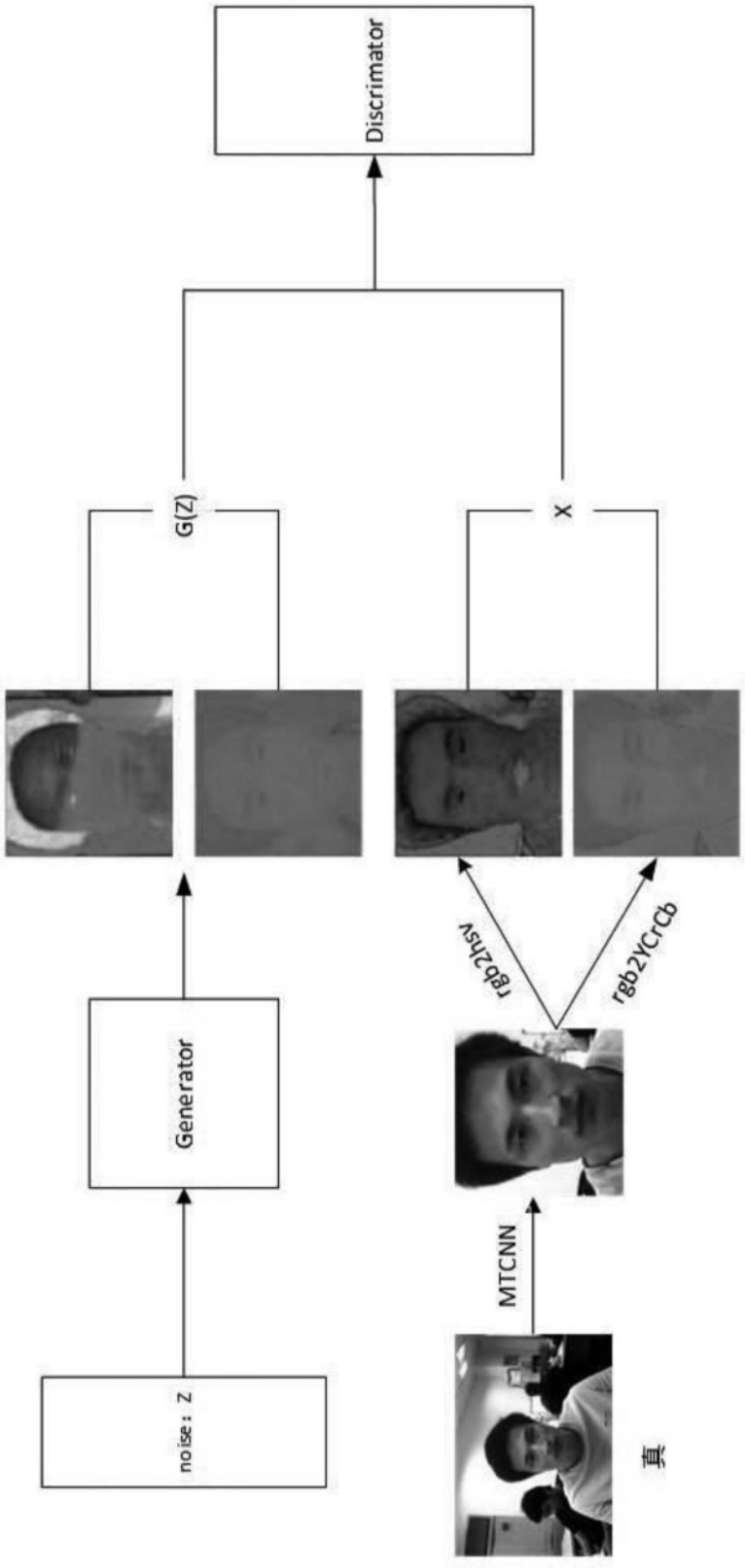

Discrimator
G(Z)
X
Generator
noise：Z
rgb2hsv
rgb2YCrCb
MTCNN
真

图1

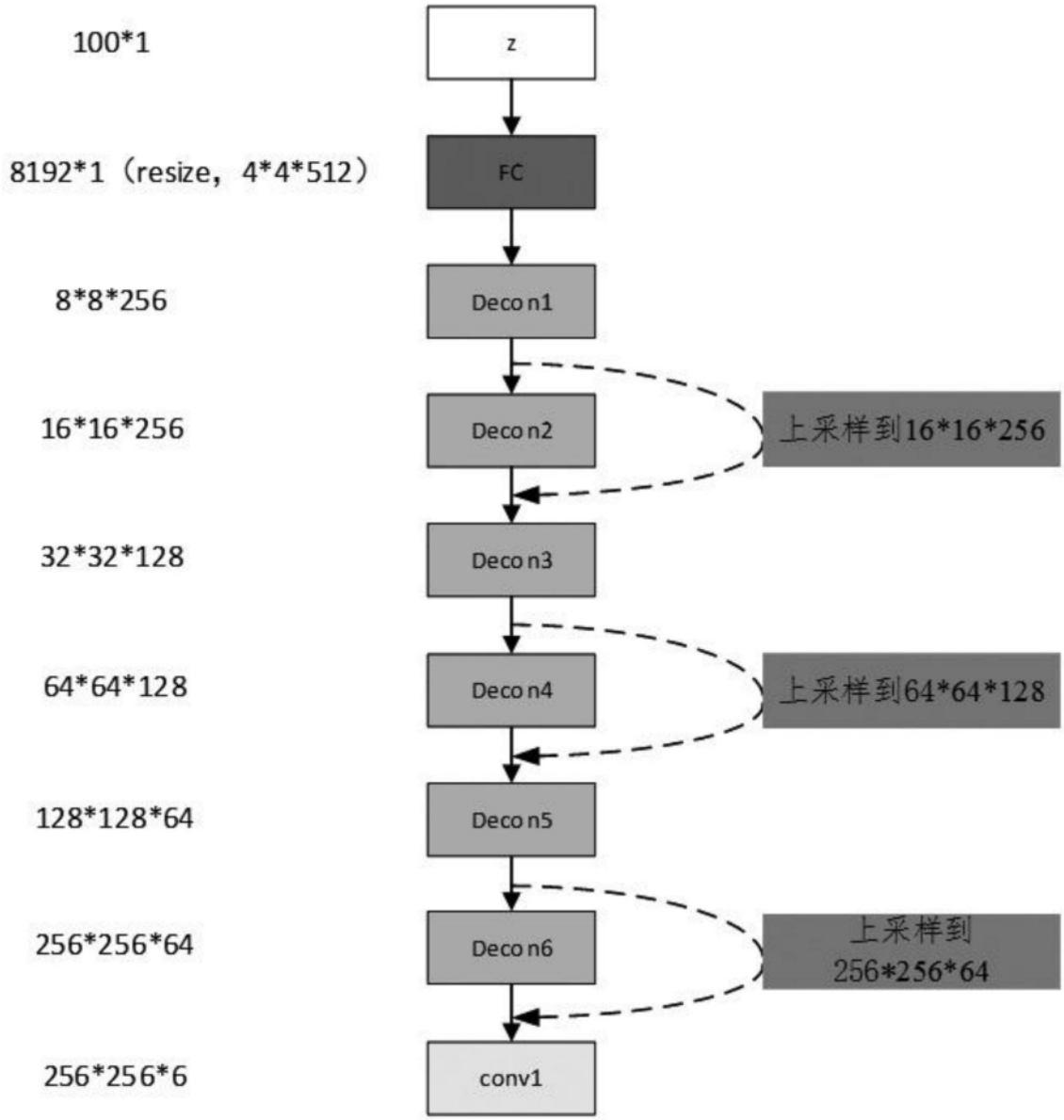

100*1
z
8192*1（resize，4*4*512）
FC
8*8*256
Decon1
16*16*256
Decon2
上采样到16*16*256
32*32*128
Decon3
64*64*128
Decon4
上采样到64*64*128
128*128*64
Decon5
256*256*64
Decon6
上采样到
256*256*64
256*256*6
conv1

图2

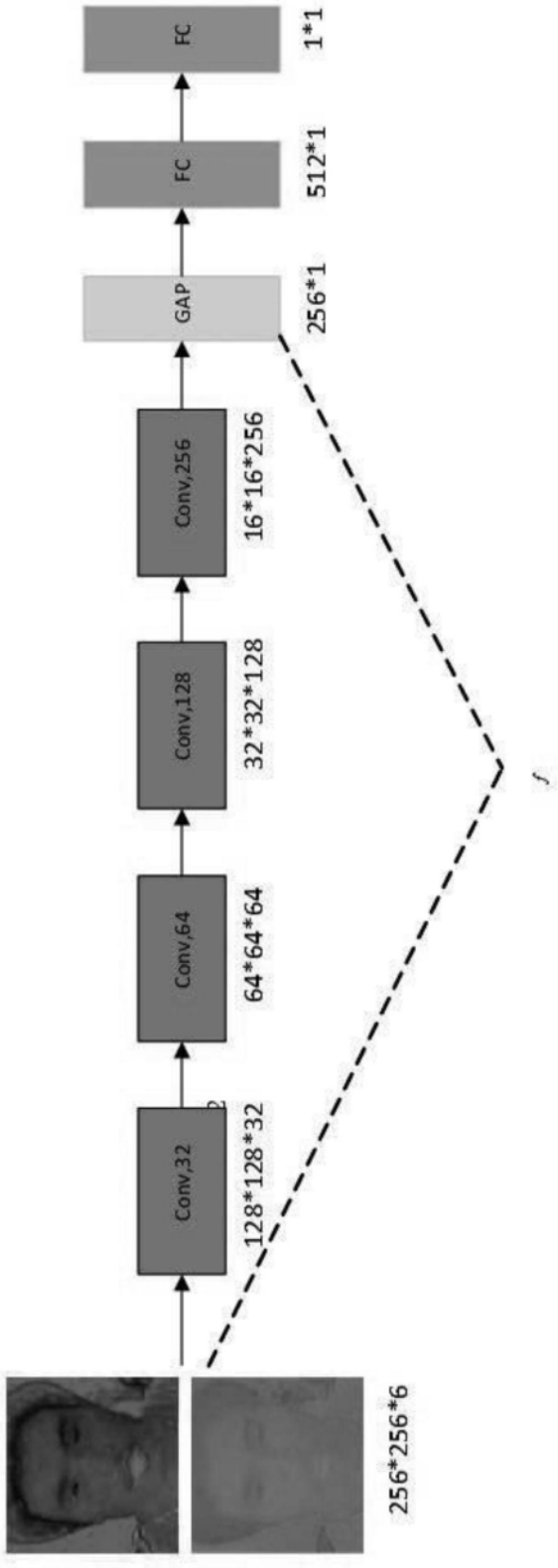

FC
1*1
FC
512*1
GAP
256*1
Conv,256
16*16*256
Conv,128
32*32*128
Conv,64
64*64*64
Conv,32
128*128*32
256*256*6
f

图3

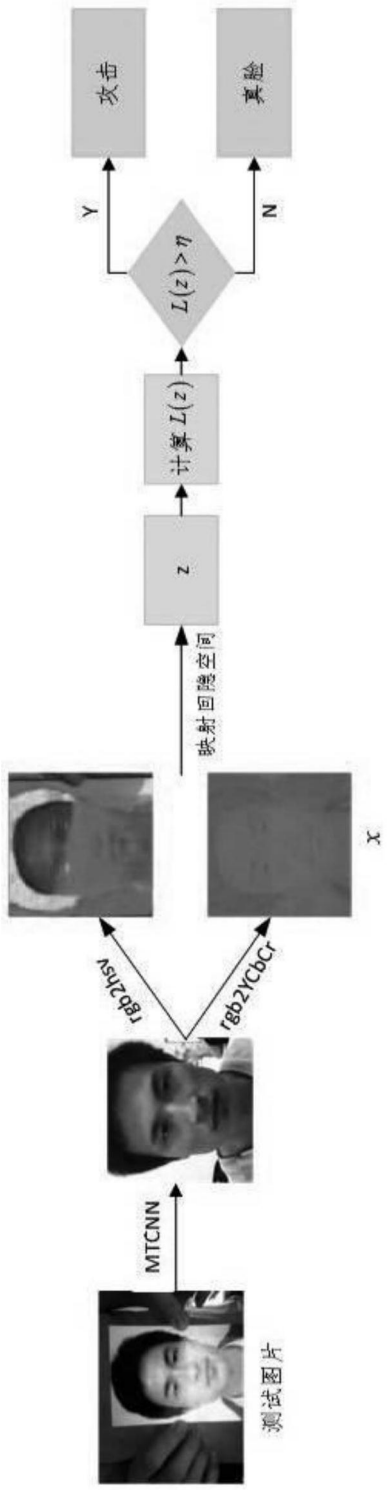


图4