



(12)发明专利申请

(10)申请公布号 CN 107590497 A

(43)申请公布日 2018.01.16

(21)申请号 201710855035.7

(22)申请日 2017.09.20

(71)申请人 重庆邮电大学

地址 400065 重庆市南岸区南山街道崇文
路2号

(72)发明人 赵辉 王艳美 刘真三

(51)Int.Cl.

G06K 9/32(2006.01)

G06K 9/46(2006.01)

G06K 9/62(2006.01)

G06N 3/02(2006.01)

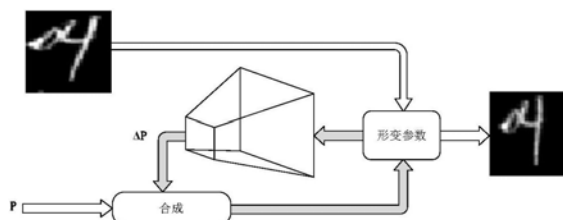
权利要求书2页 说明书7页 附图5页

(54)发明名称

基于深度卷积神经网络的脱机手写汉字识别方法

(57)摘要

本发明提出采用深度卷积神经网络的脱机手写汉字方法,具体地,先采用空间变换网络对HWDB1.1数据库的手写汉字进行旋转、平移和缩放。使得原本任意位置,比例和方向的汉字图片得到纠正。之后将纠正的手写数据集输入到卷积神经网络进行识别分类。其中对原图片进行旋转、平移和缩放的扭曲参数增量是由一个卷积神经网络或神经网络不断通过反向传播调整计算而来。扭曲参数的增量和扭曲参数用合成的方式进行更新参数。最后我们在TensorFlow深度学习框架平台上搭建了本发明的手写汉字识别的网络框架。并与单纯的采用卷积神经网络识别手写汉字网络框架作对比,经过大量数据集的训练,测试的结果表明手写汉字识别率得到显著提高。



1. 一种基于深度卷积神经网络的脱机手写汉字识别的方法,其特征包括如下步骤:

在Windows上搭建TensorFlow深度学习框架;

准备数据集,将数据集转化为TensorFlow的输入形式,并将图片进行归一化预处理;

在TensorFlow环境下搭建反向合成空间变换网络以及深度的卷积神经网络,采用反向合成空间网络对手写汉字进行扭曲形变,得到纠正和对齐。并将其输出作为卷积神经网络的输入;

最后采用大量的数据对反向合成空间变换网络和卷积神经网络形成的网络进行训练、测试。并通过TensorBoard工具对网络处理过程中的图片及最终测试结果可视化,分析其作用,并对比识别效果。

2. 根据权利要求书1所述步骤,使用反向合成空间变换对手写汉字进行纠正,其特征在于:

采用扭曲参数组成的齐次矩阵对原图像进行变换对齐,扭曲参数以仿射方式为例,参数 $p = [p_1 \ p_2 \ p_3 \ p_4 \ p_5 \ p_6]$,对应的齐次变换矩阵:

$$W(p) = \begin{bmatrix} 1+p_1 & p_2 & p_3 \\ p_4 & 1+p_5 & p_6 \\ 0 & 0 & 1 \end{bmatrix}$$

将其作用在图像上,对图像 $I(x)$ 进行扭曲形变,包括图像的旋转,平移、放缩等操作。

$$ImWarp(x) = I(x) \cdot W(p)$$

$ImWarp(x)$ 即为纠正过的图片。参数 p 是采用迭代的方式进行更新求取,而且 p 的扭曲增量 Δp 是由一个卷积神经网络或者神经网络计算而来,本发明将此网络称为几何预测器。由于反向合成空间变换原理受反向合成Lucas-Kanade算法的启发:

$$\min_{\Delta p} \|I(P) - T(\Delta P)\|_2^2$$

经过一系列的推到和转化,其解为:

$$\Delta p = R \cdot I(p) + b$$

最终解是线性回归形式,所以可以将其求解过程参数化为神经网络或者卷积神经网络;

求扭曲增量 Δp 的几何预测器的网络结构有三种方案,两个方案是卷积神经网络结构,一个为单层的神经网络。说明书中有详细的结构说明。我们最终会逐一验证每个方案并分析其有效性。

几何预测器更新的 Δp 与之前的 p 合成,更新参数 p 。之后可以将更新的 p 值从新作用在图像上,循环迭代求取最佳的 p 值。根据合成定义,更新方式如下:

$$p \leftarrow p \circ (\Delta p)^{-1}$$

采用合成的方式来更新参数的好处是不必每次都去求样本图像的梯度。大大加快了反向传播的计算速度。

最后将经过纠正的图片输入到卷积神经网络中进行特征提取与分类。卷积神经网络的结构顺序分别为:两个卷积层、一个池化层、接着两个卷积层、一个池化,最后是两个全连接

层。最终的输出神经元个数与本发明实例的类别相同。

基于深度卷积神经网络的脱机手写汉字识别方法

技术领域

[0001] 本发明属于图像分类技术领域,具体是基于深度卷积神经网络的脱机手写汉字识别方法。

背景技术

[0002] 脱机手写汉字识别是模式识别领域中的一个子方向。脱机是指所处理的手写文字是通过扫描仪或摄像头等图像捕捉设备采集到的手写文字二维图,以下均简称手写汉字识别。2011年以前发表的研究论文和技术报告大都着重讨论如何选择特征和匹配方法以适应手写体汉字字形的变化。就是在特征提取算法以及分类器的设计。传统HCCR步骤包括:图形归一化、特征提取、降维、分类器训练。而且由于汉字数量多,结构复杂、相似字多和书写风格多变等问题,这些传统方法不仅步骤复杂,而且在特征提取步骤中如果选取的特征不适合将会严重影响识别效果。虽然基于MQDF和DLQDF的方法已经取得了不错的效果,但是目前已经达到了他们的瓶颈。

[0003] 近几年由于基于深度学习的图像识别取得了重大的突破,深度卷积神经网络(Deep Convolutional Neural Network,DCNN)也应用在了手写汉字识别领域。在2013ICDAR国际比赛中,Fujitsu团队采用CNN模型以94.77%的识别率获得冠军。之后出现了采用改进的模型以及预处理手段的手写汉字识别模型。最常见的改进方法是使网络增加对不同手写汉字的形变的容忍度,其中就包括数据增强技术和空间池化,这两个方法有以下缺点,前者,根据对原始数据几何形变产生新的样本的方法是在手写汉字识别中应用最多的一种方法,如果只有样本成指数级增长的模型的学习能力才得以提升,则也是个不容忽视的问题。后者,池化操作会损坏图像细节。而且随着数据的爆炸式增长深度卷积神经网络虽然获得很大的成功,但是还没有一种准则解决同一种类数据的几何变化。比如同一个手写的“汉”字,不同的人写出来的大小和笔画的形状都是不尽相同的。数据的变换多样是影响深度卷积神经网络识别效果的关键因素。

[0004] 改进的方法有:预处理方式,对汉字进行扭曲形变来扩充数据集,提高卷积神经网络的泛化能力;特征提取,结合传统Gabor方向特征、HoG特征等,采用不同DCNN模型。目前已经采用的模型有Szegedy等组成的GoogLeNet,K He,X Zhang等人设计的ResLeNet等模型。这些模型都是为图片分类设计。与手写汉字识别还是有一定的区别,虽然达到了不错的效果但是网络结构较深,模型复杂,调试困难。而且对汉字进行扭曲形变的程度不好把握,不能自适应的选取角度。2015年,来自于Google旗下的新锐AI公司DeepMind的四位剑桥Phd

[0005] 研究员设计了空间变换网络(Spatial Transformer Network),可以实现自适应的旋转、平移、缩放。但是直接将反向空间变换网络与卷积神经网络构成的框架用于手写汉字识别,也有一些问题。比如,经过反向空间变换网络之后,虽然汉字的方向得到纠正,但是纠正过的汉字笔画较粗。而且处理过程中样本是经过多层放缩、旋转平移的样本,然后直接输入到卷积神经网络识别。因为经过放缩的样本裁剪掉了原图片的边缘信息,用在手写汉字识别中会造成字体残缺,严重影响识别效果。

[0006] 因为目前大部分网络忽略了实际手写汉字的空间畸变,因为在实际书写环境、风格、方向、位置、大小的不同导致样本集变化多样。然而,CNN仍然缺乏对输入样本的空间变化的鲁棒性。传统的归一化方法仅把样本转化为规定大小的规范化汉字,虽然它对分类任务起到了举足轻重的作用,但是归一化方法不能保证HCCR任务是最佳的。而反向合成空间变换网络能够在没有标注关键点的情况下,根据任务学习图片的特征变换参数,将输入图片或者学习的特征空间上进行对齐,从而减少由于空间的旋转、平移、尺度、扭曲等几何变换对分类和识别的影响。

发明内容

[0007] 本发明针对上面存在的问题以及汉字的特点,采用反向合成空间变换算法以及深度卷积神经网络等算法对脱机手写汉字进行识别。本方法对不同的书写风格以及不同的书写环境有良好的鲁棒性。其中我们采用了反向合成空间变换网络来解决来因书写风格多样带来的字体扭曲、变形、倾斜等问题。并设计了与之对应的深度卷积神经网络。深度卷积神经网络可以有效提取特征并分类。最后反向合成空间变换网络的输出作为卷积神经网络的输入,实现对于手写汉字的识别。

[0008] 本发明的技术方案及流程顺序如下:

[0009] (1) 搭建深度卷积神经网络的TensorFlow深度学习框架平台;

[0010] (2) 将HWDB1.1数据集的GNT格式数据转化为二进制并存储为PKL格式

[0011] (3) 读取PKL格式数据并进行归一化处理,并将其转换为训练集、交叉验证集和测试集;

[0012] (4) 在TensorFlow框架平台中实现反向合成空间变换网络和卷积神经网络,并将反相合成空间变换网络的输出作为卷积神经网络的输入;

[0013] (5) 设置方向合成空间变换网络框架模型并将其加入本发明人自己设计的卷积神经网络框架中,形成反相合成空间变换网络加卷积神经网络的模型结构;

[0014] (6) 训练和测试构建好的网络框架。

[0015] (7) 采用TensorBoard工具可视化中间结果、训练和测试的结果。对比分析本发明网络结构的有效性。

[0016] 所述步骤2) 的详细过程:HWDB1.1是CASIA-HWDB数据库的一个脱机手写汉字的数据集。其是中科院在2007-2010年间收集。存储的格式为独有的二进制GNT格式。因此我们需要首先将此数据集转化为可用的格式。因为此数据集书写字体多样,是目前验证模型识别效果最多也是最权威的数据集之一。

[0017] 所述步骤3):将图片进行反转并归一化,前者可以使图片大部分数据为零,加快运算速度,后者可以提高网络的收敛速度。

[0018] 所述步骤4):反向合成空间变换网络的具体过程为:训练样本图像与初始化的扭曲参数 p 相乘,得到第一次纠正的图片。将其输入到一个几何预测器,几何预测器是由一个小型卷积神经网络或者神经网络构成。网络的前向传播预测扭曲增量 Δp ,反向传播BP算法来更新几何预测器的参数,进一步迭代更新扭曲参数。每更新 Δp 之后,用合成的方式更新参数 p 。之后将 P 组成的矩阵与训练样本相乘结果为纠正过的图片,将经过纠正的图片再此输入的几何预测器,这样迭代更新 Δp 、 p 。直到满足提前给定的循环次数,将最后的 P 值与原

训练样本相乘,将得到的结果作为本发明的设计的卷积神经网络的输入进一步的提取特征和分类。

[0019] 所述步骤5):卷积神经网络采用前两个卷积加一个池化层,之后两个卷积层,卷积层之后是两个全连接层。

[0020] 与目前的基于深度卷积神经网络的手写汉字相比,1) 本发明采用反向合成空间变换网络可以将手写汉字集经过扭曲形变而转化为易于识别的汉字图片有效的提高了手写汉字的识别效果.而且避免的人为设计参数对图片的扭曲形变等操作。2) 卷积神经网络采用两个卷积层加一个池化层加两个卷积层的网络结构。采用较多的卷积层和小的卷积核来有效的提取汉字笔画特征。

附图说明

[0021] 图1反向合成空间变换网络流程图

[0022] 图2CNN_4网络结构

[0023] 图3方案一经过反向空间变换网络前后对比图

[0024] (a) 原手写汉字

[0025] (b) 经过ICSTN_2纠正的手写汉字

[0026] 图4方案二经过反向空间变换网络前后对比图

[0027] (a) 原手写汉字

[0028] (b) 经过ICSTN_2纠正的手写汉字

[0029] 图5方案三经过反向空间变换网络前后对比图

[0030] (a) 原手写汉字

[0031] (b) 经过ICSTN_4纠正的手写汉字

[0032] 图6方案一代价函数和测试准确率变化曲线图

[0033] (a) 采用ICSTN_1+CNN_4损失函数

[0034] (b) 采用ICSTN_1+CNN_4测试准确率

[0035] 图7方案二代价函数和测试准确率变化曲线图

[0036] (a) 采用ICSTN_2+CNN_4损失函数

[0037] (b) 采用ICSTN_2+CNN_4测试准确率

[0038] 图8方案三代价函数和测试准确率变化曲线图

[0039] (a) ICSTN_4+CNN_4损失函数

[0040] (b) 采用ICSTN_4+CNN_4测试准确率

[0041] 图9CNN_4代价函数和测试准确率变化曲线图

[0042] (a) 采用CNN_4损失函数

[0043] (b) CNN_4测试准确率

[0044] 具体的实施方式

[0045] 下面结合附图对本发明进一步的详细说明。

[0046] 以下给出网络的具体构造方法,以及本发明采用的新理论的算法原理及其效果。对本发明的实施进一步的说明。本发明所采用的实例为HWDB1.1的200个种类的汉字,每个类别300个汉字,共60000个样本,其中训练集,交叉验证集和测试机分别为:48000、6000,

6000。本发明所设计的网络运行的具体流程为：

[0047] (1) 数据预处理, 将获取的二进制图像反转, 归一化处理, 以提高网络的收敛性。并将图像尺寸统一调整为 32×32 , 然后分配为训练集、交叉验证集和测试集;

[0048] (2) 将训练样本输入到下一步形变参数, 记为 $I(x)$, $x = (x, y)$ 表示图像中任意像素点得坐。 $p = [p_1 \ p_2 \ \cdots \ p_n]$ 表示扭曲参数。本发明都以仿射变换为 $p = [p_1 \ p_2 \ p_3 \ p_4 \ p_5 \ p_6]$, 相关的其次坐标表变换矩阵为:

$$[0049] \quad W(p) = \begin{bmatrix} 1+p_1 & p_2 & p_3 \\ p_4 & 1+p_1 & p_6 \\ 0 & 0 & 1 \end{bmatrix} \quad (1)$$

[0050] (3) $W(p)$ 对原图像进行变换:

$$[0051] \quad \text{ImWarp}(x) = I(x) \cdot W(p) \quad (2)$$

[0052] (4) 因为经过扭曲变换图像坐标是非整的导致图像不连续, 本发明将经过变换的图像 $\text{ImWarp}(x)$ 进行双线性插值:

$$[0053] \quad V_i^c = \sum_n^H \sum_m^W \text{ImWarp}_{n,m}^c \max(0, |x_i - m|) \max(0, |y_i - n|) \quad (3)$$

[0054] V_i^c : 表示第 c 个样本的第 i 个像素点;

[0055] (x_i, y_i) : 表示训练样本 ImWarp 的第 i 个像素点坐标;

[0056] W : 表示图像的宽;

[0057] H : 表示图像的高;

[0058] (5) 将图像 $V(x)$ 输入到卷积神经网络或者神经网络中, 本发明称此网络为几何预测器(后面将会详细的推导为何用神经网络或者卷积神经网络学习扭曲增量), 几何预测器用来学习图像的扭曲增量 Δp , 其网络结构的组成共有三种方案。最后本发明将会逐一仿真实现每个方案的效果。三种方案分别为:

[0059] a) 方案一: 网络的结构为: $\text{Conv}(7 \times 7, 4) + \text{Conv}(7 \times 7, 8) + P + \text{FC}(48) + \text{FC}(8)$, 其中以 Conv 表示卷积层, P 表示池化层, FC 表示全连接层, 括号内分别为卷积核的大小和个数, 全连接层括号内为神经元个数;

[0060] b) 方案二: 网络的结构为: $\text{Conv}(9 \times 9, 8) + \text{FC}(48) + \text{FC}(8)$;

[0061] c) 方案三: 网络的结构为: $\text{FC}(48) + \text{FC}(8)$

[0062] 几何预测器的前向传播输出扭曲增量 Δp , 反向传播更新网络参数, 进一步产生新的 Δp 。

[0063] (6) 更新扭曲参数, 扭曲参数 p 的更新方式为合成。具体公式下:

$$[0064] \quad p \leftarrow p \circ (\Delta p)^{-1} \quad (4)$$

[0065] 与之对应的变换矩阵为:

$$[0066] \quad W(p_{\text{out}}) = W(\Delta P) \cdot W(p_{\text{in}}) \quad (5)$$

[0067] (7) 如果循环次数 $\text{recurN} > 1$, 转到步骤(3)直到循环结束。反向合成空间变换网络可以多次循环以获得对齐和扭曲图像最好的效果。

[0068] (8) 循环结束后, 将最终更新的扭曲参数 p 组成的变换矩阵 $W(p_{\text{out}})$ 作用在原图像 j 即:

[0069] $I_m = I(x) \cdot W(P_{out})$ (6)

[0070] ,之后将图像输入到卷积神经网络中;

[0071] (9) 卷积神经网络的框架如图2,网络结构为Conv(3×3,8)+Conv(3×3,16)+P+Conv(3×3,32)+Conv(3×3,64)+P+FC(100)+FC(200),每个参数具体的代表什么与上述的相同。接下来我们以一个卷积层一个池化层和一个全连接层进行实例讲解。

[0072] 卷积层,卷积层的主要内容是采用可训练的卷积核来对输入数据进行卷积操作,并将结果以某种组合形式输出,其实质上就是对输入数据的特征提取。一般情况下,为了使模型获得非线性特性并把输出限制在给定的范围内,会将卷积的输出用一个非线性函数(Non-linear Function)来转换。这个函数也称为激活函数(Activation Function)。本实例采用的是收敛速度较快的Relu激活函数。卷积层的具体运算公式为:

$$[0073] \quad U = \sum_{i \in M_j} I_m \otimes W + b$$

$$Y = f(U) \quad (7)$$

[0074] 其中 I_m 为反向空间变换网络的输出,就是经过反向空间变换网络纠正的扭曲汉字的训练样本, W 为卷积核, b 为卷积层偏置; U 为卷积层输出, f 即为Relu激活函数, Y 为卷积输出 U 经过激活函数的输出,是整个卷积层的输出,也是下一层的输入。

[0075] 池化层,主要任务是在二维空间对输入的数据样本进行采样操作,也称为下采样操作,具体的计算过程如下:

$$[0076] \quad y_j^l = f(u_j^l) = f(\text{pooling}(x_i^l) \times k_i^l + b_j^l) \quad (8)$$

[0077] 池化层可使网络对平移,旋转等具有鲁棒性。使网络有更好的容忍度。

[0078] 全连接层,全连接层的作用是将输入的二维特征矩阵降维到一维特征向量,便于输出层进行分类处理。

[0079] 输出层,输出层的作用是根据上面全连接层的输出的一维向量进行分类,本列采用的是softmax交叉熵损失

[0080] 图像的扭曲偏移量 Δp 可以由卷积神经网络模型求取是因为对于反向合成KL图像对齐算法:

$$[0081] \quad \min_{\Delta p} \|I(P) - T(\Delta P)\|_2^2 \quad (9)$$

[0082] 式(9)反向合成图像对齐的目标函数,其中 I 是原图像, T 是目标图像, p :扭曲参数、 Δp :要求取的扭曲增量; $I(p)$ 和 $T(\Delta p)$ 分别表示经过扭曲参数作用的图像。然后,将上式(9)经过一阶Taylor近似得:

$$[0083] \quad \min_{\Delta p} \left\| I(P) - T(0) - \frac{\partial T(0)}{\partial P} \Delta P \right\|_2^2 \quad (10)$$

[0084] 其最小二乘解为:

$$[0085] \quad \Delta p = \left(\frac{\partial T(0)}{\partial p} \right)^\dagger (I(p) - T(0)) \quad (11)$$

[0086] 其中 $+$ 符号代表广义逆。接下来参数 p 的更新方式为:

$$[0087] \quad p \leftarrow p \circ (\Delta p)^{-1} \quad (12)$$

[0088] 其中,为扭曲参数 p 的合成运算,式(10)的解就是个线性回归形式,可以将其转化为一般式:

$$[0089] \quad \Delta p = R \cdot I(p) + b \quad (13)$$

[0090] 其中 R 是估计图像和图像的几何形状之间的线性关系的线性回归, b 为偏置。主要是求取参数 R 和 b 。线性回归模型我们可以将其参数化为神经网络或者卷积神经网络,网络的输出即为 Δp 。

[0091] 反向合成空间变换网络不是由网络直接预测图像的几何变换,而是在分类之前多次循环使用几何预测器更新扭曲参数 p ,而且在每次循环之后,扭曲参数 p 在输出图像后得以保存,这将使得在重新采样原图像时避免边缘效应。几何预测子 $\{R, b\}$,用监督梯度下降优化和学习的目标函数,其目标函数为:

$$[0092] \quad \min_{R, b} \sum_{n=1}^N \sum_{j=1}^M \|\Delta p_{n,j} - R \cdot I_n(p_n \circ \Delta p_{n,j}) - b\|_2^2 \quad (14)$$

[0093] Δp : 扭曲增量,是几何预测器的输出

[0094] M : 图像的样本数

[0095] 由公式, (12) 和 (13) 可得:

$$[0096] \quad \begin{aligned} R \cdot I_n(p_n \circ \Delta p_{n,j}) + b &= R \cdot I_n \cdot W(P_{in}) \cdot W(\Delta P) \\ &= R \cdot I_n \cdot W(P_{out}) \end{aligned} \quad (15)$$

[0097] 采用合成算法更新参数,不用每次都更新图像对参数的一阶偏导数,反向传播中求取梯度为中,扭曲参数的偏导数是个闭式数学表达式, P_{in}, P_{out} 分别为输入输出的扭曲参数:

$$[0098] \quad \frac{\partial p_{out}}{\partial p_{in}} = \mathbf{I} + \begin{bmatrix} \Delta p_1 & 0 & 0 & \Delta p_2 & 0 & 0 \\ 0 & \Delta p_1 & 0 & 0 & \Delta p_2 & 0 \\ 0 & 0 & \Delta p_1 & 0 & 0 & \Delta p_2 \\ \Delta p_4 & 0 & 0 & \Delta p_5 & 0 & 0 \\ 0 & \Delta p_4 & 0 & 0 & \Delta p_5 & 0 \\ 0 & 0 & \Delta p_4 & 0 & 0 & \Delta p_5 \end{bmatrix} \quad (16)$$

$$[0099] \quad \frac{\partial P_{out}}{\partial p_{in}} = \mathbf{I} + \begin{bmatrix} P_{in,1} & P_{in,4} & 0 & 0 & 0 & 0 \\ P_{in,2} & P_{in,5} & 0 & 0 & 0 & 0 \\ P_{in,3} & P_{in,6} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & P_{in,1} & P_{in,4} & 0 \\ 0 & 0 & 0 & P_{in,2} & P_{in,5} & 0 \\ 0 & 0 & 0 & P_{in,3} & P_{in,6} & 0 \end{bmatrix} \quad (17)$$

[0100] \mathbf{I} 是单位矩阵,它使得梯度可以反向传播到几何预测器。

[0101] 为了直观地说明我们构造的网络框架的可行性和有效性,本发明通过实验仿真进行了验证。并逐一验证了我们设计的几何预测器的网络的三个方案。每个方案所组成的框架在训练时采用相同的参数配置:epoch=16,初始学习率为0.001,每两个epoch学习率乘以0.8的因子。为了对比采用反相合成空间的有效性,本发明单独仿真实现了仅有卷积神经网络的识别效果。

[0102] 首先本发明可视化了经过反向空间变换网络之前和之后的训练样本。方案一、二、三所对应的模型分别为:ICSTN_1+CNN_4、ICSTN_2+CNN_4和ICSTN_4+CNN_4。其中ICSTN后面的数字为模型深度。如图3、4、5分别为三个不同的几何预测器模型对样本变换前后对比图。从整体看,在经过反向合成空间变换网络之后,每个汉字大小相同、笔画均匀、排列整齐,一些扭曲的笔画得到纠正,比原汉字辨识度高了很多。不同模型间的差异也相当显著,整体趋势是几何预测器的网络越深,经过空间变换的汉字笔画就越清晰,而且旋转过度情况较少,越多的扭曲笔画得到纠正。如图5(b)中汉字“巾”、“臼”、“尽”等字得到恰当的纠正变形。而图3效果最差,有较多的汉字产生了倾斜。综上可得出,反向合成空间变换网络的几何预测器的网络越深越能够有效的纠正扭曲形变的汉字。

[0103] 其次分析各网络框架的迭代曲线图:图6、7、8分别为三个方案所组成三个模型的损失函数曲线和测试准确率曲线。纵坐标的迭代次数是以50为单位,即每50次训练迭代的结果显示一次,所以共50000次迭代,10次epoch。从三个图的(a)图来看ICSTN_2+CNN_4和ICSTN_4+CNN_4模型收敛较快,模型ICSTN_4+CNN_4的模型稳定性最好,因为曲线的尖点较少。但是ICSTN_2+CNN_4模型的loss最接近零点。从三个图中的(b)图中可看得ICSTN_2+CNN_4的识别率最高。那为什么方案三形成的模型纠正汉字的效果最好,却没有得到最好的识别率呢?原因主要是ICSTN_4+CNN_4模型的参数较多在相同的epoch下没有得到充分的训练。CNN模型与方案一、二、三组成的模型相比差距较大,图9中的(a)(b)与图6、7、8的(a)(b)图相比损失函数收敛的值最大,识别率仅有82.08%。CNN_4的识别效果不好是因为模型较浅。

[0104] 最后从表1可得,从训练模型消耗的时间上来看,反向空间变换网络越深所消耗的时间越多。所以综合效率以及识别率等因素,ICSTN_2+CNN_4的效果和料率最好。从测试的准确率来看。本发明有明显的优势。而CNN_4模型与ICSTN_1+CNN_4模型训练时间相差较小但是识别效果差距明显。ICSTN_1仅采用了一个一层的神经网络。所以本发明实现了反向空间变换网络和较浅的卷积神经网络相结合却得到了一个较高的识别的效果。

[0105] 表1不同网络结构的识别效果

[0106]

模型类别	识别准确率(%)	耗时(h)
CNN_4	82.08	5.16
ICSTN_4+CNN_4	95.83	17.72
ICSTN_2+CNN_4	96.38	13.40
ICSTN_1+CNN_4	95.32	6.92

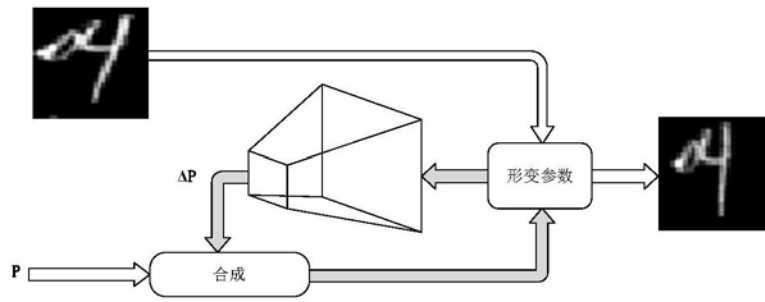


图1

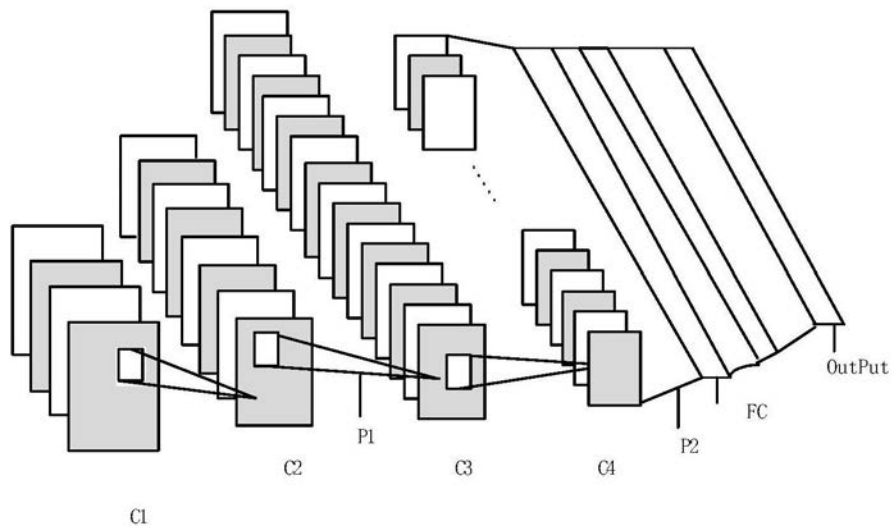
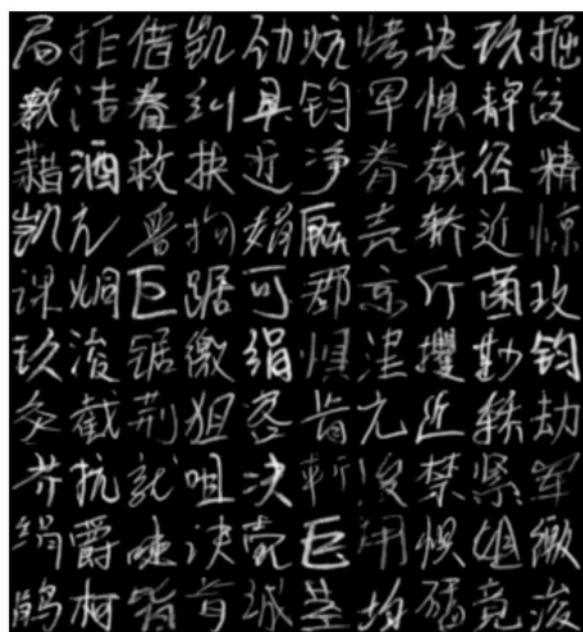


图2

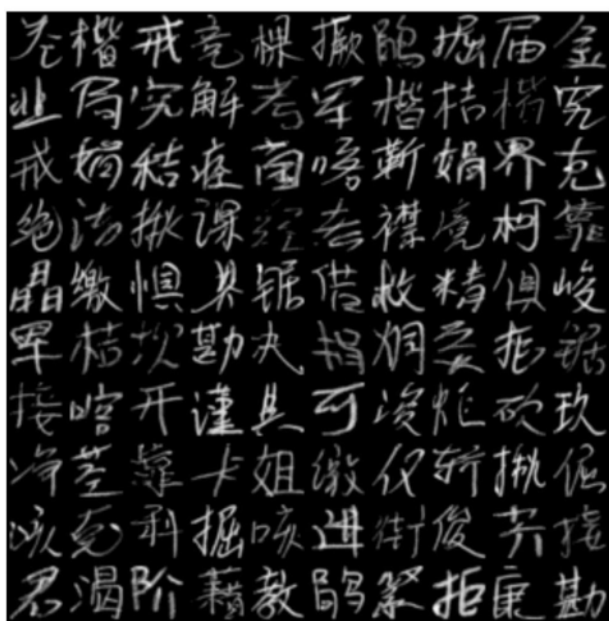


(a)



(b)

图3

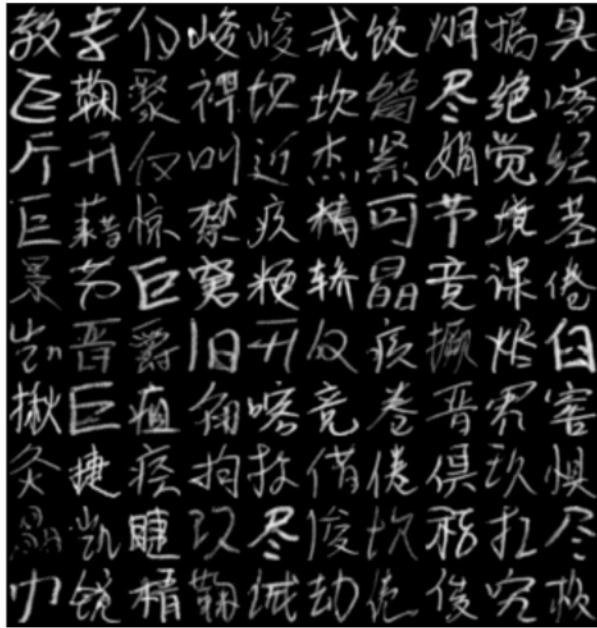


(a)



(b)

图4

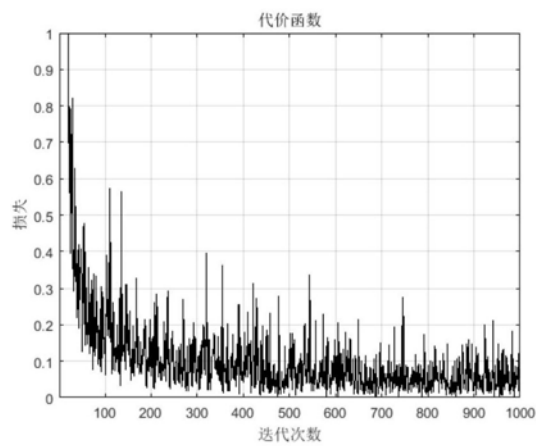


(a)

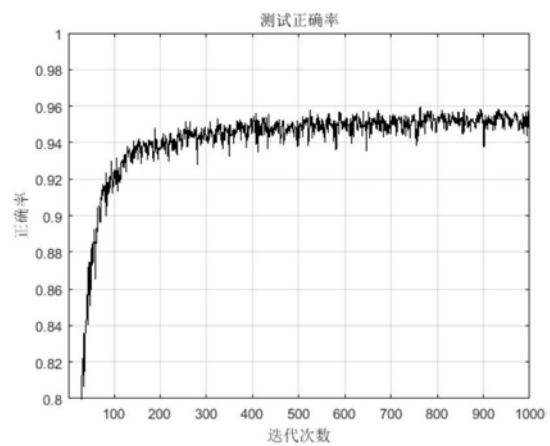


(b)

图5



(a)



(b)

图6

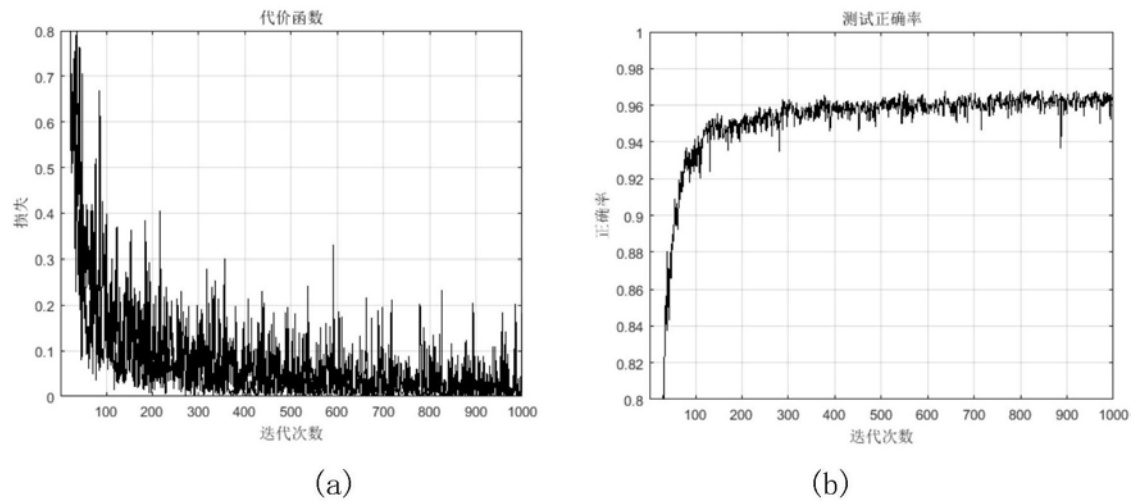


图7

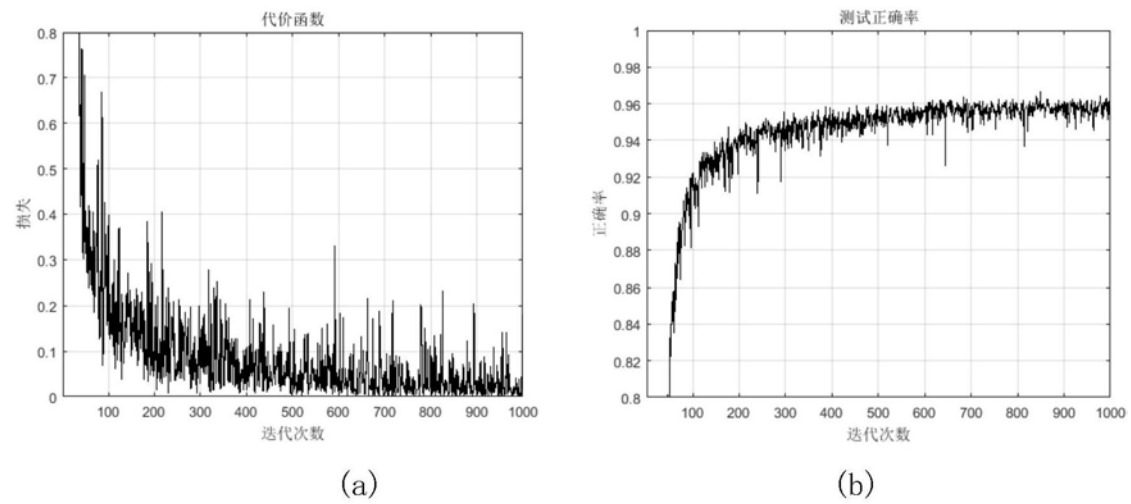


图8

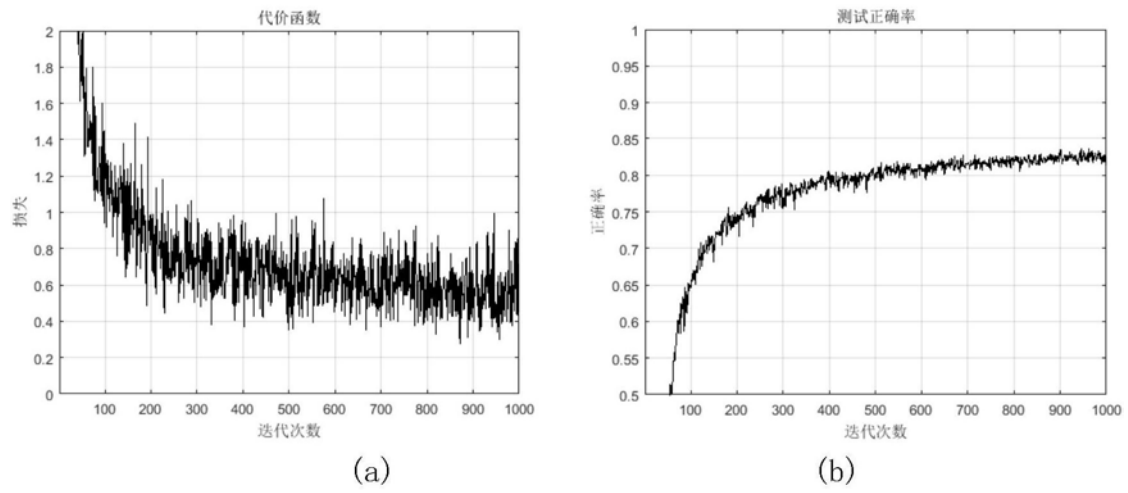


图9