



(12)发明专利申请

(10)申请公布号 CN 106469304 A
(43)申请公布日 2017. 03. 01

(21)申请号 201610841643.8

(22)申请日 2016.09.22

(71)申请人 西安理工大学

地址 710048 陕西省西安市金花南路5号

(72)发明人 张二虎 李雪薇

(74)专利代理机构 西安弘理专利事务所 61214

代理人 成丹

(51)Int.Cl.

G06K 9/00(2006.01)

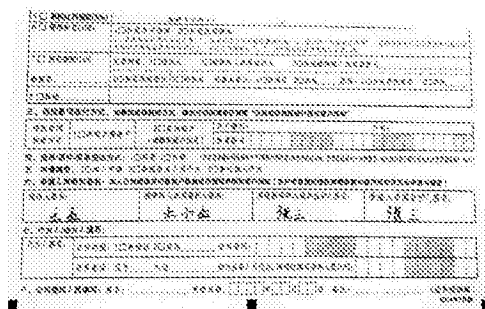
权利要求书2页 说明书6页 附图4页

(54)发明名称

基于深度卷积神经网络的票据中手写签名位置定位方法

(57)摘要

本发明公开了基于深度卷积神经网络的票据中手写签名位置定位方法,具体按照以下步骤实施:步骤1:搭建基于包含多种卷积神经网络模型的caffe深度学习框架的平台;步骤2:准备票据的数据集;步骤3:训练网络得到定位检测模型;步骤4:用步骤3得到的定位检测模型对待检测的票据定位手写的签名位置。本发明基于深度卷积神经网络的票据中手写签名位置定位方法,能够准确的标定出票据中含有手写签名的位置。



1. 基于深度卷积神经网络的票据中手写签名位置定位方法,其特征在于,具体按照以下步骤实施:

步骤1:搭建基于包含多种卷积神经网络模型的caffe深度学习框架的平台;

步骤2:准备票据的数据集;

步骤3:训练网络得到定位检测模型;

步骤4:用步骤3得到的定位检测模型对待检测的票据定位手写的签名位置。

2. 根据权利要求1所述的基于深度卷积神经网络的票据中手写签名位置定位方法,其特征在於,所述步骤2具体为:

步骤2.1:对票据进行拍照,得到票据的原始图像数据,并对原始图像数据进行样本扩充;

步骤2.2:对步骤2.1中得到的所有图像数据进行编号及标定,标定出票据图像中手写签名位置的坐标,得到该签名位置的左上角坐标(Xmin、Ymin)和右下角坐标(Xmax、Ymax),并且将所有图像的序号及相应的坐标位置信息写入到xml文件中;

步骤2.3:将所有图像数据分为训练数据集和测试数据集,再将训练数据集分为训练数据和验证数据。

3. 根据权利要求2所述的基于深度卷积神经网络的票据中手写签名位置定位方法,其特征在於,所述步骤2.1中对原始图像数据进行样本扩充,包括:

①对原始图像数据进行不同角度的旋转及不同方向的平移;

②对原始图像数据进行线性插值缩放;

③对原始图像数据增加不同强度的椒盐噪声、高斯噪声。

4. 根据权利要求1所述的基于深度卷积神经网络的票据中手写签名位置定位方法,其特征在於,所述步骤3具体为:

步骤3.1:将步骤2中得到的数据集中的图像调整成 $W_1 * H_1$ 大小的图像并放入ZF网络前5层层中做特征提取,输出256个 $\frac{W_1}{16} * \frac{H_1}{16}$ 大小的特征图;

步骤3.2:用3*3的卷积核与步骤3.1得到的256个特征图进行卷积,得到256维特征向量,作为RPN的第一层;

步骤3.3:将步骤3.2得到的256维特征向量输入到分类层和回归层两个并行卷积层,根据正样本中前景概率的得分高低,选取得分高的前300个候选框;

步骤3.4:使用ROI Pooling层将步骤3.3中的300个候选框映射到ZF网络的第五层卷积后得到的256维特征图上,得到池化归一后的大小6*6的特征图;

步骤3.5:将每一个6*6大小的特征图输入到两个连续的全连接层fc6、fc7,先经过fc6得到4096维特征,再将4096维特征输入到fc7中,最终得到1*4096维特征;

步骤3.6:将1*4096维特征输入到两个并行的全连接层cls_score层、bbox_predict层,cls_score层用于分类,输出背景的概率和K类样本的概率,其中,K是样本类别数;bbox_predict层用于调整候选区域位置,输出候选框的(x',y',w',h'),x'为调整后的候选框左上角横坐标,y'为调整后的候选框左上角纵坐标,w'为调整后的候选框的宽度,h'为调整后的候选框的高度;

步骤3.7:判断总迭代次数是否大于阈值,如果不大于阈值,则转到步骤3.2;如果大于

阈值,则结束。

5. 根据权利要求4所述的基于深度卷积神经网络的票据中手写签名位置定位方法,其特征在于,所述步骤3.1中数据集中的图像调整成 $W_1 * H_1$ 大小的图像需要根据数据集中图像大小 $W * H$ 来计算,即:

$$\frac{W}{H} = \frac{W_1}{H_1}。$$

6. 根据权利要求4所述的基于深度卷积神经网络的票据中手写签名位置定位方法,其特征在于,所述步骤3中训练网络时学习速率的初始值设定为 $lr=0.01$,每当当前迭代次数达到步长值的整数倍时,学习速率衰减一次,当迭代次数达到总迭代次数时,结束;衰减后的 $lr=lr * \text{gamma}$,其中, $\text{gamma}=0.1$,迭代次数 \leq 总迭代次数。

7. 根据权利要求4所述的基于深度卷积神经网络的票据中手写签名位置定位方法,其特征在于,所述步骤3中训练网络时mini-batch size设为256。

基于深度卷积神经网络的票据中手写签名位置定位方法

技术领域

[0001] 本发明属于图像定位检测技术领域,具体涉及一种基于深度卷积神经网络的票据中手写签名位置定位方法。

背景技术

[0002] 目前,我国对票据自动检测系统的研究大多还在开发阶段,实际应用较少,并且配置一套票据指定手写签名位置检测系统成本较高,也限制了票据自动化检测技术的发展。这也就对广大的研究机构和学者提出了需求,需要研究出一种票据自动化检测的技术。票据中的手写签名具有字符的特性,对于字符的识别主要有:统计特征识别技术、结构特征识别技术、基于神经网络的识别技术。统计特征有字符二维平面位置特征、字符在水平或垂直方向投影的直方图特征等,基于统计特征的字符识别技术对于形近字符区分能力弱,只适合字符粗分类。结构特征有笔划的走向、孤立的点以及是否含有闭合笔画等,这种方法便于区分字型变化大的字符。目前,对神经网络的研究正处于一个新的高峰时期,神经网络已经广泛应用于模式识别领域。随着学术界对于深度学习的研究,深度学习的算法越来越成熟,应用也越来越多。但是,大多数神经网络只能用于提取目标的特征,不能用于目标定位。

发明内容

[0003] 本发明的目的是提供一种基于深度卷积神经网络的票据中手写签名位置定位方法,能够准确的标定出票据中含有手写签名的位置。

[0004] 本发明所采用的技术方案是,基于深度卷积神经网络的票据中手写签名位置定位方法,具体按照以下步骤实施:

[0005] 步骤1:搭建基于包含多种卷积神经网络模型的caffe深度学习框架的平台;

[0006] 步骤2:准备票据的数据集;

[0007] 步骤3:训练网络得到定位检测模型;

[0008] 步骤4:用步骤3得到的定位检测模型对待检测的票据定位手写的签名位置。

[0009] 本发明的特点还在于:

[0010] 步骤2具体为:

[0011] 步骤2.1:对票据进行拍照,得到票据的原始图像数据,并对原始图像数据进行样本扩充;

[0012] 步骤2.2:对步骤2.1中得到的所有图像数据进行编号及标定,标定出票据图像中手写签名位置的坐标,得到该签名位置的左上角坐标(Xmin、Ymin)和右下角坐标(Xmax、Ymax),并且将所有图像的序号及相应的坐标位置信息写入到xml文件中;

[0013] 步骤2.3:将所有图像数据分为训练数据集和测试数据集,再将训练数据集分为训练数据和验证数据。

[0014] 步骤2.1中对原始图像数据进行样本扩充,包括:

[0015] ①对原始图像数据进行不同角度的旋转及不同方向的平移;

[0016] ②对原始图像数据进行线性插值缩放；

[0017] ③对原始图像数据增加不同强度的椒盐噪声、高斯噪声。

[0018] 步骤3具体为：

[0019] 步骤3.1:将步骤2中得到的数据集中的图像调整成 $W_1 * H_1$ 大小的图像并放入ZF网络前5层层中做特征提取,输出256个 $\frac{W_1}{16} * \frac{H_1}{16}$ 大小的特征图；

[0020] 步骤3.2:用3*3的卷积核与步骤3.1得到的256个特征图进行卷积,得到256维特征向量,作为RPN的第一层；

[0021] 步骤3.3:将步骤3.2得到的256维特征向量输入到分类层和回归层两个并行卷积层,根据正样本中前景概率的得分高低,选取得分高的前300个候选框；

[0022] 步骤3.4:使用ROI_Pooling层将步骤3.3中的300个候选框映射到ZF网络的第五层卷积后得到的256维特征图上,得到池化归一后的大小6*6的特征图；

[0023] 步骤3.5:将每一个6*6大小的特征图输入到两个连续的全连接层fc6、fc7,先经过fc6得到4096维特征,再将4096维特征输入到fc7中,最终得到1*4096维特征；

[0024] 步骤3.6:将1*4096维特征输入到两个并行的全连接层cls_score层、bbox_predict层,cls_score层用于分类,输出背景的概率和K类样本的概率,其中,K是样本类别数；bbox_predict层用于调整候选区域位置,输出候选框的 (x', y', w', h') , x' 为调整后的候选框左上角横坐标, y' 为调整后的候选框左上角纵坐标, w' 为调整后的候选框的宽度, h' 为调整后的候选框的高度；

[0025] 步骤3.7:判断总迭代次数是否大于阈值,如果不大于阈值,则转到步骤3.2;如果大于阈值,则结束。

[0026] 步骤3.1中数据集中的图像调整成 $W_1 * H_1$ 大小的图像需要根据数据集中图像大小 $W * H$ 来计算,即：

$$[0027] \quad \frac{W}{H} = \frac{W_1}{H_1}。$$

[0028] 步骤3中训练网络时学习速率的初始值设定为 $lr=0.01$,每当当前迭代次数达到步长值的整数倍时,学习速率衰减一次,当迭代次数达到总迭代次数时,结束;衰减后的 $lr = lr * \gamma$,其中, $\gamma=0.1$,迭代次数 \leq 总迭代次数。

[0029] 步骤3中训练网络时mini-batch size设为256。

[0030] 本发明的有益效果是:本发明基于深度卷积神经网络的票据中手写签名位置定位方法,采用的是深度神经网络来进行图像定位,比传统的方法在定位的速度和准确度上均有提高,准确度已经可以达到90.9%,速度基本可以做到实时定位,定位一张图片需要0.3s;并且本发明的票据图像数据库包含了各种各样的数据样本,数据样本具有多样性,使得定位的准确度有所上升。

附图说明

[0031] 图1是本发明方法中采集的正常票据图像；

[0032] 图2是本发明方法中采集的旋转45°票据图像；

[0033] 图3是本发明方法中采集的加椒盐噪声票据图像；

- [0034] 图4是本发明方法中旋转90°的待检测票据图像；
 [0035] 图5是图4的检测结果图；
 [0036] 图6是本发明方法中旋转180°的待检测票据图像；
 [0037] 图7是图6的检测结果图；
 [0038] 图8是本发明方法中旋转45°的待检测票据图像；
 [0039] 图9是图8的检测结果图；
 [0040] 图10是本发明方法中加高斯噪声的待检测票据图像；
 [0041] 图11是图10的检测结果图；
 [0042] 图12是本发明方法中加椒盐噪声的待检测票据图像；
 [0043] 图13是图12的检测结果图。

具体实施方式

[0044] 下面结合附图和具体实施方式对本发明进行详细说明。

[0045] 本发明基于深度卷积神经网络的票据中手写签名位置定位方法,具体按照以下步骤实施:

[0046] 步骤1:在ubuntu系统环境或者Windos环境下搭建基于包含多种卷积神经网络模型的caffe深度学习框架的平台;

[0047] 步骤2:准备票据的数据集,具体为:

[0048] 步骤2.1、由于在真实环境中的图像都是由用户自己在手机或者相机上拍照上传给票据系统的,所以在准备图片的时候,需要考虑不同分辨率的手机拍摄的照片,以及拍摄的环境光照等条件。本发明使用多种不同分辨率的手机对票据进行拍摄,将这部分手机拍摄的图像称为原始图像数据。为了使图像数据足够充分且能够符合各种实际情况,本发明对原始图像数据进行了样本扩充:①对原始图像数据进行不同角度的旋转及不同方向的平移;②对原始图像数据进行线性插值缩放,是考虑到不同相机采集到的图像大小不同;③对原始图像数据增加不同强度的椒盐噪声、高斯噪声;

[0049] 如图1-3给出采集到的部分票据样本图像,包括正常的票据图像、旋转的票据图像、含有噪声的票据图像。

[0050] 步骤2.2:对步骤2.1中得到的所有图像数据进行编号及标定,标定出票据图像中手写签名位置的坐标,得到该签名位置的左上角坐标(Xmin、Ymin)和右下角坐标(Xmax、Ymax),并且将所有图像序号及相应的坐标位置信息写入到xml文件中;

[0051] 步骤2.3:对所有准备好的图像数据随机分成训练数据集trainval和测试数据集test两个部分数据,其中设置trainval数据集占整个数据集的8/10,test数据集占整个数据集的2/10。在trainval数据集中又分成train数据和val数据,其中trian数据是用来做训练的,占trainval数据集的4/5,val数据是用来做验证的,占trainval数据集的1/5。

[0052] 步骤3:训练网络得到定位检测模型并对训练参数进行优化

[0053] 步骤3.1:将步骤2中得到的数据集中的图像调整成 $W_1 * H_1$ (本发明中为600*800)大小的图像并放入ZF网络前5层层中做特征提取,输出256个 $\frac{W_1}{16} * \frac{H_1}{16}$ (本发明中为37*50)大小的特征图。数据集中图像调整的大小需要根据数据集图像中图像的长宽比(图像最长边/图

像最短边)来计算,设输入图像大小为 $W*H$,调整后的图像大小为 W_1*H_1 ,关系式为:

$$[0054] \quad \frac{W}{H} = \frac{W_1}{H_1}。$$

[0055] 数据集中图像大小调整成600*800的原因:

[0056] 数据集中的图像大小不一,计算归一后图像大小时,选定占整个数据集数目较多的几种类型大小的图像,并且在数据集中图像最长边与图像最短边之比占整个数据集较多的几种类型的图像,符合条件的包括600*800、1200*1600、1500*2000、2000*2600、3000*4000,将这几个作为调整后的图像大小,调整后的图像需进行卷积计算,考虑到计算量的大小、GPU内存大小,需尽可能选取较小的调整比例,初步修改为600*800、1200*1600这两个大小的图像进行训练,选取600*800大小的图像进行训练准确度为90.84%,选取1200*1600大小的图像进行训练的准确度为90.86%,准确度后者比前者只提高了0.02%,但训练时间以及计算复杂度上后者比前者要复杂得多,因此最终选取600*800这个比例作为调整的比例。

[0057] 步骤3.2:用3*3的卷积核(滑动窗口)与步骤3.1得到的256个特征图进行卷积,因为这个3*3的区域上,每一个特征图上得到一个1维向量,256个特征图上即可得到256维特征向量,作为RPN(region proposal network,区域候选网络)的第一层;

[0058] 3*3滑窗中心点位置,对应预测输入图像的3种尺度(128、256、512)、3种长宽比(1:2、2:1、7:1)的目标区域候选框,这种映射的机制称为锚点,产生了 $k=9$ 个锚点。即每个3*3区域可以产生9个目标区域候选框。所以对于本发明中的37*50的特征图,总共有约20000(37*50*9)个锚点,也就是在输入图像上预测了20000个目标区域候选框。

[0059] 目标区域候选框的尺度有三个(128、256、512),这个实质是指候选框的面积为128*128、256*256、512*512,这个候选框面积大小和步骤3.1中图像调整归一化后的大小有关,尽可能候选框的面积要将图像中目标包围在内。

[0060] 目标区域候选框的长宽比选用1:2、2:1和7:1这三个比例是根据数据集中每个图像内目标候选框的宽(W_{box})和高(H_{box})的比例得出,选取宽高比例图像数最多的三个比例作

为候选框的宽高比 $\frac{W_{\text{box}}}{H_{\text{box}}}$,其中, $W_{\text{box}} = X_{\text{max}} - X_{\text{min}}$, $H_{\text{box}} = Y_{\text{max}} - Y_{\text{min}}$, $(X_{\text{min}}, Y_{\text{min}})$ 、 $(X_{\text{max}},$

$Y_{\text{max}})$ 为步骤2.2中标定的手写签名位置的左上角及右下角坐标值。

[0061] 步骤3.3:将256维特征向量输入到两个并行卷积层,即分类层和回归层,分别用于分类和边框回归。就局部来说,这两层是全连接网络;就全局来说,由于网络在所有位置(共37*50个)的参数相同,所以实际用尺寸为1*1的卷积网络实现。需要注意的是:并没有显式地提取任何候选窗口,完全使用网络自身完成判断和修正。

[0062] 对每个候选框,分类层从256维特征中输出属于前景和背景的概率,并对每个候选框进行标定:正样本为与真实区域重叠大于0.7,负样本为与真实区域重叠小于0.3,保留正样本;

[0063] 同时回归层从256维特征中输出4个平移缩放参数(x, y, w, h),其中, x 为候选框左上角横坐标, y 为候选框左上角纵坐标, w 为候选框的宽度, h 为候选框的高度,这四个坐标元素用于确定目标位置。

[0064] 经过步骤3.2,本发明在输入图像上预测得到20000个候选框,经过步骤3.3后,20000个预测候选框剩下2000个左右候选框,最后根据正样本中前景概率的得分高低,选取

得分高的前300个候选框。

[0065] 步骤3.4:使用ROI_Pooling层将步骤3.3中的300个候选框映射到ZF网络的第五层卷积后得到的256维特征图上,得到池化归一后的大小6*6的特征图。

[0066] ROI_Pooling层就是实现从原图区域映射到conv5区域最后池化到固定大小的功能。

[0067] 首先计算预测的候选框映射到特征图上的坐标,即原始坐标乘以十六分之一,然后针对每个输出来进行计算,即将特征图上已经映射好的300个不同大小的候选框进行池化,然后将池化后的结果统一归一化为大小6*6的特征图。

[0068] 步骤3.5、将每一个6*6大小的特征图(一共300个特征图)输入到两个连续的全连接层fc6、fc7,这两个全连接层是连续的不是并行的,先经过fc6得到4096维特征,再将4096维特征输入到fc7中,最终得到1*4096维特征。

[0069] 步骤3.6:将1*4096维特征输入到两个并行的全连接层cls_score层、bbox_predict层,cls_score层用于分类,输出背景的概率和K类样本的概率,其中,K是样本类别数;bbox_predict层用于调整候选区域位置,输出候选框的(x',y',w',h'),x'为调整后的候选框左上角横坐标,y'为调整后的候选框左上角纵坐标,w'为调整后的候选框的宽度,h'为调整后的候选框的高度;

[0070] 步骤3.7:判断总迭代次数(本发明中为8000)是否大于阈值,如果不大于阈值,则转到步骤3.2;如果大于阈值,则结束。

[0071] 总迭代次数的选择:在训练过程中观察loss的值,当loss的值不再大幅度下降、趋于稳定,我们就可以选择当前的迭代次数最为最终的迭代次数。

[0072] 在训练网络时学习速率的初始值设定为lr=0.01,每当当前迭代次数达到步长值(本发明为6000)的整数倍时,学习速率衰减一次,当迭代次数达到总迭代次数时,结束;衰减后的lr=lr*gamma,其中,gamma=0.1,迭代次数≤总迭代次数。

[0073] 运用梯度下降算法进行优化时,权重的更新规则中,在梯度项前会乘以一个系数,这个系数就叫学习速率lr。如果学习速率太小,则会使网络收敛过慢,如果学习速率太大,则会导致代价函数振荡,一个比较好的策略是在实践中先把学习速率设置为0.01,然后观察training cost的走向,如果training cost在减小,那可以逐步地调大学习速率,如0.1,1.0…。如果training cost在增大,那就减小学习速率,如0.001,0.0001…。经过上述方法确定学习速率的值。

[0074] 学习速率什么时候衰减与步长有关,减少多少与gamma有关。选择步长时,可以尽可能的接近总迭代次数。

[0075] 训练网络时mini-batch size设为256,采用mini-batch时的权重更新规则为:

$$[0076] \quad w \rightarrow w' = w - \eta \frac{1}{256} \sum_x \nabla C_x$$

[0077] 也就是将256个样本的梯度求均值。

[0078] 当采用mini-batch时,我们可以将一个batch里的所有样本放在一个矩阵里,利用线性代数库来加速梯度的计算,这是工程实现中的一个优化方法。

[0079] 一个大的batch size,可以利用矩阵、线性代数库来进行计算的加速,batch size越小,则加速效果可能越不明显。batch size不是越大越好,太大了,权重的更新就会

不那么频繁,导致优化过程太漫长。一般batch size大小为256,若是图片数据集不大、GPU内存4G以下,可以考虑将batch size改小。

[0080] 步骤4:用步骤3得到的定位检测模型对待检测的票据定位手写的签名位置,即包括手写签名位置的矩形框的左上角坐标和右下角坐标。

[0081] 图4-图13是带有不同旋转角度和不同噪声类型的待检测的票据图像以及使用本发明方法定位处的手写签名位置在该图中用矩形框标记图。可以看出,本发明方法的鲁棒性非常好,可以克服旋转、噪声等情况对定位结果的影响,具有定位准确、定位速度快的特点。



图1

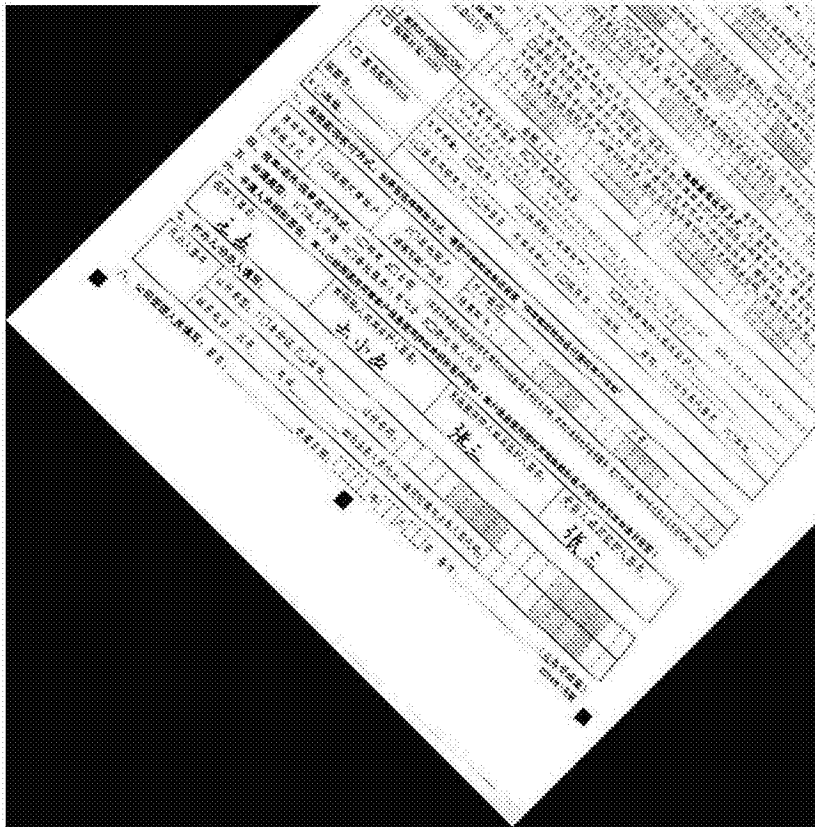


图2

Figure 3 shows a form with several sections:

- Section 1:** Contains three checkboxes: 选项1, 选项2, 选项3.
- Section 2:** A table with columns for '姓名' (Name) and '性别' (Gender). It lists '张三' (Zhang San) and '李四' (Li Si).
- Section 3:** A table with columns for '姓名' (Name) and '性别' (Gender). It lists '张三' (Zhang San) and '李四' (Li Si).
- Section 4:** A table with columns for '姓名' (Name) and '性别' (Gender). It lists '张三' (Zhang San) and '李四' (Li Si).
- Section 5:** A table with columns for '姓名' (Name) and '性别' (Gender). It lists '张三' (Zhang San) and '李四' (Li Si).
- Section 6:** A table with columns for '姓名' (Name) and '性别' (Gender). It lists '张三' (Zhang San) and '李四' (Li Si).

图3

Figure 4 shows a vertical form with a barcode at the top and several rows of data. It includes a signature '张三' (Zhang San).

图4

Figure 5 shows a form with a signature '张三' (Zhang San) and a value 'Sign : 0.998'. It includes a barcode and various data fields.

图5

Figure 6 shows a form with a signature '张三' (Zhang San) and a value 'Sign : 0.998'. It includes a barcode and various data fields.

图6

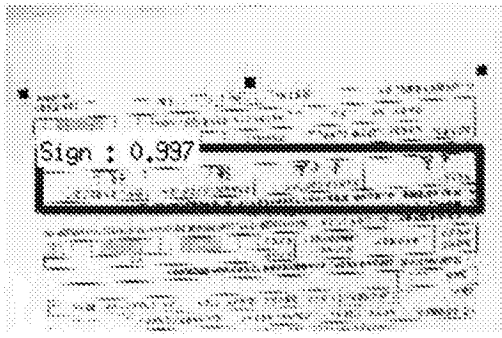


图7

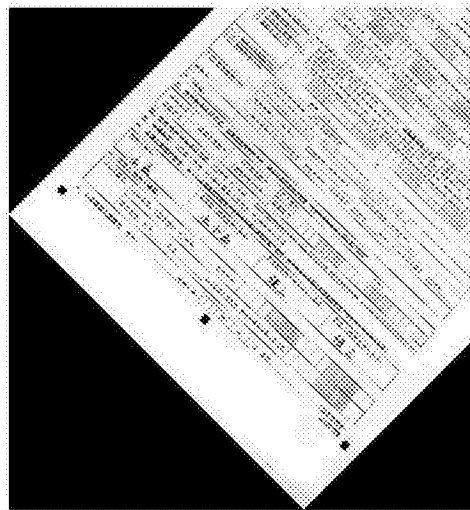


图8

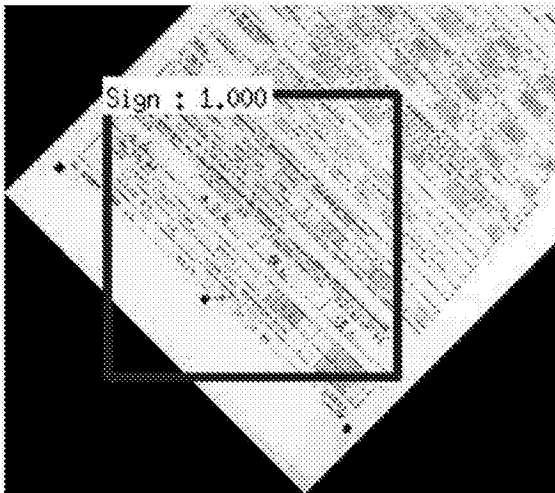


图9

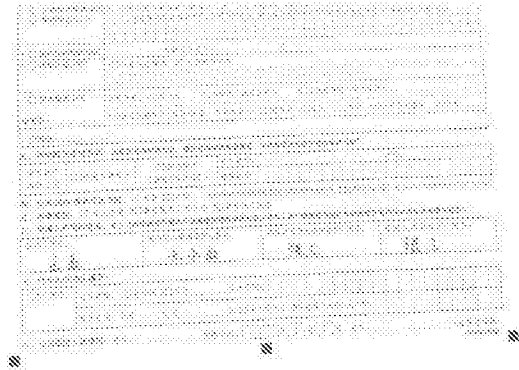


图10

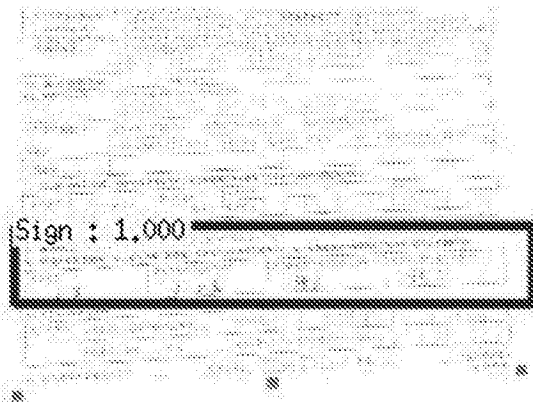


图11

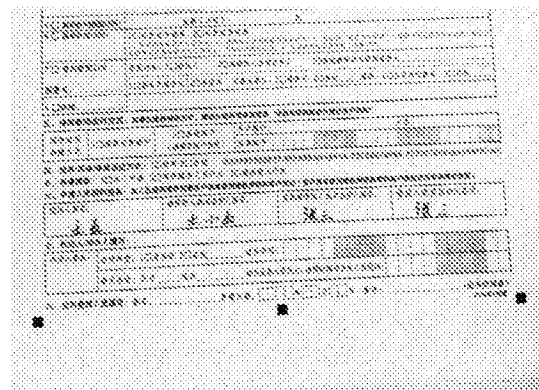


图12

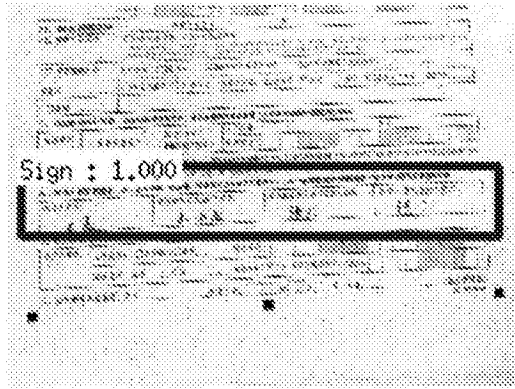


图13