



(12) 发明专利

(10) 授权公告号 CN 114495118 B

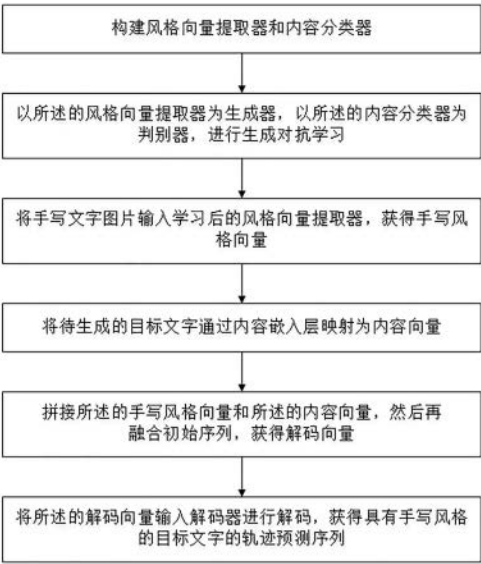
(45) 授权公告日 2022. 08. 09

(21) 申请号 202210392274.4
(22) 申请日 2022.04.15
(65) 同一申请的已公布的文献号
 申请公布号 CN 114495118 A
(43) 申请公布日 2022.05.13
(73) 专利权人 华南理工大学
 地址 510000 广东省广州市天河区五山路
 381号
 专利权人 人工智能与数字经济广东省实验
 室(广州)
(72) 发明人 黄双萍 王庆丰 代港
(74) 专利代理机构 东莞卓诚专利代理事务所
 (普通合伙) 44754
 专利代理师 朱鹏

(51) Int.Cl.
 G06V 30/32 (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)
 G06T 11/20 (2006.01)
(56) 对比文件
 US 2020151938 A1,2020.05.14
 CN 106803082 A,2017.06.06
 审查员 刘志军

权利要求书3页 说明书6页 附图3页

(54) 发明名称
 基于对抗解耦的个性化手写文字生成方法
(57) 摘要
 本发明公开了基于对抗解耦的个性化手写文字生成方法,包括以下步骤:构建风格向量提取器和内容分类器;以所述的风格向量提取器为生成器,以所述的内容分类器为判别器,进行生成对抗学习;将手写文字图片输入学习后的风格向量提取器,获得手写风格向量;将待生成的目标文字通过内容嵌入层映射为内容向量;拼接所述的手写风格向量和所述的内容向量,然后在融合初始序列,获得解码向量;将所述的解码向量输入解码器进行解码,获得具有手写风格的目标文字的轨迹预测序列。本发明方法能够通过对抗解耦的方式,利用对抗网络帮助风格提取器消除风格向量中夹杂的内容信息,提取出干净的风格向量,进而提升个性化手写文字的生成效果。



CN 114495118 B

1. 基于对抗解耦的个性化手写文字生成方法, 其特征在于, 包括以下步骤:

步骤1, 构建风格向量提取器和内容分类器;

步骤2, 以所述的风格向量提取器为生成器, 以所述的内容分类器为判别器, 进行生成对抗学习;

步骤3, 将手写文字图片输入学习后的风格向量提取器, 获得手写风格向量 V_s ;

步骤4, 将待生成的目标文字通过内容嵌入层映射为内容向量 V_c ;

步骤5, 拼接所述的手写风格向量和所述的内容向量, 然后再融合初始序列 Seq , 获得解码向量 V_D , 即 $V_D = V_s \oplus V_c \oplus Seq$, 其中 $\oplus$ 代表在第二维上拼接, 初始序列 Seq 为: 在训练时为手写文字图片对应的真实轨迹序列, 在测试时为解码器预测的轨迹序列;

步骤6, 将所述的解码向量输入解码器进行解码, 获得具有手写风格的目标文字的轨迹预测序列;

步骤2中所述的对抗学习的过程中, 所述的生成器的损失函数为:

$$G_loss = dx_loss + dy_loss + state_loss - eps \times D_loss$$

其中, dx_loss 表示 x 坐标损失函数, dy_loss 表示 y 坐标损失函数, $state_loss$ 表示状态损失函数, D_loss 表示判别器损失函数, eps 表示判别器损失函数的权重;

所述的 x 坐标损失函数的计算公式为:

$$dx_loss = \frac{\sum_{i=0}^n (|x_{gt_i} - x_{pred_i}|)}{2n}$$

其中, x_{gt_i} 表示文字对应的轨迹序列的真实 x 坐标, x_{pred_i} 表示解码器预测轨迹序列的 x 坐标, n 表示序列点长度, $||$ 表示绝对值;

所述的 y 坐标损失函数的计算公式为:

$$dy_loss = \frac{\sum_{i=0}^n (|y_{gt_i} - y_{pred_i}|)}{2n}$$

其中, y_{gt_i} 表示文字对应的轨迹序列的真实 y 坐标, y_{pred_i} 表示解码器预测轨迹序列的 y 坐标;

所述的状态损失函数的计算公式为:

$$state_loss = - \sum_{i=1}^k y_i \log \bar{y}_i$$

其中, y_i 为第*i*个真实状态标签的onehot编码, $\bar{y}_i$ 为解码器输出的经过softmax的第*i*个概率值, $k=3$, 表示三个状态, 每个轨迹点对应三个状态为: 下笔、抬笔、起笔。

2. 根据权利要求1所述的基于对抗解耦的个性化手写文字生成方法, 其特征在于, 步骤2中所述的对抗学习的过程为: 交叉多轮训练判别器和生成器; 每轮训练过程中, 判别器的训练频次大于生成器的训练频次; 训练判别器的时候仅更新判别器的参数, 训练生成器的时候更新除判别器以外的模型的参数。

3. 根据权利要求2所述的基于对抗解耦的个性化手写文字生成方法, 其特征在于, 所述的判别器损失函数的计算公式为:

$$D_loss = - \sum_{i=1}^n c_i \log \bar{c}_i$$

其中, c_i 为第*i*个真实内容标签的onehot编码, $\bar{c}_i$ 为判别器输出的经过softmax的第*i*个概率值;

所述的判别器损失函数的权重 eps 的计算公式为:

$$eps = \begin{cases} \delta * step / all_step, & step < all_step \\ \delta, & step \geq all_step \end{cases}$$

其中, $\delta=1e-5$, $all_step=10000$, $step$ 为当前训练轮次。

4. 根据权利要求1所述的基于对抗解耦的个性化手写文字生成方法, 其特征在于, 所述的风格向量提取器的结构为M1层CNN网络和N1层全连接层构成, 所述的风格向量提取器的输出为Batchsize个128维的风格向量, 其中, M1和N1分别表示CNN网络和全连接层的层数, Batchsize表示一次训练所抓取的数据样本数量。

5. 根据权利要求4所述的基于对抗解耦的个性化手写文字生成方法, 其特征在于, M1取值为7, N1取值为2。

6. 根据权利要求4或5所述的基于对抗解耦的个性化手写文字生成方法, 其特征在于, 所述的风格向量提取器为经过预训练的风格向量提取器, 预训练采用的数据集为Casia1.1数据集, 预训练过程中采用风格分类任务训练, 所述的风格向量提取器的最后一层全连接层的数量根据手写风格的类别数量而确定。

7. 根据权利要求1所述的基于对抗解耦的个性化手写文字生成方法, 其特征在于, 所述的内容分类器的结构为M2层CNN网络和N2层全连接层构成, 所述的内容分类器的任务为文字分类, 即判别输入的手写文字是哪个字, 其中M2和N2分别为CNN网络和全连接层的层数。

8. 根据权利要求7所述的基于对抗解耦的个性化手写文字生成方法, 其特征在于, M2取值为6, N2取值为1。

9. 根据权利要求8所述的基于对抗解耦的个性化手写文字生成方法,其特征在于,所述的解码器的结构为5层LSTM网络。

基于对抗解耦的个性化手写文字生成方法

技术领域

[0001] 本发明涉及生成对抗神经网络技术领域,尤其涉及基于对抗解耦的个性化手写文字生成方法。

背景技术

[0002] 手写文字是人们展示自我的信息载体,属于人的生物特征。手写文字具有商业应用价值,高效的手写文字生成方法可以大幅减少字体设计师的成本。同时,手写文字趣味性十足,能满足人们的个性化需求。

[0003] 近年来,随着深度学习领域的发展,手写文字生成领域也得到了一定的发展。在字母型语言中,手写文字生成领域已经提出了较多方法,但是由于中文庞大的常用字典,以及汉字复杂的拓扑结构,中文的手写文字生成还未能达到非常好的效果。因此,如何能够有效地提高中文手写文字的生成效果,仍是待研究的问题。

[0004] 现有的手写文字生成技术基本上都是基于风格提取的方法,即使用CNN网络单独对手写者的少量手写文字图片进行风格向量的提取,后续再对此向量做进一步的融合利用。但是仅仅用CNN网络提取风格向量并不能保证提取出的向量仅包含风格信息。提取出的风格信息中可能夹杂了部分内容信息,这会干扰后续对风格信息的利用。因此,如果能够消除其中的内容信息,提取出干净的风格信息,将会对后续的生成步骤带来很大的提升。

发明内容

[0005] 有鉴于此,有必要针对上述技术问题,提供基于对抗解耦的个性化手写文字生成方法,所述方法通过书写者的少量手写文字图片,利用生成对抗网络帮助风格提取器对内容信息和风格信息进行对抗解耦,再将风格提取器提取出的干净的风格信息和输入的内容信息进行融合,送入解码器,最终生成目标文字更加精准的在线轨迹序列。

[0006] 基于对抗解耦的个性化手写文字生成方法,包括以下步骤:

[0007] 步骤1,构建风格向量提取器和内容分类器;

[0008] 步骤2,以所述的风格向量提取器为生成器,以所述的内容分类器为判别器,进行生成对抗学习;

[0009] 步骤3,将手写文字图片输入学习后的风格向量提取器,获得手写风格向量 V_s ;

[0010] 步骤4,将待生成的目标文字通过内容嵌入层映射为内容向量 V_c ;

[0011] 步骤5,拼接所述的手写风格向量和所述的内容向量,然后再融合初始序列 Seq ,获得解码向量 V_D ,即 $V_D = V_s \oplus V_c \oplus Seq$,其中 $\oplus$ 代表在第二维上拼接,初始序列 Seq 为:在训练时为手写文字图片对应的真实轨迹序列,在测试时为解码器预测的轨迹序列;

[0012] 步骤6,将所述的解码向量输入解码器进行解码,获得具有手写风格的目标文字的

轨迹预测序列。

[0013] 具体地,步骤2中所述的对抗学习的过程为:

[0014] 交叉多轮训练判别器和生成器;每轮训练过程中,判别器的训练频次大于生成器的训练频次;训练判别器的时候仅更新判别器的参数,训练生成器的时候更新除判别器以外的模型的参数。

[0015] 优选地,步骤2中所述的对抗学习的过程中,所述的生成器的损失函数为:

$$[0016] \quad G_loss = dx_loss + dy_loss + state_loss - eps \times D_loss$$

其中, dx_loss 表示 x 坐标损失函数, dy_loss 表示 y 坐标损失函数, $state_loss$ 表示状态损失函数, D_loss 表示判别器损失函数, eps 表示判别器损失函数的权重;

[0017] 所述的 x 坐标损失函数的计算公式为:

$$[0018] \quad dx_loss = \frac{\sum_{i=0}^n (|x_{gt_i} - x_{pred_i}|)}{2n}$$

[0019] 其中, x_{gt_i} 表示文字对应的轨迹序列的真实 x 坐标, x_{pred_i} 表示解码器预测轨迹序列的 x 坐标, n 表示序列点长度;

[0020] 所述的 y 坐标损失函数的计算公式为:

$$[0021] \quad dy_loss = \frac{\sum_{i=0}^n (|y_{gt_i} - y_{pred_i}|)}{2n}$$

[0022] 其中, y_{gt_i} 表示文字对应的轨迹序列的真实 y 坐标, y_{pred_i} 表示解码器预测轨迹序列的 y 坐标;

[0023] 所述的状态损失函数的计算公式为:

$$[0024] \quad state_loss = - \sum_{i=1}^k y_i \log \bar{y}_i$$

[0025] 其中, y_i 为第 i 个真实状态标签的 onehot 编码, $\bar{y}_i$ 为解码器输出的经过 softmax 的第 i 个概率值, $k=3$, 表示三个状态, 每个轨迹点对应三个状态为: 下笔、抬笔、起笔;

[0026] 所述的判别器损失函数的计算公式为:

$$[0027] \quad D_loss = - \sum_{i=1}^n c_i \log \bar{c}_i$$

[0028] 其中, C_i 为第 i 个真实内容标签的 onehot 编码, $\bar{C}_i$ 为判别器输出的经过 softmax 的第 i 个概率值;

[0029] 所述的判别器损失函数的权重 eps 的计算公式为:

$$[0030] \quad eps = \begin{cases} \delta * step / all_step, & step < all_step; \\ \delta, & step \geq all_step. \end{cases}$$

[0031] 其中, $\delta = 1e-5$, $all_step = 10000$, $step$ 为当前训练轮次。

[0032] 具体地, 所述的风格向量提取器的结构为 $M1$ 层 CNN 网络和 $N1$ 层全连接层构成, 所述的风格向量提取器的输出为 $Batchsize$ 个 128 维的风格向量, 其中, $M1$ 和 $N1$ 分别表示 CNN 网络 and 全连接层的层数, $Batchsize$ 表示一次训练所抓取的数据样本数量;

[0033] 优选地, $M1$ 取值为 7, $N1$ 取值为 2;

[0034] 更进一步地, 所述的风格向量提取器为经过预训练的风格向量提取器, 预训练采用的数据集为 Casial.1 数据集, 预训练过程中采用风格分类任务训练, 所述的风格向量提取器的最后一层全连接层的数量根据手写风格的类别数量而确定。

[0035] 具体地, 所述的内容分类器的结构为 $M2$ 层 CNN 网络和 $N2$ 层全连接层构成, 所述的内容分类器的任务为文字分类, 即判别输入的手写文字是哪个字, 其中 $M2$ 和 $N2$ 分别为 CNN 网络 and 全连接层的层数;

[0036] 优选地, $M2$ 值为 6, $N2$ 为 1;

[0037] 具体地, 所述的解码器的结构为 5 层 LSTM 网络。

[0038] 优选地, 所述的手写风格向量 V_S 为 128 维向量, 内容向量 V_C 为 128 维向量, 初始序列 Seq 为 5 维向量。

[0039] 具体地, 所述的轨迹预测序列指的是输出的手写文字的相对坐标序列, 所述的相对坐标序列包括 x 、 y 坐标, 以及每个点对应的状态, 所述的状态包括下笔状态、起笔状态和抬笔状态。

[0040] 与现有技术相比, 本发明的有益效果在于, 本发明方法能够通过对抗解耦的方式, 利用对抗网络帮助风格提取器消除风格向量中夹杂的内容信息, 提取出干净的风格向量, 进而提升个性化手写文字的生成效果。

附图说明

[0041] 图 1 示出了本发明实施方法的流程示意图;

[0042] 图 2 示出了本发明实施例的生成对抗学习的过程示意图;

[0043] 图 3 示出了本发明实施例的手写文字生成的流程示意图。

具体实施方式

[0044] 为了使本发明的目的、技术方案和优点更加清楚, 下面将结合附图对本发明作进一步地详细描述, 显然, 所描述的实施例仅仅是本发明一部份实施例, 而不是全部的实施

例。基于本发明中的实施例，本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其它实施例，都属于本发明保护的范围。应当理解，此处所描述的具体实施例仅仅用以解释本发明，并不用于限定本发明。

[0045] 图1示出了本发明实施例的流程示意图。基于对抗解耦的个性化手写文字生成方法，包括以下步骤：

[0046] 步骤1，构建风格向量提取器和内容分类器；

[0047] 步骤2，以所述的风格向量提取器为生成器，以所述的内容分类器为判别器，进行生成对抗学习；

[0048] 步骤3，将手写文字图片输入学习后的风格向量提取器，获得手写风格向量 V_S ；

[0049] 步骤4，将待生成的目标文字通过内容嵌入层映射为内容向量 V_C ；

[0050] 步骤5，拼接所述的手写风格向量和所述的内容向量，然后再融合初始序列 Seq ，获得解码向量 V_D ，即 $V_D = V_S \oplus V_C \oplus Seq$ ，其中 $\oplus$ 代表在第二维上拼接，初始序列 Seq 为：在训练时为手写文字图片对应的真实轨迹序列，在测试时为解码器预测的轨迹序列；

[0051] 步骤6，将所述的解码向量输入解码器进行解码，获得具有手写风格的目标文字的轨迹预测序列。

[0052] 具体地，在本实施例中采用以下方法步骤：

[0053] (1) 使用Casial.1数据集预训练风格向量提取器，将最后一层全连接层改为300，用于300个书写者的风格分类任务，借用风格分类任务对风格向量提取器进行预训练。

[0054] (2) 加载预训练的风格向量提取器权重(除全连接层)，如图2所示，以内容分类器为判别器D，风格向量提取器为生成器G，进行生成对抗学习。具体为：交叉训练判别器和生成器，训练5次D，训练1次G。训练D时仅更新判别器的参数，训练G时更新除判别器以外的模型的参数。

[0055] 具体的损失函数设置如下：

[0056] (a) 坐标损失函数， x_{gt_i}, y_{gt_i} 为文字对应的轨迹序列的第i个真实x、y坐标， x_{pred_i}, y_{pred_i} 则分别是解码器预测轨迹序列的第i个x、y坐标

$$[0057] \quad dx_loss = \frac{\sum_{i=0}^n (|x_{gt_i} - x_{pred_i}|)}{2n}$$

$$[0058] \quad dy_loss = \frac{\sum_{i=0}^n (|y_{gt_i} - y_{pred_i}|)}{2n}$$

[0059] (b) 状态损失函数，每个轨迹点对应三个状态下笔、抬笔、起笔。这三个状态的损失函数使用交叉熵损失函数来衡量。 y_i 为第i个真实状态标签的onehot编码， $\bar{y}_i$ 为解码器输

出的经过softmax的第i个概率值。

$$[0060] \quad state_loss = - \sum_{i=1}^n y_i \log \bar{y}_i$$

[0061] (c)判别器损失函数,由于判别器是做内容分类,故直接使用交叉熵损失函数。 C_i 为真实内容标签的onehot编码, $\bar{c}_i$ 为判别器输出的经过softmax的概率值。

$$[0062] \quad D_loss = - \sum_{i=1}^n c_i \log \bar{c}_i$$

[0063] (d)生成器损失函数,由坐标损失函数和状态损失函数、判别器损失函数加权得到,eps为D_loss的权重,具体为 $eps = \delta * step / all_step, \delta = 1e-5, all_step = 10000$,step为当前训练轮次。

$$[0064] \quad G_loss = dx_loss + dy_loss + state_loss - eps \times D_loss$$

[0065] (3)当生成文字时,如图2所示,需要通过扫描或拍照,将书写者的少量手写文字图片输入风格向量提取器,通过这部分文字图片提取出书写者的128维手写风格向量 V_s 。

[0066] (4)输入要生成的目标文字,并根据其在字典中的ID将其通过嵌入层映射为128维的内容向量 V_c 。

[0067] (5)将前两步提取的风格向量和内容向量通过拼接的方式融合为256维的向量,再将融合向量和初始序列 Seq 融合,得到261维的解码向量 V_D 。即, $V_D = V_s \oplus V_c \oplus Seq$ 。其中 $\oplus$ 代表在第二维上拼接。初始序列 Seq 为:在训练时为手写文字图片对应的真实轨迹序列,在测试时为解码器预测的轨迹序列。

[0068] (6)将解码向量输入解码器进行解码,最终解码得到具有书写者手写风格的目标文字的在线轨迹序列。本实施例方法的简易流程图如图3所示。

[0069] 本发明方法的技术原理是:

[0070] 本发明利用内容分类器作为判别器,与风格编码器进行对抗,进而使得编码器在对抗过程中逐步消除内容信息,进而提取出干净的风格信息,实现内容和风格的解耦。通过该对抗解耦过程,本发明解决了在现有方法中风格编码器存在内容干扰的问题,进而增强了最终手写文字的生成效果,使得生成的文字更加多样化,更符合书写者的风格特性。

[0071] 以上实施例的各技术特征可以进行任意的组合,为使描述简洁,未对上述实施例中的各个技术特征所有可能的组合都进行描述,然而,只要这些技术特征的组合不存在矛盾,都应当认为是本说明书记载的范围。

[0072] 本领域技术人员应明白,本申请的实施例可提供为方法、系统或计算机程序产品。因此,本申请可采用完全硬件实施例、完全软件实施例或结合软件和硬件方面的实施例的

形式。而且,本申请可采用在一个或多个其中包含有计算机可用程序代码的计算机可用存储介质(包括但不限于磁盘存储器、CD-ROM、光学存储器等)上实施的计算机程序产品的形式。

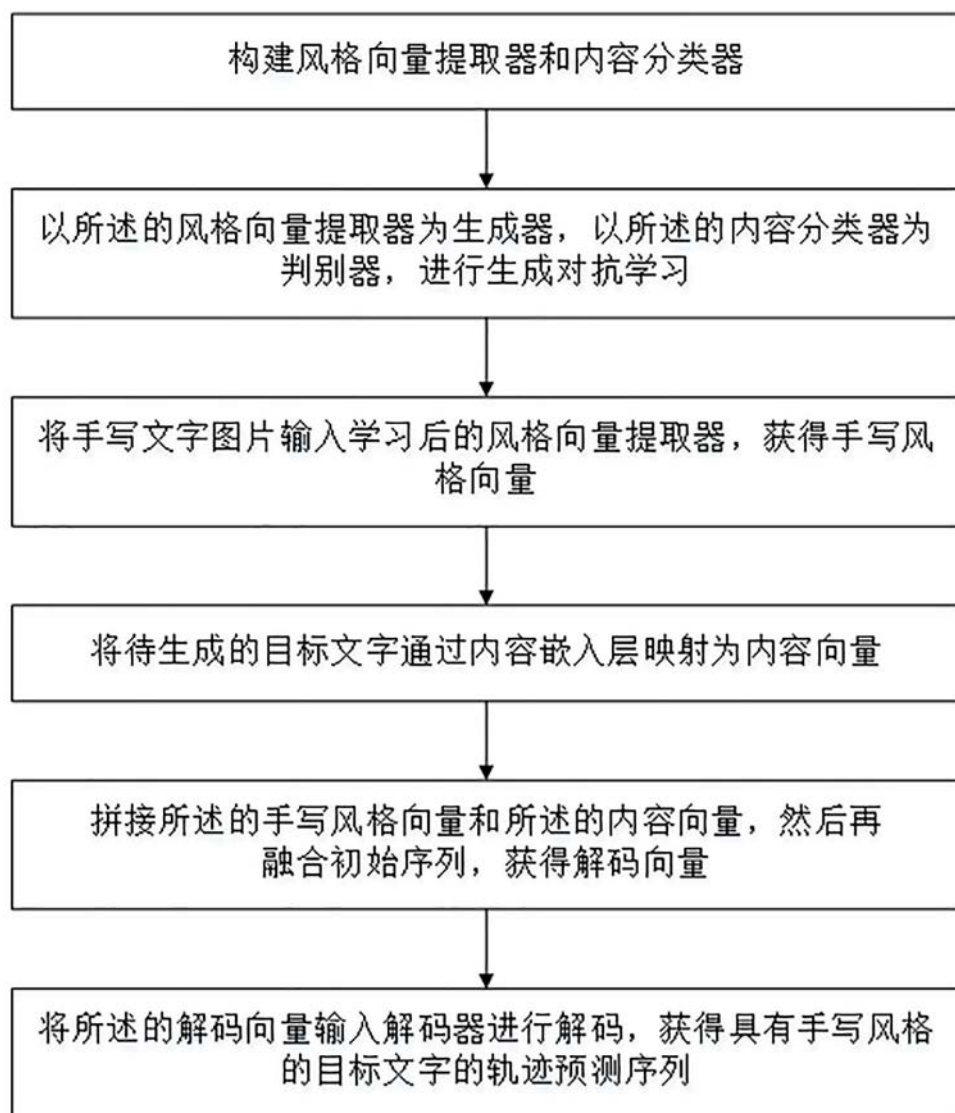


图1

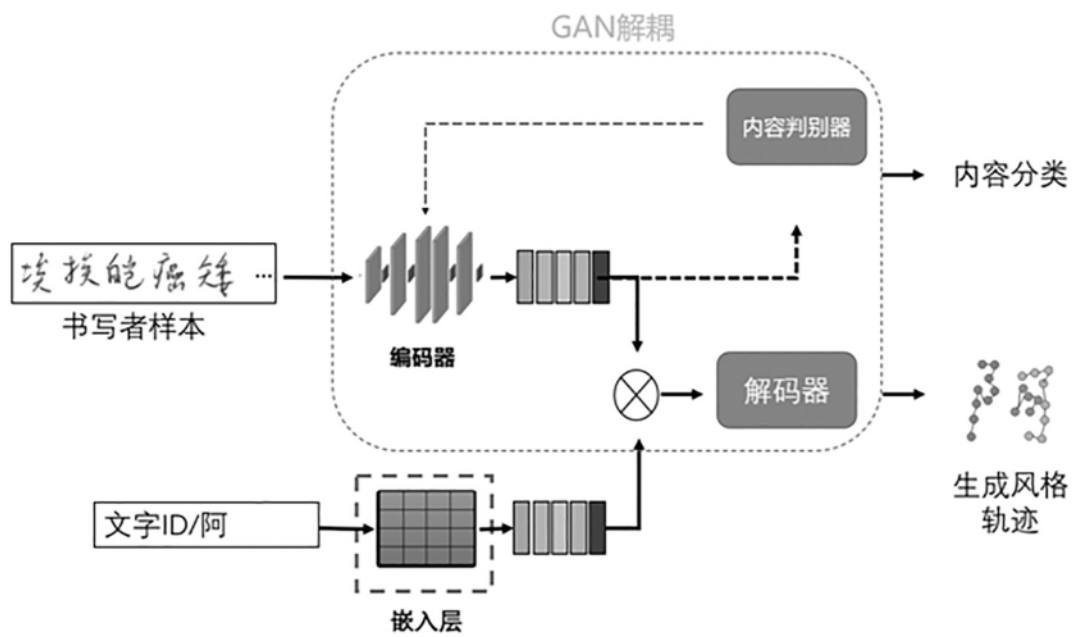


图2

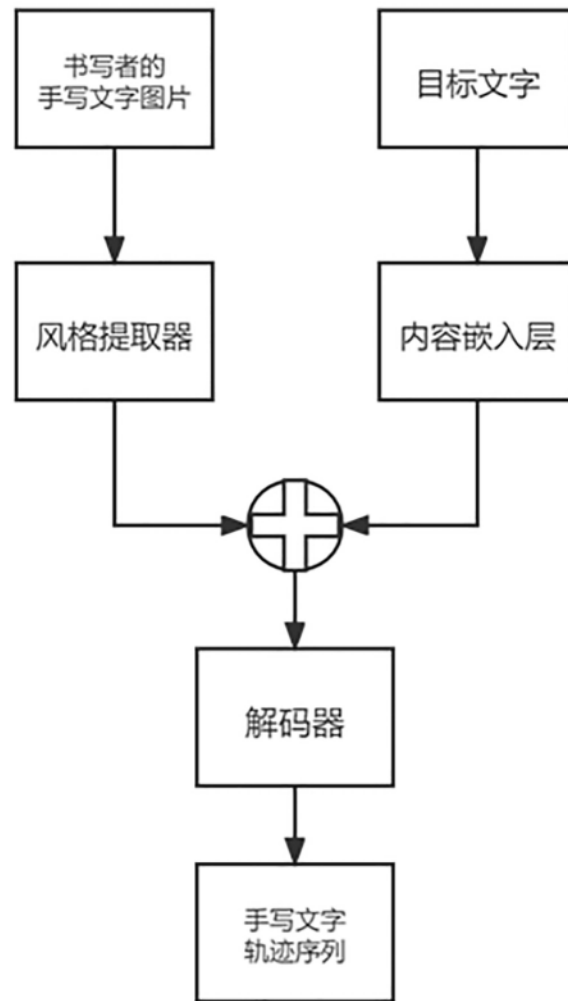


图3