



(12) 发明专利

(10) 授权公告号 CN 109376582 B

(45) 授权公告日 2022. 07. 29

(21) 申请号 201811024740.3

G06V 10/82 (2022.01)

(22) 申请日 2018.09.04

G06K 9/62 (2022.01)

(65) 同一申请的已公布的文献号

G06N 3/04 (2006.01)

申请公布号 CN 109376582 A

G06N 3/08 (2006.01)

(43) 申请公布日 2019.02.22

(56) 对比文件

(73) 专利权人 电子科技大学

CN 108446667 A, 2018.08.24

地址 611731 四川省成都市高新区(西区)

CN 108038823 A, 2018.05.15

西源大道2006号

US 2009087035 A1, 2009.04.02

(72) 发明人 李宏亮 梁小娟 邓志康 颜海强

CN 108229349 A, 2018.06.29

尹康 袁欢

CN 108182657 A, 2018.06.19

(74) 专利代理机构 电子科技大学专利中心

孙晶晶等.基于人像字典集的卡通自动生成方法.《系统仿真学报》.2015,第27卷(第4期),

51203

Yifan Liu等.Auto-painter: Cartoon

专利代理师 周刘英

Image Generation from Sketch by Using

(51) Int. Cl.

Conditional Generative Adversarial

G06V 40/16 (2022.01)

Networks.《https://arxiv.org/pdf/

G06V 10/26 (2022.01)

1705.01908.pdf》.2017,

G06V 10/44 (2022.01)

审查员 刘穗君

G06V 10/764 (2022.01)

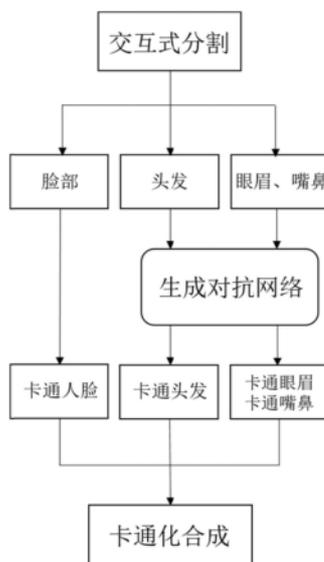
权利要求书2页 说明书5页 附图1页

(54) 发明名称

一种基于生成对抗网络的交互式人脸卡通方法

(57) 摘要

本发明公开了一种基于生成对抗网络的交互式人脸卡通方法。本发明对待处理的图像,首先进交互式分割处理,得到眉眼、嘴鼻、头发和脸部图像,再将眉眼、嘴鼻、头发分别输入三个训练好的眼、嘴鼻、头发生成模型,输出对应的卡通五官图像;基于脸部图像的卡通化处理,直接得到卡通人脸;再将人脸五官合成在卡通人脸,并叠加上头发效果,得到最终的卡通图像。本发明利用交互式和生成对抗网络的优势,通过交互式分割得到人物头发、脸型与人脸五官,消除训练样本之间由于背景不同导致的差异性,再通过生成对抗网络对各个部分进行风格转化,尽可能多地保留眼角、嘴角等细节部分的信息。



1. 一种基于生成对抗网络的交互式人脸卡通方法,其特征在于,包括下列步骤:

步骤S1:基于一一对应的真实人脸图像和卡通人脸图像,采用生成对抗网络训练关于头发、眉眼、嘴鼻的卡通生成模型:

对真实人脸图像和卡通人脸图像进行尺寸归一化处理;再对所有图像进行交互式分割处理,获取眉眼数据库、嘴鼻数据库和头发数据库;其中,卡通人脸图像的获取方式为:将专业画师绘画的同一种风格的卡通头像作为对应真实人脸图像的卡通映射图;

构建三个生成对抗网络,每个生成对抗网络包括两对生成器和判别器,生成器 G_A 用于将真实人脸图像生成卡通人脸图像,生成器 G_B 用于将卡通图像生成真实人脸图像,判别器 D_A 的输入为生成器 G_A 生成的卡通人脸图像 $G_A(y)$ 和卡通人脸图像 y ;判别器 D_B 的输入为生成器 G_B 生成的真实人脸图像 $G_B(x)$ 和真实人脸图像 x ;

训练时,将两个生成器输出的生成图像作为正样本增加到彼此的正训练样本集中,进行循环训练,直到判别精度达到精度需求;再将训练好的生成对抗网络中的生成器 G_A 作为一个卡通生成模型,从而得到关于头发、眉眼、嘴鼻的卡通生成模型;

训练时,判别器的损失函数为:

$$\text{loss}_{D_A} = E_{y \sim P_{data}(y)} [\log(1 - D_A(G_A(y)))] + E_{x \sim P_{data}(x)} [\log(D_A(x))]$$

$$\text{loss}_{D_B} = E_{x \sim P_{data}(x)} [\log(1 - D_B(G_B(x)))] + E_{y \sim P_{data}(y)} [\log(D_B(y))]$$

且在训练过程中保留上一次生成的图像,记为fake',每一个epoch更新它的数据,同时,计入下一个epoch的损失函数中: $\text{loss}'_D = \text{loss}_D + E[\log(\text{fake}')]]$,其中,D表示判别器 D_A 或 D_B ,epoch表示每次训练量, $E[]$ 表示取均值;

所述生成器是根据判别器的输出来进行训练更新,训练时,将两个生成器输出的生成图像作为正样本增加到彼此的正训练样本集中,进行循环训练,直到判别精度达到精度需求;再将训练好的生成对抗网络中的生成器 G_A 作为一个卡通生成模型,从而得到关于头发、眉眼、嘴鼻的卡通生成模型;

所述生成器在训练时的损失函数为:

$$\text{loss}_{G_A} = E_{y \sim P_{data}(y)} [\log(D_A(G_A(y)))]$$

$$\text{loss}_{G_B} = E_{x \sim P_{data}(x)} [\log(D_B(G_B(x)))]$$

各卡通生成模型的最终损失函数为 $L_G = \text{loss}_{G_A} + \text{loss}_{G_B} + \lambda_1 L_{cyc}(G_A, G_B) + \lambda_2 L_{idt}$,其中 loss_{G_A} 、 loss_{G_B} 分别表示生成器 G_A 、 G_B 的最终损失函数, $L_{cyc}(G_A, G_B)$ 表示循环损失函数, L_{idt} 表示重构损失函数;

所述重构损失函数 L_{idt} 具体为:

$$L_{idt} = E_{y \sim p_{data}(y)} [||\text{decoder}_B(\text{encoder}_B(y)) || - y] + E_{x \sim p_{data}(x)} [||\text{decoder}_A(\text{encoder}_A(x)) || - x],$$

其中,x表示生成器 G_A 的输入图像,即真实人脸图像,y表示生成器 G_B 的输入图像,即卡通人脸图像;

$\text{encoder}_B(\cdot)$ 表示提取括号中的卡通图像的图像特征, $\text{encoder}_A(\cdot)$ 表示提取括号中的真实的图像特征;

$\text{decoder}_B(\cdot)$ 表示生成器 G_B 生成的真实人脸图像, $\text{decoder}_A(\cdot)$ 表示生成器 G_A 生成的卡通人脸图像;

$E_{\zeta \sim p_{data}(\zeta)}[\]$ 表示取均值, 下标 $\zeta \sim p_{data}(\zeta)$ 用于表示图像 ζ 所属的图像类型, 其中 $\zeta \in \{x, y\}$, $p_{data}(x)$ 表示真实图像, $p_{data}(y)$ 表示卡通图像;

步骤S2: 对待处理的真实人脸图像进行交互式分割处理, 获取眉眼、嘴鼻、头发和人脸轮廓图像, 再将眉眼、嘴鼻、头发分别输入对应的生成模型, 输出对应的卡通图像: 卡通眉眼、卡通嘴鼻和卡通头发;

对人脸轮廓图像, 取脸部肤色的平均像素值进行填充, 得到卡通人脸;

将生成的卡通眉眼、卡通嘴鼻合成在卡通人脸上, 并叠加上卡通头发, 得到最终的卡通图像。

2. 如权利要求1所述的方法, 其特征在于, 所述交互式分割处理具体为:

用不同颜色的线条分别标注头发区域、背景区域, 用不同颜色的点标注双眼位置以及嘴唇位置, 其中双眼和嘴唇的颜色相同, 但与头发、背景的颜色不同;

以头发的标注颜色的色彩区域作为前景, 其他区域作为背景, 分割出头发图像;

以双眼的标注颜色的色彩点的相连区域作为前景, 其他区域作为背景, 分割出人脸部分, 从而得到人脸轮廓图像;

再通过双眼和嘴唇标注点剪切出左右眼眉图像和嘴唇图像。

一种基于生成对抗网络的交互式人脸卡通方法

技术领域

[0001] 本发明属于图像处理和计算机视觉技术领域,是一种基于生成对抗网络的交互式人脸卡通方法。

背景技术

[0002] 近年来,伴随着移动网络和手机的快速发展,特别是移动互联网的蓬勃发展,数字娱乐凭借其特殊优势在产业内规模越来越大,人脸卡通画技术应运而生,这一技术主要是利用图像处理和计算机程序自动地生成卡通化结果,能够极大地减少绘画师的工作量,提高卡通化效率。

[0003] 目前人脸卡通化方法主要有基于交互式的方法和无监督学习的方法。基于交互式的方法采用了人机交互的方式,提取出人脸轮廓线条,然后对人脸特征突出区域进行人为的夸张变形来生成个性卡通画,虽然交互式能够更加准确的得到人脸关键点信息,但是这种方法只能生成单一的卡通线条画,没有颜色渲染,并且耗时比较长。无监督学习的方法主要是利用机器学习或深度学习的方法对人脸图像进行卡通化转换。基于机器学习的方法中,一类是使用变形或颜色变换的方法对输入人脸图像像素进行处理,从而得到一定的卡通效果;另一类是基于样本学习的方法,对人脸图像整体进行特征提取,学习从人脸图像到风格图像的映射关系,然后生成卡通人脸。深度学习作为机器学习的分支,沿用了机器学习的思想,通过建立神经网络,能够学习到更具有高层语义特征,从而具有更强的学习能力,被广泛应用于物体检测、分类、分割与图像生成等领域。

[0004] 生成对抗网络也属于深度神经网络的一种,实质上是一种生成模型与判别模型的结合,生成模型是通过训练样本学习到输入到输出的条件概率分布,即一种生成关系,判别模型是由数据直接学习决策函数或预测函数,即对结果的决策或预测。生成对抗网络将两种模型相结合,并建立一种竞争的关系,利用生成模型将输入噪声或图片转换成风格图像,利用判别模型来判断生成图像的真假,真实样本即为真,生成图像记为假。训练过程既是一种相互博弈的过程,生成模型想要尽可能的欺骗判别网络,而判别网络想要尽可能的识别出假的图片。基于这一原理,生成对抗网络广泛应用于图像生成与风格迁移,相较于一般的生成模型,该网络能够生成更加清晰的图像,但是细节仍有不足,同时依赖于大量的风格类似的训练样本。

发明内容

[0005] 本发明的发明目的在于:针对上述存在的问题,提供

[0006] 一般的卡通化转换受到复杂背景的影响,生成的图像质量不佳。为了将训练模型能应用于数据集以外的图像当中,本发明尝试在对人脸头像进行卡通化之前进行交互式分割方法得到人物头发、脸部轮廓以及人脸五官,并分别进行卡通化转换,提升卡通画的细节处理,最后合成为完整的卡通图像。

[0007] 本发明的基于生成对抗网络的交互式人脸卡通方法,包括下列步骤:

[0008] 步骤S1:基于一一对应的真实人脸图像和卡通人脸图像,采用生成对抗网络训练关于头发、眉眼、嘴鼻的卡通生成模型:

[0009] 对真实人脸图像和卡通人脸图像进行尺寸归一化处理;再对所有图像进行交互式分割处理,获取眉眼数据库、嘴鼻数据库和头发数据库;

[0010] 构建三个生成对抗网络,每个生成对抗网络包括两对生成器和判别器,生成器 G_A 用于将真实人脸图像生成卡通人脸图像,生成器 G_B 用于将卡通图像生成真实人脸图像;

[0011] 训练时,将两个生成器输出的生成图像作为正样本增加到彼此的正训练样本集中,进行循环训练,直到判别精度达到精度需求;再将训练好的生成对抗网络中的生成器 G_A 作为一个卡通生成模型,从而得到关于头发、眉眼、嘴鼻的卡通生成模型;

[0012] 各卡通生成模型的最终损失函数为 $L_G = \text{loss}_{G_A} + \text{loss}_{G_B} + \lambda_1 L_{\text{cyc}}(G_A, G_B) + \lambda_2 L_{\text{idt}}$,其中 loss_{G_A} 、 loss_{G_B} 分别表示生成器 G_A 、 G_B 的最终损失函数, $L_{\text{cyc}}(G_A, G_B)$ 表示循环损失函数, L_{idt} 表示重构损失函数;

[0013] 其中 L_{idt} 具体为:

[0014] $L_{\text{idt}} = E_{y \sim \text{pdata}(y)} [|| \text{decoder}_B(\text{encoder}_B(y)) || - y] + E_{x \sim \text{pdata}(x)} [|| \text{decoder}_A(\text{encoder}_A(x)) || - x]$,

[0015] 其中, x 表示生成器 G_A 的输入图像,即真实人脸图像, y 表示生成器 G_B 的输入图像,即卡通人脸图像;

[0016] $\text{encoder}_B(\cdot)$ 表示提取括号中的卡通图像的图像特征, $\text{encoder}_A(\cdot)$ 表示提取括号中的真实的图像特征;

[0017] $\text{decoder}_B(\cdot)$ 表示生成器 G_B 生成的真实人脸图像, $\text{decoder}_A(\cdot)$ 表示生成器 G_A 生成的卡通人脸图像;

[0018] $E_{\zeta \sim \text{pdata}(\zeta)} []$ 表示取均值,下标 $\zeta \sim \text{pdata}(\zeta)$ 用于表示图像 ζ 所属的图像类型,其中 $\zeta \in \{x, y\}$, $\text{pdata}(x)$ 表示真实图像, $\text{pdata}(y)$ 表示卡通图像;

[0019] 步骤S2:对待处理的真实人脸图像进行交互式分割处理,获取眉眼、嘴鼻、头发和人脸轮廓图像,再将眉眼、嘴鼻、头发分别输入对应的生成模型,输出对应的卡通图像:卡通眉眼、卡通嘴鼻和卡通头发;

[0020] 对人脸轮廓图像,取脸部肤色的平均像素值进行填充,得到卡通人脸;

[0021] 最后,将生成的卡通眉眼、卡通嘴鼻合成在卡通人脸上,并叠加上卡通头发,得到最终的卡通图像。

[0022] 本发明中的交互式分割处理具体为:

[0023] 用不同颜色的线条分别标注头发区域、背景区域,用不同颜色的点标注双眼位置以及嘴唇位置,其中双眼和嘴唇的颜色相同,但与头发、背景的颜色不同;

[0024] 以头发的标注颜色的该色彩区域作为前景,其他区域作为背景,分割出头发图像;

[0025] 以双眼的标注颜色的该色彩点的相连区域作为前景,其他区域作为背景,分割出人脸部分,从而得到人脸轮廓图像(脸部图像);

[0026] 再通过双眼和嘴唇标注点剪切出左右眼眉图像和嘴唇图像。

[0027] 综上所述,由于采用了上述技术方案,本发明的有益效果是:本发明通过同时利用交互式和生成对抗网络的优势,通过交互式分割得到人物头发、脸型与人脸五官,消除训练

样本之间由于背景不同导致的差异性,再通过生成对抗网络对各个部分进行风格转化,尽可能多地保留眼角、嘴角等细节部分的信息。

附图说明

[0028] 图1:本发明流程示意图。

[0029] 图2:本发明中生成对抗网络结构图。

具体实施方式

[0030] 为使本发明的目的、技术方案和优点更加清楚,下面结合实施方式和附图,对本发明作进一步地详细描述。

[0031] 本发明提出一种基于生成对抗网络的交互式人脸卡通化方法,同时利用交互式 and 生成对抗网络的优势,通过交互式分割得到人物头发、脸型与人脸五官,消除训练样本之间由于背景不同导致的差异性,再通过生成对抗网络对各个部分进行风格转化,尽可能多地保留眼角、嘴角等细节部分的信息,最后合成为人脸卡通化。

[0032] 首先,构建卡通数据库,包括:眼眉数据库(可只构建左或者右,再将其翻转得到对称的眉眼数据库)、嘴鼻数据库、头发数据库。

[0033] 本具体实施中,为了追求更好的效果,将专业画师为100位志愿者所绘画的同一种风格的卡通头像作为的对应人脸图像的卡通映射图。

[0034] 首先将所有人脸图像与卡通头像图片裁剪到同一尺寸,例如256*256。然后利用经典交互式分割方法对所有图像进行分割,同时标注头发、背景、双眼,具体方法如下:

[0035] 1、用不同颜色的粗线条分别标注头发区域(蓝色)、背景区域(绿色),用圆点标注双眼位置以及嘴唇位置(红点)。

[0036] 2、以蓝色区域作为前景,其他区域作为背景,分割出头发。

[0037] 3、以红点相连区域作为前景,其他区域作为背景,分割出人脸部分,从而得到人脸轮廓,并取脸部肤色的平均像素值进行填充,作为后续卡通化的脸部。

[0038] 4、通过红点位置剪切出左右眼眉和嘴唇部分。

[0039] 5、将以上得到的部分分别构成以下数据集:眼眉数据库(将右眼翻转)、嘴鼻数据库、头发数据库,用于后续的训练。

[0040] 得到数据库之后,利用生成对抗网络实现卡通画的转换。针对以上三个数据库设置三个生成对抗网络,分别记为: G_{eye} 、 G_{mouth} 、 G_{hair} 。每个网络训练过程一致,以训练眼眉图为例,具体方法如下:

[0041] 生成对抗网络一般由一个生成器G和一个判别器D构成,本发明的网络结构中包括两个生成器与两个判别器,分别实现人脸到卡通的转换和卡通到人脸的转换;生成器采用编码-解码(Encoder-Decoder)的结构。网络结构图如图2所示。网络的意义可以解读为:对于两类风格图像A与B,Encoder A表示将提取A类图像的特征,Decoder A表示将特征转成A类图像,Encoder B表示将提取B类图像的特征,Decoder B表示将特征生成B类图像。

[0042] 首先将眼眉图 x 输入生成对抗网络中,通过EncoderA->Decoder B(G_B)得到生成图像fake_y,将生成的图像y与真实卡通图像同时输入判别器 D_B 中,如果是真实图像则输出1,反之输出为0,据此设计损失函数如下,最小化损失函数反向传播更新 D_B 参数。其中 $G_B(x)$

即为fake_y,表示生成器生成的卡通图像,G_B(y)即为fake_x,表示生成器生成的真实图像,E[]表示取均值,下标用于区分对象所属的图像类型。

$$\begin{aligned} \text{[0043]} \quad \text{loss}_{D_B} = & E_{x \sim P_{data}(x)} \left[\log \left(1 - D_B(G_B(x)) \right) \right] \\ & + E_{y \sim P_{data}(y)} \left[\log(D_B(y)) \right] \end{aligned} \quad (1)$$

[0044] 同理,将真实卡通图像y输入生成对抗网络,通过Encoder B->Decoder A(G_A)得到生成图像fake_x,将生成的图像与真实眼眉同时输入判别器D_A中计算损失函数,最小化损失函数反向传播更新D_A参数。损失函数如下:(其中G_A(y)即为fake_x,表示生成器生成的真实眼眉图像。其中G_A(x)即为fake_y,表示生成器生成的卡通眼眉图像。)

$$\begin{aligned} \text{[0045]} \quad \text{loss}_{D_A} = & E_{y \sim P_{data}(y)} \left[\log \left(1 - D_A(G_A(y)) \right) \right] \\ & + E_{x \sim P_{data}(x)} \left[\log(D_A(x)) \right] \end{aligned} \quad (2)$$

[0046] 从公式可以得到,每次迭代更新时,将训练样本中的数据当作正样本,而生成的图像当作负样本,最后为了达到一种平衡,负样本可以尽可能的与正样本相似,但是最后生成的图像尽管整体与真实图像非常相似,但容易非常的模糊,导致判别器仍然无法正确分辨,因此本发明在训练过程中保留上一次生成的图像,记为fake',每一个epoch(每次训练量)更新它的数据,同时,计入下一个epoch的损失函数中:(其中D表示D_A或D_B)

$$\text{[0047]} \quad \text{loss}'_D = \text{loss}_D + E[\log(\text{fake}')] \quad (3)$$

[0048] 对于生成器而言,是根据判别器的输出来进行训练更新,为了使生成的图像尽可能地接近真实图像,应该把生成的图像当作正样本来训练,因此生成器的损失函数可以表示为:

$$\text{[0049]} \quad \text{loss}_{G_A} = E_{y \sim P_{data}(y)} \left[\log \left(D_A(G_A(y)) \right) \right] \quad (4)$$

$$\text{[0050]} \quad \text{loss}_{G_B} = E_{x \sim P_{data}(x)} \left[\log \left(D_B(G_B(x)) \right) \right] \quad (5)$$

[0051] 为了使生成的图像更加接近真实图像,本发明也使用了cycleGAN中的循环损失函数:

$$\begin{aligned} \text{[0052]} \quad L_{\text{cyc}}(G_A, G_B) = & E_{y \sim P_{data}(y)} \left[\|G_B(G_A(y)) - y\| \right] \\ \text{[0053]} \quad & + E_{x \sim P_{data}(x)} \left[\|G_A(G_B(x)) - x\| \right] \end{aligned} \quad (6)$$

[0054] 其中,循环损失函数具体可参考文献《Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks》。

[0055] 同时本发明增加了重构损失,也就是说任何一张图像,经过编码器之后,能够利用该类的解码器还原为原始图像,来证明特征的有用性。损失函数为:

$$\begin{aligned} \text{[0056]} \quad L_{\text{idt}} = & E_{y \sim P_{data}(y)} \left[\| \text{decoder}_B(\text{encoder}_B(y)) - y \| \right] \\ \text{[0057]} \quad & + E_{x \sim P_{data}(x)} \left[\| \text{decoder}_A(\text{encoder}_A(x)) - x \| \right] \end{aligned} \quad (7)$$

[0058] 其中,encoder_B(•)表示提取括号中的卡通图像的图像特征,encoder_A(•)表示提取括号中的真实的图像特征;decoder_B(•)表示生成器G_B生成的真实人脸图像,decoder_A(•)表示生成器G_A生成的卡通人脸图像;即decoder_B(encoder_B(y))表示原始图像y通过该

类的解码器再通过该类的编码器得到的图像,应与原始图像尽可能的一致,即尽可能的还原出原始图像。

[0059] 因此,最终的生成器损失可以表示为:(其中 λ_1, λ_2 为训练设置的参数)

$$L_G = loss_{G_A} + loss_{G_B} + \lambda_1 L_{cyc}(G_A, G_B) + \lambda_2 L_{idt} \quad (8)$$

[0061] 基于本发明构建的三个卡通生成模型,实时对待处理的人脸图像进行人脸卡通化处理得三个部分的卡通图像:眉眼、嘴鼻、头发的卡通图像之后,根据定位进行合成,得到最终的卡通图像。

[0062] 实施例

[0063] 步骤S1:构建训练集。

[0064] 步骤S101:对原始数据进行裁剪成统一尺寸256*256;

[0065] 步骤S102:进行交互式分割处理,用蓝色粗线标记头发区域、绿色粗线标记背景区域,用红点表示左眼、右眼、嘴唇;

[0066] 步骤S103:根据红点坐标裁剪眼眉图像与嘴唇图像(正方形剪裁),并记录左上坐标位置,与剪切图像大小。并尺寸归一化到64*64,构成训练数据库;

[0067] 步骤S104:根据脸部分割图像得到脸部轮廓,并进行肤色填充,作为卡通人脸图像;。

[0068] 步骤S105:将分割的头发图像背景填充为白色,构成头发训练集;

[0069] 步骤S2:训生成对抗网络,得到三个生成模型。

[0070] 分别用上述三个数据训练各自对应的生成对抗网络。得到三个生成模型。

[0071] 步骤S3:对待处理的图像,首先行交互式分割处理,得到眉眼、嘴鼻、头发和脸部图像,再将眉眼、嘴鼻、头发分别输入三个对应的生成模型,输出对应的卡通五官图像;基于脸部图像的卡通化处理,直接得到卡通人脸;

[0072] 再将生成的卡通图像还原成裁剪时的尺寸,并根据坐标位置,将人脸五官合成在卡通人脸上,并叠加上头发效果,得到最终的卡通图像。

[0073] 以上所述,仅为本发明的具体实施方式,本说明书中所公开的任一特征,除非特别叙述,均可被其他等效或具有类似目的的替代特征加以替换;所公开的所有特征、或所有方法或过程中的步骤,除了互相排斥的特征和/或步骤以外,均可以任何方式组合。

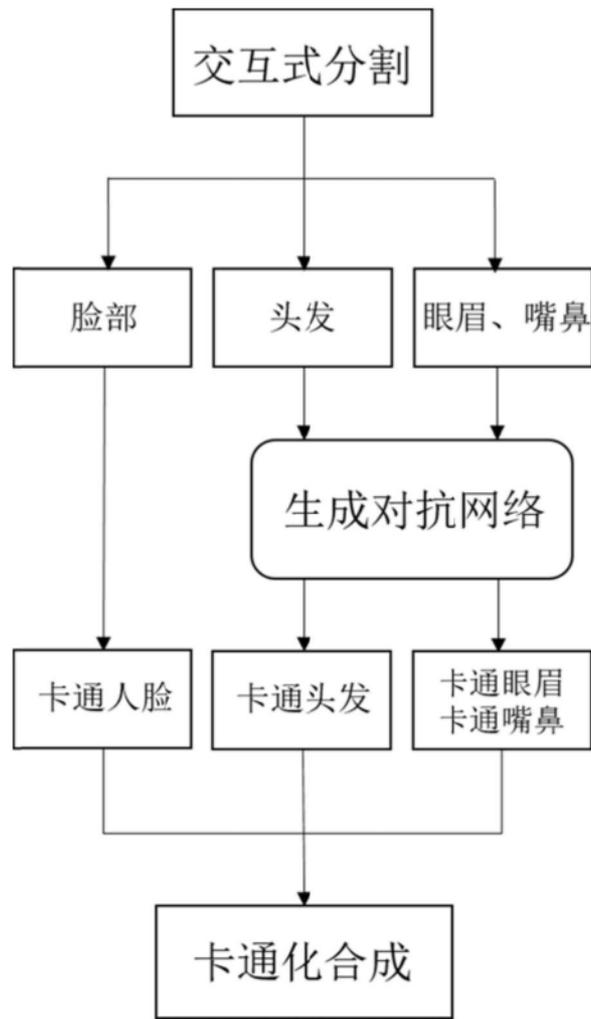


图1

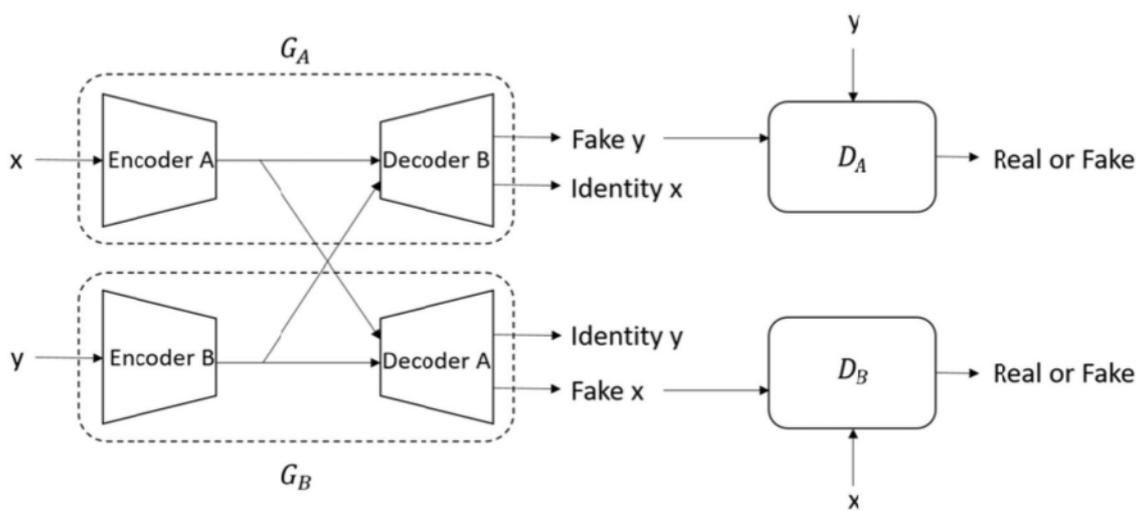


图2