(19)国家知识产权局

(12)发明专利

(10)授权公告号 CN 114782961 B
(45)授权公告日 2023.04.18

(21)申请号 202210285238.8
(22)申请日 2022.03.23
(65)同一申请的已公布的文献号
申请公布号 CN 114782961 A
(43)申请公布日 2022.07.22
(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
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 黄鸿翔 杨代辉
(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 30/19(2022.01)
G06T 3/00(2006.01)
G06N 3/0464(2023.01)
G06N 3/048(2023.01)
G06N 3/094(2023.01)
(56)对比文件
CN 108399408 A,2018.08.14
CN 111652332 A,2020.09.11
审查员 王杰

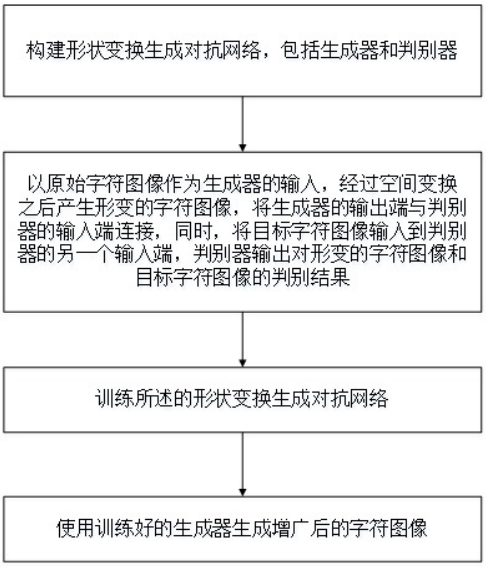
权利要求书3页 说明书9页 附图1页

(54)发明名称

一种基于形状变换的字符图像增广方法

(57)摘要

本发明公开了一种基于形状变换的字符图像增广方法,包括以下步骤:构建形状变换生成对抗网络,包括生成器和判别器;以原始字符图像作为生成器的输入,经过空间变换之后产生形变的字符图像,将生成器的输出端与判别器的输入端连接,同时,将目标字符图像输入到判别器的另一个输入端,判别器输出对形变的字符图像和目标字符图像的判别结果;训练所述的形状变换生成对抗网络;使用训练好的生成器生成增广后的字符图像。本发明方法结合仿射矩阵和TPS变换采样网格参数使STN能够同时产生全局和局部的形状变化,能够更好的拟合字符的形状特征,使产生的字符真实性和多样性更好,使用增广后的数据所训练的分类器的分类性能进一步提升。



CN 114782961 B

1.一种基于形状变换的字符图像增广方法,其特征在于,包括以下步骤:

步骤1,构建形状变换生成对抗网络,包括生成器和判别器;

步骤2,以原始字符图像作为生成器的输入,经过空间变换之后产生形变的字符图像,将生成器的输出端与判别器的输入端连接;将目标字符图像输入到判别器的另一个输入端,判别器输出对形变的字符图像和目标字符图像的判别结果;

步骤3,训练所述的形状变换生成对抗网络;

步骤4,使用训练好的生成器生成增广后的字符图像;

所述的生成器为一个空间变换网络,包括编码器、预测器、采样器、噪声重建网络和图像重建网络;

所述的编码器由多个卷积模块顺序连接,每一个卷积模块包括一个二维卷积层、一个非线性激活层和一个汇聚层顺序连接;

所述的预测器由多个全连接模块和最后一个全连接层顺序连接,每一个全连接模块包括一个全连接层和一个非线性激活层,最后一个全连接层的输出通道数量设置为所需预测的形变参数的个数;

所述的采样器通过在采样网格上应用矩阵乘法将形变的字符图像像素区域映射到原始字符图像像素区域;

所述的图像重建网络由多个全连接模块、一个全连接层和多个转置卷积模块顺序连接,每一个转置卷积模块包含依次连接的一个转置卷积层和一个非线性激活层;

所述的噪声重建网络由多层全连接模块和一个全连接层顺序连接;

所述的判别器基于patchGAN的结构,由依次连接的五个卷积模块组成,前四个卷积模块中每个卷积模块包括一个二维卷积层、一个实例归一化层、一个leakyReLU激活层,最后一个卷积模块包括一个padding层和一个二维卷积层;

首先,原始字符图像作为编码器的输入,编码器从原始字符图像中提取形状特征,输出一个形状特征向量,然后从标准正态分布中随机选取一个噪声隐向量,将形状特征向量和噪声隐向量进行融合,将融合后的隐向量输入到预测器中,预测器负责预测出TPS变换参数和仿射变换参数,其中,TPS变换参数为TPS变换采样网格匹配点的坐标值,将仿射变换参数转化为仿射变换采样网格,接着,将TPS变换采样网格和仿射变换采样网格以及原始字符图像输入到采样器中,输出形变的字符图像,同时,将预测器输出端与图像重建网络和噪声重建网络的输入端连接,分别重建出原始字符图像和噪声隐向量,接着,将生成器输出的形变的字符图像和目标字符图像分别输入到判别器,判别器输出对形变的字符图像和目标字符图像的判别结果。

2.根据权利要求1所述的一种基于形状变换的字符图像增广方法,其特征在于,构建最小二乘生成对抗损失函数作为形状变换生成对抗网络训练的优化目标,最小二乘生成对抗损失函数计算公式为:

$$\mathrm{L}_{\mathrm{G}_g}=\mathbb{E}_x\left[D_g\left(G_g(x)\right)-1\right]^2+\mathrm{L}_{\mathrm{snr}}\left(G_g\right)+\mathrm{L}_{\mathrm{div}}(E,P)$$

$$\mathrm{L}_{\mathrm{D}_g}=\frac{1}{2}\mathbb{E}_x\left[D_g\left(G_g(x)\right)\right]^2+\frac{1}{2}\mathbb{E}_y\left[D_g(y)-1\right]^2$$

L_{G_g}表示生成器损失函数，D_g表示判别器，G_g表示生成器，$L_{snr}(G_g)$表示信噪重建损失函数，$L_{div}(E,P)$表示多样性损失函数，L_{D_g}表示判别器损失函数，x表示原始字符图像，y表示目标字符图像，$\mathbb{E}_x$和$\mathbb{E}_y$表示求相应的数学期望。

3.根据权利要求2所述的一种基于形状变换的字符图像增广方法，其特征在于，所述的信噪重建损失函数包括信号重建子项、噪声重建子项和重建误差比值项，计算公式如下：

$$L_{snr}(G_g) = L_1(x, x_{rec}) + L_1(z, z_{rec}) + \alpha \cdot \log\left(\frac{L_1(z, z_{rec})}{L_1(x, x_{rec})}\right)$$

L_1表示平均绝对误差，z表示噪声隐向量，x_{rec}和z_{rec}分别表示图像重建网络R_x和R_z重建后的原始图像和噪声隐向量，α是动态系数，如果重建误差比值项大于logM，设α＝1，当重建误差比值项小于-logM时，设α＝-1，如果重建误差比值项在理想范围内，即[-logM,logM]，则设置α＝0，M表示超参数；

所述的多样性损失函数的计算公式如下：

$L_{div}(E,P) = -L_1(P(E(x),z_1),P(E(x),z_2))$

其中P表示预测器，E表示编码器，z_1和z_2分别表示取自相同高斯分布的不同噪声隐向量。

4.根据权利要求1所述的一种基于形状变换的字符图像增广方法，其特征在于，编码器中所述的每一个卷积模块还包括一个批归一化层，位于二维卷积层和非线性激活层中间；编码器中所述的非线性激活函数选择ReLU函数，所述的汇聚层的汇聚运算选择最大化汇聚；

预测器中所述的每一个全连接模块还包括一个批归一化层，位于全连接层和非线性激活层中间；预测器中所述的非线性激活函数选择ReLU函数，所述的汇聚层的汇聚运算选择最大化汇聚；

所述的转置卷积模块还包括一个批归一化层，位于转置卷积层和非线性激活层中间；转置卷积模块中所述的非线性激活函数选择ReLU函数，所述的汇聚层的汇聚运算选择最大化汇聚。

5.根据权利要求1所述的一种基于形状变换的字符图像增广方法，其特征在于，预测器中所述的最后一个全连接层的输出通道数量设置为132，所需预测的形变参数中128个参数为64个TPS变换采样网格匹配点的坐标，4个参数为仿射变换矩阵的元素值。

6.根据权利要求1或5所述的一种基于形状变换的字符图像增广方法，其特征在于，TPS变换的目标是求解一个形变函数f，使得$f(p_i(x_i,y_i))=p_i(x_i',y_i')$，$(1\leqslant i\leqslant n)$，且弯曲能量函数最小，$p_i(x_i,y_i)$表示原始字符图像上的TPS变换采样网格匹配点的坐标，$p_i(x_i',y_i')$表示形变的字符图像的TPS变换采样网格匹配点的坐标，n为一个TPS变换采样网格匹配点的个数，假设已经获取到两张图像的n组TPS变换采样网格匹配点坐标对：$(p_1(x_1,y_1),p_1(x_1',y_1'))$、$(p_2(x_2,y_2),p_2(x_2',y_2'))$、...、$(p_n(x_n,y_n),p_n(x_n',y_n'))$，把形变函数想象成弯折一块薄金属板，使这块板穿过给定的n个点，弯曲薄板的能量函数表示为：

$$E(f)=\iint\left(\left(\frac{\partial^2 f}{\partial x^2}\right)^2+2\left(\frac{\partial^2 f}{\partial x\partial y}\right)^2+\left(\frac{\partial^2 f}{\partial y^2}\right)^2\right)dxdy$$

能够证明薄板样条函数就是弯曲能量最小的函数,薄板的样条函数为:

$$f(x,y)=a_1+a_2x+a_3y+\sum_{i=1}^{n}w_i\,U(|p(x_i,y_i)-p(x,y)|)$$

其中U为基函数:

$$U(r_{ij})=\begin{cases}r_{ij}^{\,2}\log r_{ij}^{\,2} & ,r_{ij}\neq 0\\ 0 & ,r_{ij}=0\end{cases}$$

$$r_{ij}=|p_i(x_i,y_i)-p_j(x_j,y_j)|=\sqrt{(x_i-x_j)^2+(y_i-y_j)^2}$$

a_1,a_2,a_3和w_i通过n个TPS变换采样网格匹配点的坐标的预设值和预测器预测的偏移量来求解,由此获得f(x,y)具体表达式。

7.根据权利要求1或5所述的一种基于形状变换的字符图像增广方法,其特征在于,所述的仿射变换采样网格的采样公式如下:

$$\begin{bmatrix}x\\y\\1\end{bmatrix}=\begin{bmatrix}scale\times cos(theta) & -scale\times sin(theta) & t_x\\ scale\times sin(theta) & scale\times cos(theta) & t_y\\ 0 & 0 & 1\end{bmatrix}\begin{bmatrix}x'\\y'\\1\end{bmatrix}$$

其中,scale,theta,t_x,t_y分别表示预测器预测出的仿射变换参数,(x,y)和(x′,y′)分别为变换前后的像素点的位置坐标。

8.根据权利要求1所述的一种基于形状变换的字符图像增广方法,其特征在于,所述的采样器通过pytorch中的torch.nn.functional.grid_sample()方法实现,同时,通过pytorch中的torch.nn.functional.affine_grid()方法将仿射参数转化为仿射变换采样网格。

9.根据权利要求1所述的一种基于形状变换的字符图像增广方法,其特征在于,所有图像的像素尺寸为64*64,批大小为64,初始学习率为0.0001,对抗网络的迭代次数为5000,在2500次迭代后学习率开始线性衰减至$1e^{-5}$,对抗网络采用adam优化器进行优化。

一种基于形状变换的字符图像增广方法

技术领域

[0001] 本发明属于图像处理与人工智能技术领域,特别是涉及一种基于形状变换的字符图像增广方法。

背景技术

[0002] 基于深度学习的文本图像识别方法展现出巨大的潜力,然而,训练一个高性能的文字图像识别模型需要大量且尽可能多样的标注数据。由于人工收集并标注文字图像数据是一件极为昂贵且耗时的工作,这对于古文字、手写文字、手写公式等形状复杂的文字图像更是如此。相比之下,数据增广技术是一种成本较低的增加数据多样性的有效途径。

[0003] 数据增广的方式包括对图像进行翻转、旋转、缩放、裁剪、平移等基于形状变换的增广,以及色彩抖动、噪声注入等基于非形状变换的增广。其中,基于形状变换的增广方式在文字图像领域中属于主流的增广方式,传统的基于形状变换的增广算法主要从人为预设的概率分布中随机地采样形变的参数来控制图像内容的形状变换,例如最常见的做法是采用仿射矩阵(affine matrix)来产生仿射变换,例如基于仿射变换的空间变换网络(spatial transformation network,STN)算法,又或者是通过其他非刚性形变算法,例如薄板样条(thin plate spline,TPS)变换来产生局部形变。然而,这种传统的增广算法依赖人为预设的概率分布进行采样,不仅分布的计算和选取复杂,而且所选定的分布难以完全拟合实际的文字形状的分布,导致人力成本较大,形变的真实性较差。

[0004] 传统的基于形状变换的增广算法主要从人为预设的概率分布中随机地采样形变的参数来控制图像内容的形状变换,不仅分布的计算和选取复杂,而且所选定的分布难以完全拟合实际的文字形状的分布。因此,这种做法既耗费大量的人工成本,产生的形变的真实性也较差。虽然通过生成对抗网络可以解决人工计算选择分布的缺陷,但是这些技术仍存在形变精细度不够的问题,原因在于实际的文字形状同时具有全局(比如旋转角度、平移距离等)和局部(比如笔画的扭曲程度、长度、粗细等)的特征,而现有技术仅仅产生单独的全局形变或局部形变,导致变换后的形状真实性和多样性较差,最终造成增广后的数据对于下游任务的性能提升有限。基于神经网络的增广技术采用生成对抗损失函数作为优化目标,但由于生成对抗损失函数只计算真假两种标签的损失值,仅提供较弱的监督,而文字图像既有全局又有局部的形状特征,为使变换文字的形状真实性和多样性得到保证,需要使损失函数提供更强的监督。例如,有研究指明,生成对抗网络可能会抑制生成器中语义向量或噪声向量的作用,这会导致文字的形变程度过小(甚至没有任何变化)或过大(形状扭曲),破坏文字形状多样性和真实性之间的平衡,使得数据增广只能带来有限的性能提升,甚至产生反作用。

发明内容

[0005] 有鉴于此,有必要针对上述技术问题,提供一种新的基于形状变换的字符图像增广方法,所述方法结合仿射矩阵和TPS变换采样网格参数使STN同时产生全局和局部的形状

变化，提高形变的精细程度；在STN中引入噪声向量注入技术用于丰富成样本的多样性，同时，设计了多样性损失函数和信噪重建损失函数来加强对网络的监督。其中，多样性损失函数用于促进形变参数的差异性，从而增加形变的多样性。而信噪重建损失函数则用于保证信噪平衡，从而使形变程度维持在一个合理的范围之内。

[0006] 本发明公开了一种基于形状变换的字符图像增广方法，包括以下步骤：

[0007] 步骤1，构建形状变换生成对抗网络，包括生成器和判别器；

[0008] 步骤2，以原始字符图像作为生成器的输入，经过空间变换之后产生形变的字符图像，将生成器的输出端与判别器的输入端连接，将目标字符图像输入到判别器的另一个输入端，判别器输出对形变的字符图像和目标字符图像的判别结果；

[0009] 步骤3，训练所述的形状变换生成对抗网络；

[0010] 步骤4，使用训练好的生成器生成增广后的字符图像。

[0011] 具体地，所述的生成器为一个空间变换网络，包括编码器、预测器、采样器、噪声重建网络和图像重建网络；

[0012] 所述的编码器由多个卷积模块顺序连接，每一个卷积模块包括一个二维卷积层、一个非线性激活层和一个汇聚层顺序连接；

[0013] 所述的预测器由多个全连接模块和最后一个全连接层顺序连接，每一个全连接模块包括一个全连接层和一个非线性激活层，最后一个全连接层的输出通道数量设置为所需预测的形变参数的个数；

[0014] 所述的采样器通过在采样网格上应用矩阵乘法将形变的字符图像像素区域映射到原始字符图像像素区域；

[0015] 所述的图像重建网络由多个全连接模块、一个全连接层和多个转置卷积模块顺序连接，每一个转置卷积模块包含依次连接的一个转置卷积层和一个非线性激活层；

[0016] 所述的噪声重建网络由多层全连接模块和一个全连接层顺序连接；

[0017] 首先，原始字符图像作为编码器的输入，编码器从原始字符图像中提取形状特征，输出一个形状特征向量，然后从标准正态分布中随机选取一个噪声隐向量，将形状特征向量和噪声隐向量进行融合，将融合后的隐向量输入到预测器中，预测器负责预测出TPS变换参数和仿射变换参数，其中，TPS变换参数为TPS变换采样网格匹配点的坐标值，将仿射变换参数转化为仿射变换采样网格，接着，将TPS变换采样网格和仿射变换采样网格以及原始字符图像输入到采样器中，输出变换后字符图像，同时，将预测器输出端与图像重建网络和噪声重建网络的输入端连接，分别重建出原始字符图像和噪声隐向量，接着，将生成器输出的形变的字符图像和目标字符图像分别输入到判别器，判别器输出对形变的字符图像和目标字符图像的判别结果；

[0018] 所述的判别器基于patchGAN的结构，由依次连接的五个卷积模块组成，前四个卷积模块中每个卷积模块包括一个二维卷积层、一个实例归一化层、一个leakyReLU激活层，最后一个卷积模块包括一个padding层和一个二维卷积层。

[0019] 优选地，构建最小二乘生成对抗损失函数作为形状变换生成对抗网络训练的优化目标，最小二乘生成对抗损失函数计算公式为：

[0020] $$\mathrm{L}_{G_g} = \mathbb{E}_x\left[D_g\left(G_g(x)\right) - 1\right]^2 + \mathrm{L}_{\mathrm{snr}}\left(G_g\right) + \mathrm{L}_{\mathrm{div}}(E,P)$$

[0021] $$L_{D_g}=\frac{1}{2}\mathbb{E}_x[D_g(G_g(x))]^2+\frac{1}{2}\mathbb{E}_y[D_g(y)-1]^2$$

[0022] L_{G_g}表示生成器损失函数，D_g表示判别器，G_g表示生成器，$L_{snr}(G_g)$表示信噪重建损失函数，$L_{div}(E,P)$表示多样性损失函数，L_{D_g}表示判别器损失函数，x表示原始字符图像，y表示目标字符图像，$\mathbb{E}_x$和$\mathbb{E}_y$表示求相应的数学期望。

[0023] 所述的信噪重建损失函数包括信号重建子项、噪声重建子项和重建误差比值项，计算公式如下：

[0024] $$L_{snr}(G_g)=L_1(x,x_{rec})+L_1(z,z_{rec})+\alpha\cdot\log\left(\frac{L_1(z,z_{rec})}{L_1(x,x_{rec})}\right)$$

[0025] L_1表示平均绝对误差，x表示原始字符图像，z表示噪声隐向量，x_{rec}和z_{rec}分别表示图像重建网络R_x和R_z重建后的原始图像和噪声隐向量，α是动态系数，如果重建误差比值项大于logM，设α=1，当重建误差比值项小于-logM时，设α=-1，如果重建误差比值项在理想范围内，即[-logM,logM]，则设置α=0，M表示超参数；

[0026] 所述的多样性损失函数的计算公式如下：

[0027] $$L_{div}(E,P)=-L_1\left(P(E(x),z_1),P(E(x),z_2)\right)$$

[0028] 其中*P*表示预测器，*E*表示编码器，z_1和z_2分别表示取自相同高斯分布的不同噪声隐向量。

[0029] 可选地，编码器中所述的每一个卷积模块还包括一个批归一化层，位于二维卷积层和非线性激活层中间；编码器中所述的非线性激活层的非线性激活函数选择ReLU函数，所述的汇聚层的汇聚运算选择最大化汇聚。

[0030] 可选地，预测器中所述的每一个全连接模块还包括一个批归一化层，位于全连接层和非线性激活层中间；预测器中所述的非线性激活函数选择ReLU函数，所述的汇聚层的汇聚运算选择最大化汇聚。

[0031] 优选地，预测器中所述的最后一个全连接层的输出通道数量设置为132，所需预测的形变参数中128个参数为64个TPS变换采样网格匹配点的坐标，4个参数为仿射变换矩阵的元素值。

[0032] 可选地，所述的转置卷积模块还包括一个批归一化层，位于转置卷积层和非线性激活层中间；转置卷积模块中所述的非线性激活函数选择ReLU函数，所述的汇聚层的汇聚运算选择最大化汇聚。

[0033] 具体地，所述的采样器通过pytorch中的torch.nn.functional.grid_sample()方法实现，同时，通过pytorch中的torch.nn.functional.affine_grid()方法将仿射参数转化为仿射变换采样网格。

[0034] 更加具体地，TPS变换的目标是求解一个形变函数f，所述的形变函数使得$f(p_i(x_i,y_i))=p_i(x_i',y_i'),\ (1\le i\le n)$，且弯曲能量函数最小，$p_i(x_i,y_i)$表示原始字

符图像上的TPS变换采样网格匹配点的坐标，$p_i(x_i', y_i')$表示形变的字符图像的TPS变换采样网格匹配点的坐标，n为一个TPS变换采样网格匹配点的个数，假设已经获取到两张图像的n组匹配点对：$(p_1(x_1, y_1),\ p_1(x_1', y_1'))$、$(p_2(x_2, y_2), p_2(x_2', y_2'))$、…、$(p_n(x_n, y_n), p_n(x_n', y_n'))$，把形变函数想象成弯折一块薄金属板，使这块板穿过给定的n个点，弯曲薄板的能量函数表示为：

[0035] $$E(f) = \iint \left(\left(\frac{\partial^2 f}{\partial x^2} \right)^2 + 2 \left(\frac{\partial^2 f}{\partial x \partial y} \right)^2 + \left(\frac{\partial^2 f}{\partial y^2} \right)^2 \right) dxdy$$

[0036] 能够证明薄板样条函数就是弯曲能量最小的函数，薄板的样条函数为：

[0037] $$f(x, y) = a_1 + a_2 x + a_3 y + \sum_{i=1}^{n} w_i U(|p(x_i, y_i) - p(x, y)|)$$

[0038] 其中U为基函数：

[0039] $$U(r_{ij}) = \begin{cases} r_{ij}^2 \log r_{ij}^2 & , r_{ij} \neq 0 \\ 0 & , r_{ij} = 0 \end{cases}$$

[0040] $$r_{ij} = |p_i(x_i, y_i) - p_j(x_j, y_j)| = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \quad (4)$$

[0041] a_1，a_2，a_3和w_i，$(1 \leq i \leq n)$通过n个TPS变换采样网格匹配点坐标的预设值和预测器预测的偏移量来求解，由此可以获得$f(x, y)$具体表达式。

[0042] 所述的仿射变换采样网格的采样公式如下：

[0043] $$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} scale \times cos(theta) & -scale \times sin(theta) & t_x \\ scale \times sin(theta) & scale \times cos(theta) & t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix}$$

[0044] 其中，$scale$，$theta$，t_x，t_y分别表示预测器预测出的仿射变换参数，(x, y)和(x', y')分别为变换前后的像素点的位置坐标。

[0045] 优选地，所有图像的像素尺寸为64*64，批大小为64，初始学习率为0.0001，对抗网络的迭代次数为5000，在2500次迭代后学习率开始线性衰减至$1e^{-5}$，对抗网络采用adam优化器进行优化。

[0046] 与现有技术相比，本发明的有益效果在于：

[0047] 本发明方法结合仿射矩阵和TPS变换采样网格参数使STN能够同时产生全局和局部的形状变化，能够更好的拟合字符的形状特征，使产生的字符真实性和多样性更好，使用增广后的数据所训练的分类器的分类性能进一步提升。

[0048] 本发明方法在STN中引入噪声向量注入技术用于促进生成样本的多样性，同时设计了一种能够保证信噪平衡的信噪重建损失函数和一种能够产生丰富的形状变换的多样

性损失函数，对STN的训练提供更强的监督，使样本的形变程度更加合理丰富，将增广后的数据用于训练分类器，能够提高分类器的分类性能。

附图说明

[0049] 图1示出了本发明实施方法的流程示意图；

[0050] 图2示出了本发明实施例各模块工作的结构示意图。

具体实施方式

[0051] 为使本发明实施例的目的、技术方案和优点更加清楚，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、完整地描述，显然，所描述的实施例是本发明一部分实施例，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他实施例，都属于本发明保护的范围。

[0052] 为了引用和清楚起见，下文中使用的技术名词、简写或缩写总结解释如下：

[0053] STN:spatialtransformationnetwork空间变换网络。

[0054] TPS: thin plate spline, 薄板样条函数。

[0055] CNN: convolutional neural network 卷积神经网络。

[0056] FC network:fully connected network 全连接网络。

[0057] Pytorch:一种主流的深度学习框架，封装了许多常用的深度学习相关的函数和类。

[0058] ReLU/leakyReLU:非线性激活函数。

[0059] 生成对抗网络:一种基于零和博弈思想的生成式网络训练框架，包括生成器和判别器两种网络模块。

[0060] 隐向量:随机变量空间中的向量。

[0061] 本发明公开了一种基于形状变换的字符图像增广方法，以解决现有技术中存在的诸多问题。

[0062] 图1示出了本发明实施例的流程示意图。一种基于形状变换的字符图像增广方法，包括以下步骤：

[0063] 步骤1，构建形状变换生成对抗网络，包括生成器和判别器；

[0064] 步骤2，以原始字符图像作为生成器的输入，经过空间变换之后产生形变的字符图像，将生成器的输出端与判别器的输入端连接，同时，将目标字符图像输入到判别器的另一个输入端，判别器输出对形变的字符图像和目标字符图像的判别结果；

[0065] 步骤3，训练所述的形状变换生成对抗网络；

[0066] 步骤4，使用训练好的生成器生成增广后的字符图像。

[0067] 如图2所示，所述的生成器为一个空间变换网络，包括编码器、预测器、采样器、噪声重建网络和图像重建网络。首先，原始字符图像作为编码器的输入，编码器从原始字符图像中提取形状特征，输出一个形状特征向量，然后从标准正态分布中随机选取一个噪声隐向量，将形状特征向量和噪声隐向量进行融合，将融合后的隐向量输入到预测器中，预测器负责映射出TPS变换参数和仿射变换参数，TPS变换参数为TPS变换采样网格匹配点的坐标值，将仿射变换参数转化为仿射变换采样网格，接着，将TPS变换采样网格和仿射变换采样

网格以及原始字符图像输入到采样器中，输出变换后字符图像，同时，将预测器输出端与图像重建网络和噪声重建网络的输入端连接，分别重建出原始字符图像和噪声隐向量。接着，将生成器输出的形变的字符图像和目标字符图像分别输入到判别器，判别器输出对形变的字符图像和目标字符图像的判别结果。

[0068] 具体地，本实施例采用如下步骤进行发明方法的实施。

[0069] 1.分别构建待增广的原始字符数据集和具有目标形状特征的目标字符数据集。

[0070] 2.搭建空间变换网络，作为对抗训练中的生成器，具体步骤如下：

[0071] (1)空间变换网络包含三个模块，分别是编码器、预测器和采样器。首先构造编码器，编码器由相连接的卷积神经网络(CNN)构成，CNN的卷积层层数一般选择超过3层，在我们实现的实例中选择含有依次连接的4个卷积模块，每个卷积模块包括一个二维卷积层、一个批归一化层、一个非线性激活层和一个汇聚(pooling，也称池化)层，其中批归一化层是可选的，非线性激活函数可以选择ReLU函数或者leakyReLU函数，汇聚运算可以选择最大化汇聚、平均汇聚或者自适应汇聚，本实例采用的是ReLU函数和最大化汇聚。

[0072] (2)其次构建预测器，预测器由全连接(FC)神经网络构成，FC层的层数一般选择超过2层，本实例的FC网络包含依次连接的3个FC模块，最后连接一个FC层。每个FC模块包含一个FC层，一个批归一化层和一个非线性激活层，其中批归一化层是可选的，非线性激活函数可以选择ReLU函数或者leakyReLU函数，汇聚运算可以选择最大化汇聚、平均汇聚或者自适应汇聚，本实例采用的是ReLU函数和最大化汇聚，最后一个FC层的输出通道数设置为所需预测的形变参数的个数，本实例采用的形变参数个数为132个，其中128个参数为8*8=64个TPS变换采样网格匹配点的坐标，4个为仿射变换矩阵的元素值。注意，TPS变换采样网格匹配点的个数可以替代为小于原始图像高度或宽度的任意整数的平方。

[0073] (3)接着构建采样器，采样器通过在采样网格上应用矩阵乘法将形变的字符图像像素区域映射到原始字符图像像素区域，本实例采用深度学习框架pytorch中的torch.nn.functional.grid_sample()方法实现。

[0074] (4)最后，构建图像重建网络R_x和噪声重建网络R_y，R_x由依次连接的3个FC模块、1个FC层、4个转置卷积模块组成，R_y由依次连接的3个FC模块、1个FC层组成，其中转置卷积模块包含依次连接的1个转置卷积层、1个批归一化层、1个非线性激活层，其中批归一化层是可选的，非线性激活函数可以选择ReLU函数或者leakyReLU函数，汇聚运算可以选择最大化汇聚、平均汇聚或者自适应汇聚，本实例采用的是ReLU函数和最大化汇聚。重建网络的作用详见第4点。

[0075] (5)空间变换网络的工作原理如下：首先以原始字符图像作为编码器的输入，编码器从原始字符图像中提取出形状特征，输出一个形状特征向量x_v，接着，我们从标准正态分布中随机选取一个噪声隐向量z，将x_v和z进行融合，本实例的融合方式是直接进行叠加求和，其中x_v包含字形特征信息，具有保证输出真实性的作用，z则能够带来一定的随机性，保证了输出的多样性。将融合和后的隐向量输入到预测器中，预测器负责映射出TPS变换参数和仿射变换参数，其中TPS变换参数为TPS变换采样网格匹配点的坐标值，该采样网格具有8*8=64个网格匹配点，而仿射变换参数为仿射变换矩阵的元素值，共有4个参数。接着，通过pytorch中的torch.nn.functional.affine_grid()方法将4个仿射变换参

数转化为仿射变换采样网格。接着，将TPS变换采样网格和仿射变换采样网格以及原始图像输入到采样器中，输出形变的字符图像。我们假设已经获取到两张图像的n组TPS变换采样网格匹配点对：$(p_1(x_1,y_1),p_1(x_1',y_1'))$、$(p_2(x_2,y_2),p_2(x_2',y_2'))$、…、$(p_n(x_n,y_n),p_n(x_n,y_n))$，本实例n取64。那么使用TPS变换计算的坐标对应关系的过程如下：TPS变换的目标是求解一个函数f，使得$f(p_i(x_i,y_i))=p_i(x_i',y_i'),\ (1\le i\le n)$，且弯曲能量函数最小，这样图像上的其它点也可以通过插值得到很好的变换结果。可以把形变函数想象成弯折一块薄金属板，使这块板穿过给定的n个点，弯曲薄板的能量函数可表示为：

[0076] $$E(f)=\iint\left(\left(\frac{\partial^2 f}{\partial x^2}\right)^2+2\left(\frac{\partial^2 f}{\partial x\partial y}\right)^2+\left(\frac{\partial^2 f}{\partial y^2}\right)^2\right)dxdy \quad (1)$$

[0077] 可以证明薄板样条函数就是弯曲能量最小的函数，薄板的样条函数为：

[0078] $$f(x,y)=a_1+a_2x+a_3y+\sum_{i=1}^{n}w_i\,U(|p(x_i,y_i)-p(x,y)|) \quad (2)$$

[0079] 其中U为基函数：

[0080] $$U(r_{ij})=\begin{cases}r_{ij}^2\log r_{ij}^2 & ,r_{ij}\neq 0\\ 0 & ,r_{ij}=0\end{cases} \quad (3)$$

[0081] $$r_{ij}=|p_i(x_i,y_i)-p_j(x_j,y_j)|=\sqrt{(x_i-x_j)^2+(y_i-y_j)^2} \quad (4)$$

[0082] 上式中只要求出a_1，a_2，a_3和$w_i,(1\le i\le n)$，就可以确定$f(x,y)$，a_1，a_2，a_3和$w_i,(1\le i\le n)$可以通过64个TPS变换采样网格匹配点坐标的预设值和预测器预测的偏移量来求解。

[0083] 类似地，假设(x,y)和(x',y')分别为变换前后的像素点的位置，则仿射变换的采样公式如下：

[0084] $$\begin{bmatrix}x\\y\\1\end{bmatrix}=\begin{bmatrix}scale\times cos(theta) & -scale\times sin(theta) & t_x\\ scale\times sin(theta) & scale\times cos(theta) & t_y\\ 0 & 0 & 1\end{bmatrix}\begin{bmatrix}x'\\y'\\1\end{bmatrix} \quad (5)$$

[0085] 其中$scale$，$theta$，t_x，t_y分别表示预测器预测出的4个仿射变换参数。

[0086] 最后，将预测器输出端与图像重建网络和噪声重建网络的输入端连接，分别重建出原始图像和噪声隐向量。

[0087] 3.构建对抗训练中的判别器，判别器是基于patchGAN的结构，由依次连接的5个卷积模块组成，前4个卷积模块中每个模块包括一个二维卷积层、一个实例归一化层、一个leakyReLU激活层。最后一个卷积模块包括一个可选的padding层和一个二维卷积层。

[0088] 4.生成器以原始字符图像x作为输入，经过空间变换网络之后产生形变的字符图像x_t，将生成器的输出端与判别器的输入端连接，同时，把目标字符图像y输入到判

别器的另一个输出端，判别器输出对形变的字符图像和目标字符图像的判别结果。

[0089] 5.构建信噪重建损失函数，该损失函数由三个子项构成，分别为信号重建子项、噪声重建子项和重建误差比值项，公式的计算如下：

[0090] $$\mathrm{L_{snr}}(G_g)=\mathrm{L_1}(x,x_{rec})+\mathrm{L_1}(z,z_{rec})+\alpha\cdot\log\left(\frac{\mathrm{L_1}(z,z_{rec})}{\mathrm{L_1}(x,x_{rec})}\right) \quad (6)$$

[0091] 其中，x_{rec} 和z_{rec} 分别表示重建网络 R_x 和R_z 重建后的原始图像和噪声隐向量。在缺乏强监督的情况下，神经网络学习过程中可能会抑制输入信息产生的作用，为了避免这种情况，我们分别为信息和噪声向量分别设计重建损失项，从而使形状信息保证真实性和噪声保证多样性的作用不会被抑制。此外，为了保证变换字形的变换程度更加合理可控，我们还设计了一个重建损失比值来寻求形状信息和噪声各自作用的平衡，我们进一步用超参数M > 1来约束这一项。其中α是动态系数，在训练过程中，如果该项大于logM，我们设α = 1。在这种情况下，噪声重构比信号重构差得多。这意味着噪声的作用被网络抑制了，我们使用梯度下降法对正比值项进行优化。相反，当该项小于-logM时，我们设α =-1。在这种情况下，信号重构比噪声重构差得多。这意味着噪音的作用过于突出，形状信息的作用被抑制，我们使用梯度下降法来优化负比值项。如果该项在理想范围内，即[-logM,logM]，则设置α = 0，即没有任何额外优化。

[0092] 6.构建多样性损失函数，多样性与不同噪声隐向量对应的变换参数之间的差异呈正相关，这两个变换参数分别是从两个分别注入噪声的信噪混合隐向量中估计出来的。多样性损失的定义如下：

[0093] $$\mathrm{L_{div}}(E,P)=-\mathrm{L_1}\left(P(E(x),z_1),P(E(x),z_2)\right) \quad (7)$$

[0094] 其中P表示预测器，E表示编码器，z_1 和z_2 分别表示取自相同高斯分布的不同噪声隐向量。

[0095] 7.构建最小二乘生成对抗损失函数作为对抗训练的优化目标，对抗损失是为了拉近形变的字符图像和目标字符图像之间的分布，具体公式如下：

[0096] $$\mathrm{L_{G_g}}=\mathbb{E}_x\left[D_g\left(G_g(x)\right)-1\right]^2+\mathrm{L_{snr}}(G_g)+\mathrm{L_{div}}(E,P)$$

[0097] $$\mathrm{L_{D_g}}=\frac{1}{2}\mathbb{E}_x\left[D_g\left(G_g(x)\right)\right]^2+\frac{1}{2}\mathbb{E}_y\left[D_g(y)-1\right]^2 \quad (8)$$

[0098] 通过对抗训练，空间变换网络能够产生逼近目标字符图像的样本。

[0099] 8.所有图像像素尺寸设置为64*64，批大小为64，初始学习率为0.0001，网络的迭代次数为5000，在2500次迭代后学习率开始线性衰减至1e-5，网络采用adam优化器进行优化。

[0100] 9.按照7中的设定训练整个形状变换生成对抗网络，得到训练好的生成器，可用于产生多样的增广样本。

[0101] 本领域普通技术人员可以意识到，结合本文中所公开的实施例描述的各示例的单元及算法步骤，能够以电子硬件、或者计算机软件和电子硬件的结合来实现。这些功能究竟以硬件还是软件方式来执行，取决于技术方案的特定应用和设计约束条件。专业技术人员可以对每个特定的应用来使用不同方法来实现所描述的功能，但是这种实现不应认为超出

本发明的范围。

[0102] 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。

[0103] 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。

[0104] 以上所述仅是本发明的优选实施方式，应当指出，对于本技术领域的普通技术人员来说，在不脱离本发明原理的前提下，还可以做出若干改进和润饰，这些改进和润饰也应视为本发明的保护范围。

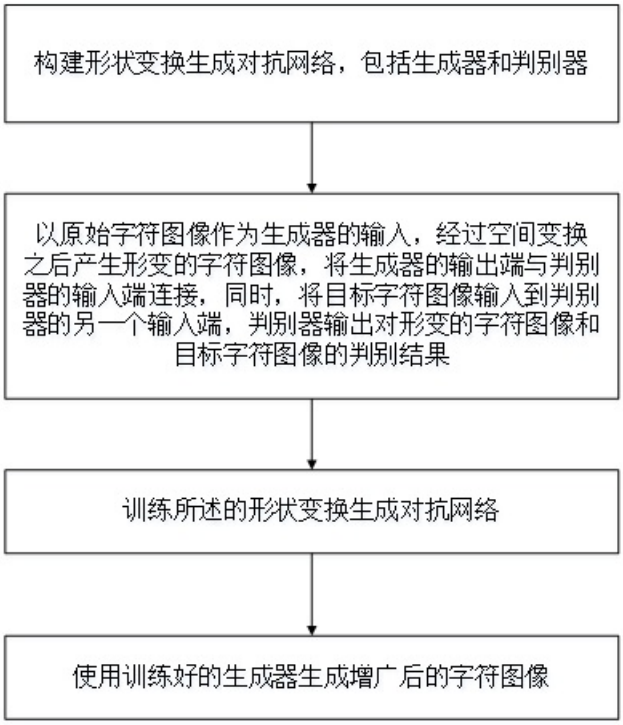

构建形状变换生成对抗网络，包括生成器和判别器
以原始字符图像作为生成器的输入，经过空间变换之后产生形变的字符图像，将生成器的输出端与判别器的输入端连接，同时，将目标字符图像输入到判别器的另一个输入端，判别器输出对形变的字符图像和目标字符图像的判别结果
训练所述的形状变换生成对抗网络
使用训练好的生成器生成增广后的字符图像

图1

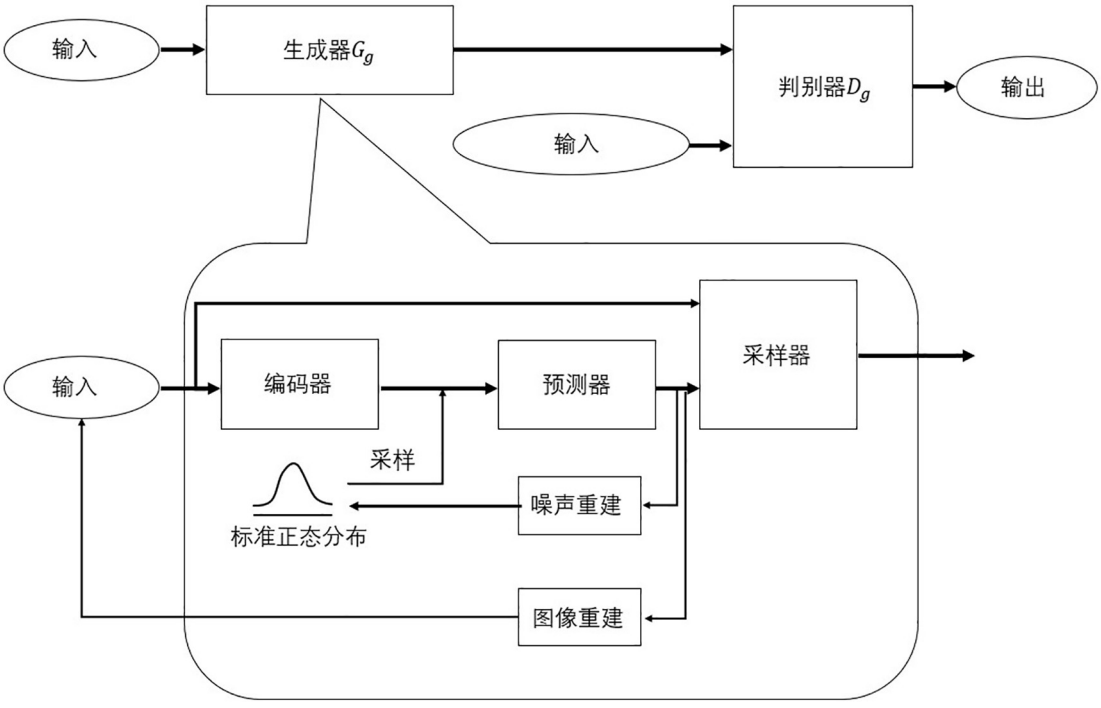

输入
生成器G_g
判别器D_g
输出
输入
采样器
输入
编码器
预测器
采样
噪声重建
标准正态分布
图像重建

图2