



(12) 发明专利

(10) 授权公告号 CN 111898606 B

(45) 授权公告日 2023.04.07

(21) 申请号 202010422410.0

(22) 申请日 2020.05.19

(65) 同一申请的已公布的文献号  
申请公布号 CN 111898606 A

(43) 申请公布日 2020.11.06

(73) 专利权人 武汉东智科技股份有限公司  
地址 430000 湖北省武汉市武昌区友谊大道6号广达大厦附楼四楼

(72) 发明人 聂晖 杨小波 李军

(74) 专利代理机构 北京众达德权知识产权代理有限公司 11570  
专利代理师 刘杰

(51) Int. Cl.  
G06V 30/148 (2022.01)  
G06V 10/82 (2022.01)  
G06N 3/0442 (2023.01)  
G06N 3/0464 (2023.01)  
G06N 3/08 (2023.01)

(56) 对比文件

CN 111027562 A, 2020.04.17  
CN 102663675 A, 2012.09.12  
CN 109344831 A, 2019.02.15  
CN 110781881 A, 2020.02.11  
CN 110533030 A, 2019.12.03  
CN 110378334 A, 2019.10.25  
LING-QUN ZUO 等.Natural Scene Text Recognition Based on Encoder-Decoder Framework.《IEEE Access》.2019,  
Dinh Viet Sang 等.Improving CRNN with EfficientNet-like feature extractor and multi-head attention for text recognition.《SoICT 2019》.2019,  
徐富勇.基于注意力增强网络的场景文字识别.《现代计算机》.2020, (第09期), (续)

审查员 党英霞

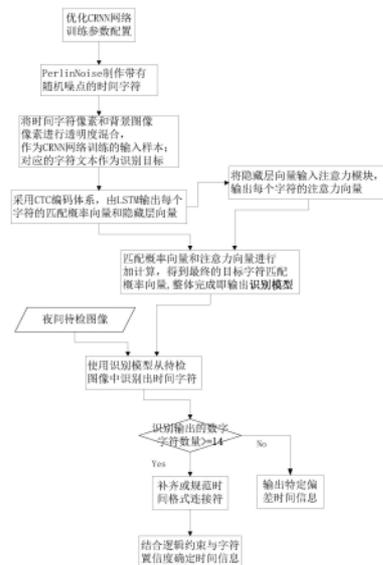
权利要求书1页 说明书6页 附图3页

(54) 发明名称

一种视频图像中叠加透明时间字符的夜间成像识别方法

(57) 摘要

本发明属于计算机视觉领域,涉及一种夜间视频图像中透明时间标注信息的识别方法,包括:改进CRNN的训练过程配置;制作透明时间字符叠加夜间背景图像作为训练输入样本;引入二维注意力机制模块训练出适合夜间透明时间字符的识别模型;使用识别模型从夜间待检图像中输出时间识别结果并进行逻辑校验。本发明针对主流摄像机设备上透明时间标注信息的识别需求,使用夜间成像环境,初步提升透明时间字符与背景图像的差异性,并进一步增强透明时间字符的特征提取能力,构造了自然场景‘无衬底’叠加透明特征文本的识别模型,实现了夜间成像环境下、具备特征加权抽取能力的透明时间字符识别方法。



CN 111898606 B

[接上页]

(56) 对比文件

和文杰 等. 基于Attention-CTC的自然场景

文本识别算法.《电子科技》.2019,第32卷(第12期),

1. 一种夜间图像中叠加透明时间字符的字符识别方法,其特征在于,包括以下步骤:
  - 步骤i、改进CRNN通用文本识别网络的训练方法;
  - 步骤ii、制作CRNN夜间图像透明时间字符训练样本;
  - 步骤iii、引入注意力模块强化训练夜间图像透明时间字符识别模型;
  - 步骤iv、识别待检夜间图像中的透明时间字符并校验输出值的合理性;所述步骤i,改进CRNN通用文本识别网络的训练方法的具体步骤包括:
  - 1-1) 在主干ResNet分类网络的特征提取环节,采用3层采样,以保留字符的更多横向特征;
  - 1-2) 在提取字符高度特征时,使用最大值池化;
  - 1-3) 选择单层LSTM,将其输出的隐藏层向量,用于注意力向量计算;所述步骤iii、引入注意力模块强化训练夜间图像透明时间字符识别模型的具体步骤包括:
  - 3-1) 采用CTC编码体系,记录从特征图解码出的每一个时间字符特征与其目标时间字符文本的匹配概率向量;
  - 3-2) 在解码输出匹配概率向量的同时,将LSTM内部的隐藏层向量叠加到处理过的特征图上生成一个注意力权重图,对相应位置时间字符进行特征增强;
  - 3-3) 在注意力模块中,将注意力权重图降维成注意力向量,和匹配概率向量进行加计算,得到最终的目标字符匹配概率向量;
  - 3-4) 当匹配概率整体最优时,即得到时间字符识别模型。
2. 根据权利要求1所述的夜间图像中叠加透明时间字符的字符识别方法,其特征在于,所述步骤ii,制作CRNN夜间图像透明时间字符训练样本的具体步骤包括:
  - 2-1) 在背景透明图像上,使用字符像素渲染函数在其上生成黑白颜色、多种字体、多种时间格式的时间字符;
  - 2-2) 使用PerlinNoise噪点图对上述时间字符笔画像素生成随机斑点;
  - 2-3) 对带有随机斑点的时间字符进行透明度处理,叠加到随机生成的夜间背景图像上作为训练的输入样本;
  - 2-4) 取2-1)中时间字符的文本形式,将其中非标准时间格式的时间字符进行忽略和强制替换处理,作为与2-3)匹配的识别目标样本。
3. 根据权利要求2所述的夜间图像中叠加透明时间字符的字符识别方法,其特征在于,所述步骤iv,识别夜间待检图像中的时间字符并校验输出值的合理性的具体步骤包括:
  - 4-1) 在夜间时段采集视频监控设备的待检图像,输入识别模型;
  - 4-2) 当识别输出的数字字符数量不足标准的时候,判定为无法校验,强制转化为特定时间值;
  - 4-3) 对识别输出的数字字符数量超出标准的,补齐或删除多余的时间连接符号‘-和:’;
  - 4-4) 按照‘年月日’和‘时分秒’的逻辑约束,并结合3-3)中记录的匹配概率,完成时间分割和修正,转化为最终时间值。

## 一种视频图像中叠加透明时间字符的夜间成像识别方法

### 技术领域

[0001] 本发明属于计算机视觉领域,可用于公安及相关行业视频监控系统画面中叠加时间字符的检测。尤其涉及一种夜间视频图像中透明时间标注信息的识别方法。

### 背景技术

[0002] 在针对视频监控图像中叠加时间字符的识别方案设计过程中,发现某些主流品牌的摄像机在叠加字符时可能采用笔画透明(即非黑非白的灰色)显示的风格,这给本已面临诸多困难的‘无衬底’背景的字符识别带来了新的挑战。有实验数据显示,在相同的自然场景下,使用字符笔画透明风格比不透明(纯黑色或纯白色)风格,其识别准确率下降20%以上,严重影响了方案的预期应用效果。

[0003] 基于深度学习技术的图像中叠加时间字符的识别算法,为了获得较理想的效果,一般都会假定应用场景的约束条件——即在图像中采用规范的、纯黑色或纯白色、非透明风格标注叠加的字符。由于透明字符固有的低对比度属性,极易与图像的自然背景产生混淆,而变得更加不易辨认,直接影响训练模型的性能。

[0004] 通过实验数据分析,单纯依靠算法训练来提升笔画透明字符的识别准确率,已经遇到了瓶颈。

### 发明内容

[0005] 本发明所要解决的技术问题是,针对笔画透明时间字符的识别需求,提供一种夜间视频图像中透明时间字符的识别方法。

[0006] 本发明的基本技术构思是,改进CRNN(一种基于图像序列的通用文本识别神经网络)的训练过程配置;制作透明时间字符叠加夜间背景图像作为训练输入样本;引入二维注意力机制模块训练出适合夜间透明时间字符的识别模型;使用识别模型从夜间待检图像中输出时间识别结果并进行逻辑校验。

[0007] 为解决上述技术问题,本发明提出一种满足夜间成像中透明时间字符的识别方法,包括以下步骤:

[0008] 步骤i、改进CRNN通用文本识别网络的训练方法;

[0009] 步骤ii、制作CRNN夜间透明时间字符训练样本;

[0010] 步骤iii、引入注意力模块强化训练夜间透明时间字符识别模型;

[0011] 步骤iv、识别夜间待检图像中的透明时间字符并校验输出值的合理性。

[0012] 优选的,所述步骤i,改进CRNN通用文本识别网络的训练方法的具体步骤包括:

[0013] 1-1) 在主干ResNet分类网络(一种主流CNN卷积网络)的特征提取环节,采用3层采样,以保留字符的更多横向特征;

[0014] 1-2) 在提取字符高度特征时,使用最大值池化;

[0015] 1-3) 选择单层LSTM(长短时记忆网络),将其输出的隐藏层向量,用于注意力向量计算。

- [0016] 优选的,所述步骤ii,制作CRNN夜间透明时间字符训练样本的具体步骤包括:
- [0017] 2-1) 在背景透明图像上,使用字符像素渲染函数在其上生成黑白颜色、多种字体、多种时间格式的时间字符;
- [0018] 2-2) 使用PerlinNoise噪点图对上述时间字符笔画像素生成随机斑点;
- [0019] 2-3) 对带有随机斑点的时间字符进行透明度处理,叠加到随机的夜间背景图像上作为训练的输入样本;
- [0020] 2-4) 取2-1)中时间字符的文本形式,将其中非标准时间格式的字符进行忽略和强制替换处理,作为与2-3)匹配的识别目标样本;
- [0021] 优选的,所述步骤iii、引入注意力模块强化训练夜间透明时间字符识别模型的具体步骤包括:
- [0022] 3-1) 采用CTC编码体系,记录从特征图解码出的每一个时间字符特征与其目标时间字符文本的匹配概率向量;
- [0023] 3-2) 在解码输出匹配概率向量的同时,将LSTM内部的隐藏层向量叠加到处理过的特征图上生成一个注意力权重图,对相应位置时间字符进行特征增强;
- [0024] 3-3) 在注意力模块中,将注意力权重图降维成注意力向量,和匹配概率向量进行简单加计算,得到最终的目标字符匹配概率向量;
- [0025] 3-4) 当匹配概率整体最优时,即得到时间字符识别模型。
- [0026] 优选的,所述步骤vi,识别夜间待检图像中的时间字符并校验输出值的合理性的具体步骤包括:
- [0027] 4-1) 在夜间时段采集视频监控设备的待检图像,输入识别模型;
- [0028] 4-2) 当识别输出的数字字符数量不足标准的时候,判定为无法校验,强制转化为特定时间值;
- [0029] 4-3) 对识别输出的数字字符数量超出标准的,补齐或删除多余的时间连接符号‘-和:’;
- [0030] 4-4) 按照‘年月日’和‘时分秒’的逻辑约束,并结合3-3)中记录的匹配概率,完成时间分割和修正,转化为最终时间值。
- [0031] 至此,完成CRNN夜间文字识别模型的训练,输入夜间待检图像,得到时间字符识别结果,实现本发明的技术方案。
- [0032] 本发明的有益效果包括:
- [0033] 1) 针对主流摄像机设备上一种透明时间标注信息的识别需求,实现了自然场景‘无衬底’叠加透明特征文本的识别模型,形成了专用解决方案。
- [0034] 2) 将对透明时间字符的特征学习问题,转向在特定成像模式下,进一步提升字符与背景区分度的新思路,为类似应用领域的识别需求提供了一种关键技术。

## 附图说明

- [0035] 下面结合附图和具体实施方式对本发明的技术方案作进一步具体说明。
- [0036] 图1为本发明方法的基本流程图。
- [0037] 图2为CRNN网络夜间识别模型的训练输入样本。
- [0038] 图3为夜间待检图像中代表性时间字符的显示效果图。

[0039] 图4为引入注意力机制增强字符特征提取的示意图。

[0040] 图5为识别结果的时间字符格式标准化示例图样,其中5(a)表示补充未显现的“:”符号的时间字符格式示意图,5(b)表示删除掉多余的“:”符号的时间字符格式示意图。

[0041] 图6为识别结果的时间字符逻辑校验示例图样。

### 具体实施方式

[0042] 为了使本发明的目的、技术方案及优点更加清楚明白,以下结合附图及实施例,对本发明进行进一步详细说明。应当理解,此处所描述的具体实施例仅用以解释本发明,并不用于限定本发明。

[0043] 如图1所示,本发明提出一种满足夜间成像中透明时间字符的识别方法的整体流程图,主要步骤如下:

[0044] 步骤i、自定义CRNN文本识别网络结构;

[0045] 步骤ii、制作夜间透明时间字符训练样本;

[0046] 步骤iii、引入注意力模块强化训练识别模型;

[0047] 步骤iv、识别夜间待检图像中的透明时间字符并校验输出值的合理性。

[0048] 更具体的,步骤i包括如下细分步骤:

[0049] 1-1) 在主干ResNet分类网络(一种主流CNN卷积网络)的特征提取环节,采用3层采样设置,以保留字符的更多横向特征;

[0050] 本实施例中,3层采样从512像素宽度的原始图像,得到64像素宽度特征图。

[0051] 1-2) 在提取字符高度特征时,使用最大值池化;

[0052]  $mf = \max(\{f_{0,y} | y \in \{0, 1, 2, 3\}\})$

[0053] 从高度方向上的4个特征值中选取最大的一个特征,提升字符识别精度。

[0054] 1-3) 选择单层LSTM(长短时记忆网络),将其输出的隐藏层向量,用于注意力向量计算。

[0055] 步骤ii包括如下细分步骤:

[0056] 2-1) 在背景透明图像上,使用字符像素渲染函数在其上生成黑白颜色、多种字体、多种时间格式的时间字符;

[0057]  $\Omega_i = \{0, 1, \dots, K\}$

[0058] 绘制K个字符,其中 $\Omega_i$ 代表第i个字符将要绘制的像素范围(仅包含该字符的笔画像素), $\Omega_0$ 代表除字符像素之外的背景像素。

[0059] 
$$I = \frac{1}{3}(R + G + B)$$

[0060] 
$$D(i) = \frac{1}{|\Omega_i|} \sum_{(x,y) \in \Omega_i} I[x,y], \quad i \in \{1, \dots, K\}$$

[0061]  $I[x,y]$ 表示图像中像素 $[x,y]$ 的RGB平均亮度值, $\Omega_i$ 表示第i个字符所覆盖的背景像素数量, $D(i)$ 表示第i个字符的所覆盖的背景像素的平均亮度。

[0062] 字符生成的渲染函数如下:

$$[0063] \quad \mathbf{f}[x, y] = \begin{cases} [0, 0, 0, 0], & \text{if } (x, y) \in \Omega_0 \\ [0, 0, 0, 1], & \text{elif } (x, y) \in \Omega_1 \wedge D(i) > 0.7 \\ [1, 1, 1, 1], & \text{otherwise} \end{cases}$$

[0064] 使用RGBA编码, [0, 0, 0, 0]代表一个无色透明像素, [0, 0, 0, 1]代表一个黑色像素, [1, 1, 1, 1]代表一个白色像素。

[0065] 2-2) 使用PerlinNoise噪点图对上述时间字符笔画像素生成随机斑点;

$$[0066] \quad \mathbf{f}' = \mathbf{f} + \frac{1}{k}(1 - \eta)(0.3 - \mathbf{f})$$

[0067] 其中 $\eta$ 是生成的噪点图,  $\eta[x, y] \in \{0, 1\}$ 。

[0068] 衰减控制系数 $k > 0$ ,  $k$ 越大, 衰减的效果越明显;

[0069] 本实施例中 $k = 8$ 。

[0070] 2-3) 对带有随机斑点的时间字符进行透明度处理, 叠加到随机的夜间背景图像上作为训练的输入样本;

[0071] 把时间字符像素和背景像素 $b$ 进行透明度混合, 使用公式:

$$[0072] \quad \mathbf{f}'' = (1 - \lambda\alpha)ob + \lambda\mathbf{f}'$$

[0073] 其中 $\alpha$ 代表 $\mathbf{f}$ 透明度通道中值为1的字符图像,  $\lambda$ 为混合参数,  $\lambda$ 越小, 输出图像中字符越透明不可见;

[0074] 本实施例中 $\lambda = 0.9$ ,

[0075] 如图2所示, 训练样本时间字符笔画透明且带有的残缺效果, 模拟了实际应用中, 如图3所示, 待检图像透明时间字符笔画与背景像素混淆的情况, 典型的见: '0'、'3'、'5'的部分笔画区域。

[0076] 2-4) 取2-1)中时间字符的文本形式, 将其中非标准时间格式的字符进行忽略和强制替换处理, 作为与2-3)匹配的识别目标样本;

$$[0077] \quad \Gamma = \{ "0", "1", "2", "3", "4", "5", "6", "7", "8", "9", "-", ":" \}$$

$$[0078] \quad \Gamma_1 = \{ "年", "月" \}$$

$$[0079] \quad \Gamma_2 = \{ "时", "分" \}$$

$$[0080] \quad \Gamma_3 = U - \Gamma - \Gamma_1 - \Gamma_2$$

[0081]  $\Gamma$ 表示支持的标准字符集合,  $\Gamma_1$ 表示应该被替换成“-”的字符集合,  $\Gamma_2$ 表示应该被替换成“:”的字符集合,  $\Gamma_3$ 表示所有不支持的字符集合, 包括中文字符。

[0082] 字符替换公式如下:

$$[0083] \quad s' = \text{replace}(s, \Gamma_1, "-")$$

$$[0084] \quad s'' = \text{replace}(s', \Gamma_2, ":")$$

$$[0085] \quad s''' = \text{delete}(s'', \Gamma_3)$$

[0086] 步骤iii包括如下细分步骤:

[0087] 3-1) 采用CTC编码体系, 记录从特征图解码出的每一个时间字符特征与其目标时间字符文本的匹配概率向量;

$$[0088] \quad p[i, c]$$

[0089] 其中,  $i$ 表示输出的预测字符串第 $i$ 个字符,  $c$ 表示第 $i$ 个字符预测为字符 $c$ 的概率,

例如p[3,“8”]表示第3个字符预测是字符“8”的概率。

[0090] 3-2) 在解码输出匹配概率向量的同时,将LSTM内部的隐藏层向量叠加到处理过的特征图上生成一个注意力权重图,对相应位置时间字符进行特征增强;

$$[0091] \quad \begin{cases} e_{i,j} = \tanh \left( W_v v_{i,j} + \sum_{p,q \in N_{i,j}} \hat{W}_{p-i,q-j} \cdot v_{p,q} + W_h h_t \right) \\ a_{i,j} = \text{softmax} (w_e^T \cdot e_{i,j}) \end{cases}$$

[0092] 其中, $W_v$ 、 $\hat{W}$ 、 $W_h$ 、 $W_e^T$ 表示需要学习的线性变换, $v$ 表示特征图, $N_{i,j}$ 表示特征图中 $(i,j)$ 坐标附近的八个相邻域,该八个相邻域即九宫格图中中心格外围的八个宫格区域, $h_t$ 表示第 $t$ 个字符对应的隐藏层向量; $a_{i,j}$ 是注意力权重图中 $(i,j)$ 坐标对应的注意力向量,所有的 $a_{i,j}$ 组成的整体即注意力权重图 $a$ 。

[0093] 如图4所示,虚线框内的处理部分。

[0094] 3-3) 在注意力模块中,将注意力权重图降维成注意力向量,和匹配概率向量进行简单加计算,得到最终的目标字符匹配概率向量;

$$[0095] \quad h'_t = \sum_{i,j} h_t + a_{i,j}$$

[0096] 如图4所示,Attention模块输出注意力向量。

[0097] 3-4) 当匹配概率整体最优时,即得到时间字符识别模型。

[0098] 步骤iv包括如下细分步骤:

[0099] 4-1) 在夜间时段采集视频监控设备的待检图像,输入识别模型;

[0100] 本实施例中,夜间时段是指当日20:00——次日04:00。

[0101] 4-2) 当识别输出的数字字符数量不足标准14(标准格式YYYY-MM-DD hh:mm:ss中数字应达到14个)的时候,判定为无法校验,强制转化为特定时间值;

[0102] 使用3-4)中的模型对夜间待检图像进行识别,输出字符集合 $s$ ,

[0103]  $z = |\text{delete}(s, U-N)|$

[0104]  $Z$ 表示识别输出的数字字符数量,其中, $N$ 代表数字字符集合, $U-N$ 即除了数字之外的所有字符的集合。

[0105] 当 $z < 14$ 时,判定为无法校验,强制转化为特定时间值,本实施例统一指定为时间信息“2000-01-01 00:00:00”输出。

[0106] 4-3) 对达到14个以上数字输出的,补齐或删除多余的时间连接符号“-和:”;

[0107]  $s' = \text{pre}(s, c_{\text{dash}}, 2)$

[0108] 删除多余的“-”连接符,其中 $\text{pre}$ 函数表示删除字符,仅保留最前的2个指定字符。

$$[0109] \quad s = \begin{cases} \text{ist}(s, c_{\text{dash}}, 7), & \text{if } d(s[7]) \wedge s[4] == c_{\text{dash}}, \\ \text{ist}(s, c_{\text{dash}}, 4), & \text{elif } d(s[4]) \wedge s[7] == c_{\text{dash}}, \\ \text{ist}(\text{ist}(s, c_{\text{dash}}, 4), c_{\text{dash}}, 7), & \text{otherwise} \end{cases}$$

[0110] 补齐“-”连接符,其中 $\text{ist}$ 表示在字符串 $s$ 的指定序号插入字符, $c_{\text{dash}}$ 表示“-”字符,

d(c) 函数判断字符c是否为数字。

[0111] 同理，

[0112]  $s' = \text{re}(\text{pre}(\text{re}(s), c_{\text{colon}}, 2))$

[0113] 删除多余“:”字符，其中re函数表示将字符串倒序，将字符串顺序颠倒再恢复，即可保留最后两个“:”字符， $c_{\text{colon}}$ 表示“:”字符。

[0114]  $m = \text{re}(s)$

[0115] 
$$n = \begin{cases} \text{ist}(m, c_{\text{colon}}, 5), & \text{if } d(m[5]) \wedge m[2] = c_{\text{colon}}, \\ \text{ist}(m, c_{\text{colon}}, 2), & \text{elif } d(m[2]) \wedge m[4] = c_{\text{colon}}, \\ \text{ist}(\text{ist}(m, c_{\text{colon}}, 2), c_{\text{colon}}, 5), & \text{otherwise} \end{cases}$$

[0116]  $s' = \text{re}(n)$

[0117] 补齐“:”连接符，其中两次倒序操作，即以最后的“:”为基准来补齐的。

[0118] 如图5所示，其中5(a)表示补齐‘:’号后的，5(b)表示删除‘:’号。

[0119] 4-4) 按照‘年月日’和‘时分秒’的逻辑约束，并结合3-3)中得到的匹配概率，完成时间分割和修正，转化为最终时间值。

[0120]  $N_1 \in \{c_1 | (10c_1 + c_2) \in [a, b], c_1, c_2 \in \{0, 1, \dots, 9\}\}$

[0121]  $\hat{c}_1 = \arg \max_{c_1 \in N_1} (p[i, c_1])$

[0122]  $N_2 \in \{c_2 | (10\hat{c}_1 + c_2) \in [a, b], c_2 \in \{0, 1, \dots, 9\}\}$

[0123]  $\hat{c}_2 = \arg \max_{c_2 \in N_2} (p[i+1, c_2])$

[0124] 二位数字的修正方法，其中，a、b表示二位数字的范围， $c_1 c_2$ 表示当前预测的十位和个位上的数字， $p[i, c_1]$ 表示预测第i位字符为 $c_1$ 的概率情况，修正过程就是在这个合理的数字集合 $N_1$ 中，重新找到概率最大的预测字符作为修正结果。当 $c_1$ 修正完毕之后，以 $\hat{c}_1$ 为正确的十位，确定合理数字集合 $N_2$ ，再进行 $c_2$ 的修正。

[0125] 如图6所示，经过校验后，最终输出正确的时间值。

[0126] 至此，完成夜间文字识别模型的训练，输入夜间待检图像，得到时间字符识别结果，实现本发明的技术方案。

[0127] 本实施例中，对8215张夜间随机图像识别结果统计，获得了91.86%的准确度(全匹配率)。

[0128] 使用本方法达到了比较理想的应用效果。

[0129] 本领域技术人员清楚，上述参数或阈值的具体数值可根据样本训练方法和规范执行的严格程度进行调整，不构成对本发明的限定。

[0130] 最后需要说明的是，以上所述仅为本发明的优选实施例而已，并不用于限制本发明，尽管按照前述实施例对本发明进行了详细的说明，对于本领域的技术人员来说，其依然可以对前述实施例所记载的技术方案进行修改，或者对其中部分技术特征进行等同替换。凡在本发明的精神和原则之内，所作的任何修改、等同替换、改进等，均应包含在本发明的保护范围之内。

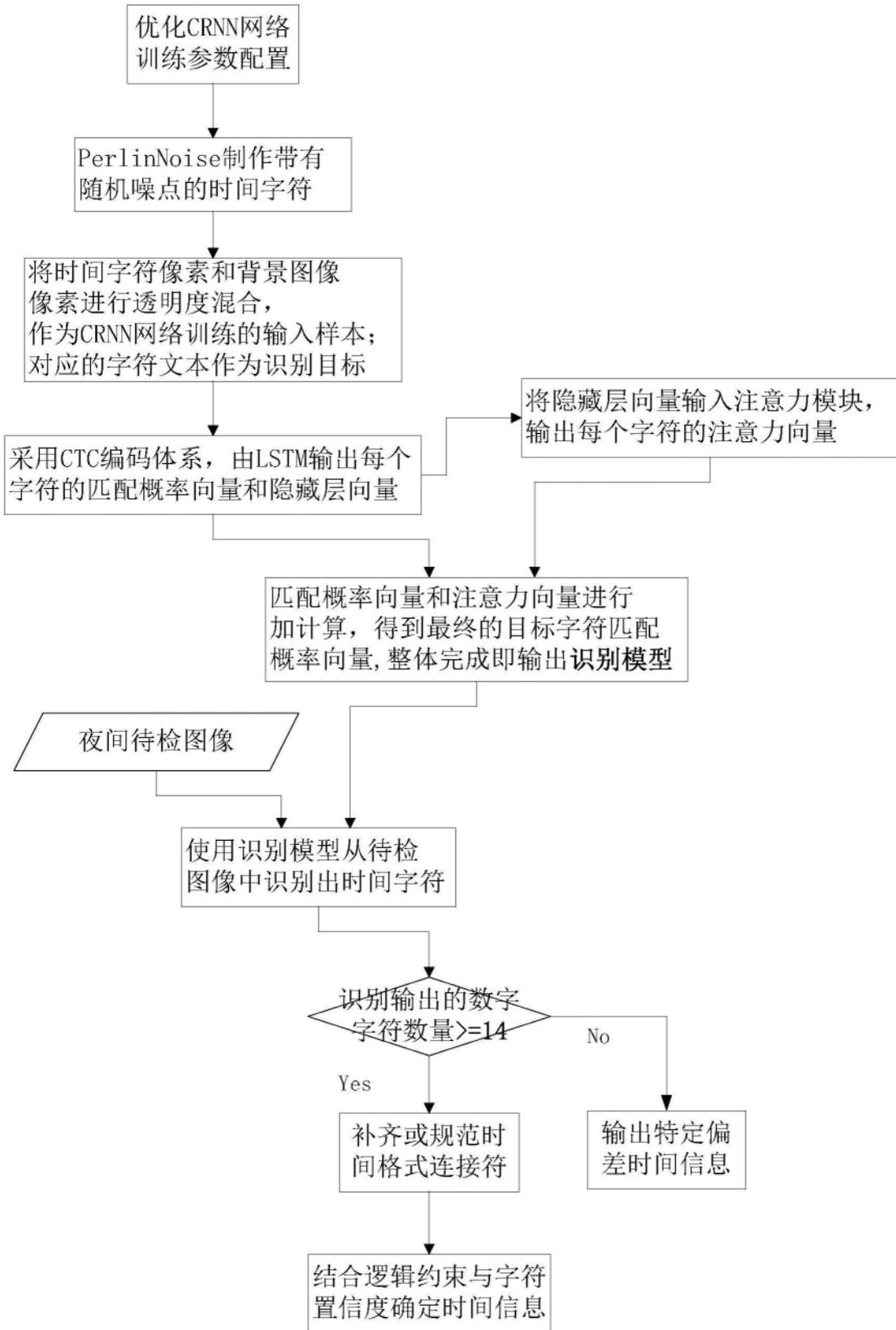


图1



图2



图3

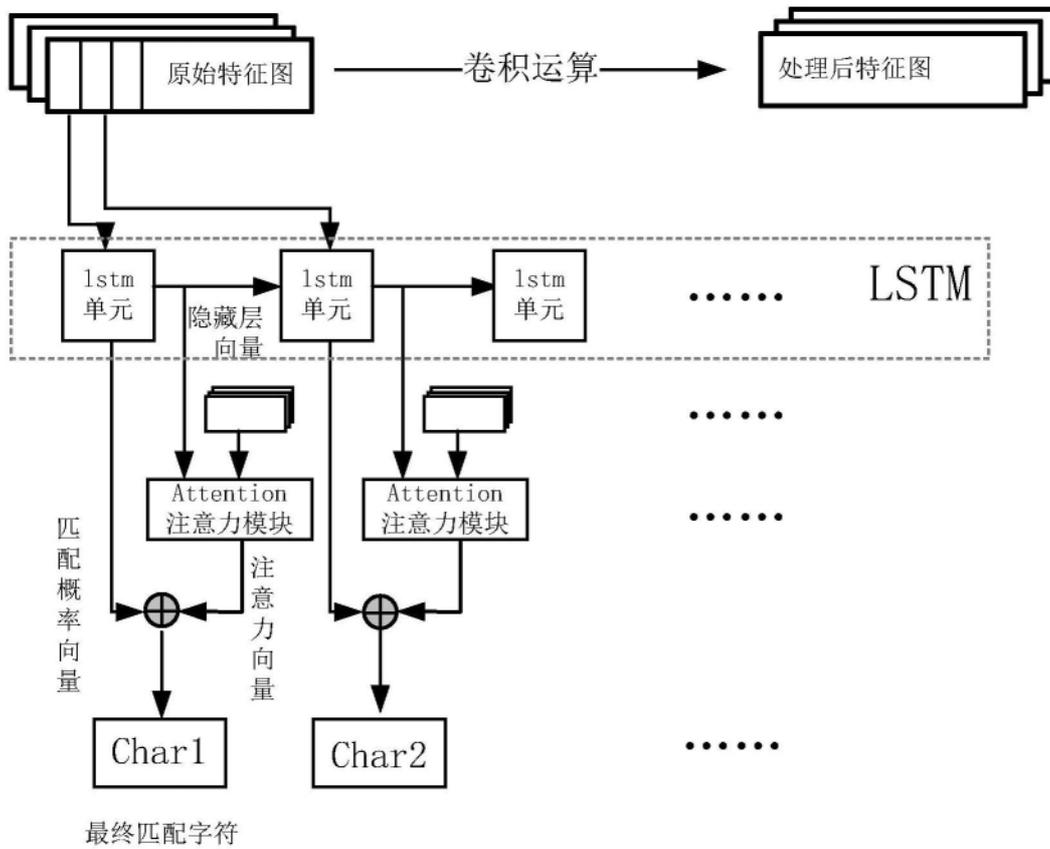
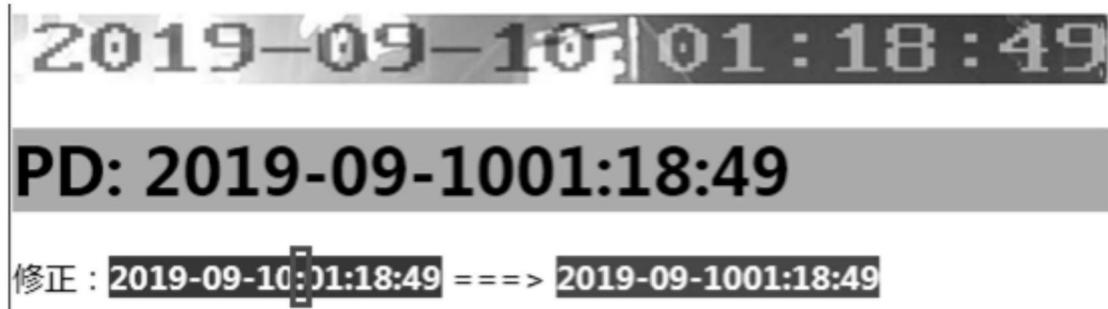


图4



(a)



(b)

图5



图6