



(12) 发明专利申请

(10) 申请公布号 CN 112329767 A

(43) 申请公布日 2021.02.05

(21) 申请号 202011106010.5

(22) 申请日 2020.10.15

(71) 申请人 方正株式(武汉)科技开发有限公司

地址 430000 湖北省武汉市东湖新技术开发
区东信路数码港E幢二层2256、
2266-56室

(72) 发明人 杨威

(74) 专利代理机构 武汉蓝宝石专利代理事务所

(特殊普通合伙) 42242

代理人 廉海涛

(51) Int. Cl.

G06K 9/32 (2006.01)

G06K 9/62 (2006.01)

G06N 3/04 (2006.01)

G06N 3/08 (2006.01)

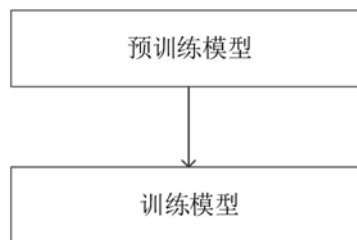
权利要求书2页 说明书7页 附图4页

(54) 发明名称

基于联合预训练的合同文本图像关键信息提取系统和方法

(57) 摘要

本发明涉及一种基于联合预训练的合同文本图像关键信息提取系统和方法,该系统包括:预训练模型,预训练模型通过输入多个合同文本图像进行预训练任务训练后得到,预训练任务包括基于图像位置的文本预测;训练模型,训练模型通过输入多个标注有提取信息的位置的合同文本图像进行训练任务训练后得到,训练任务包括利用预训练模型进行信息提取;将待检测合同文本图像的输入训练完成的训练模型,得到训练模型的预定义的提取信息的所在位置及文字;新增的预训练任务不仅融合了图像特征,还融合了文字预测任务,这不仅使得模型学习到更多的先验知识而且由于预训练阶段无需标注数据所以节省了大量的人力,且信息抽取的正确率更高。



1. 一种基于联合预训练的合同文本图像关键信息提取系统,其特征在于,所述系统包括:预训练模型和训练模型;

所述预训练模型通过输入多个合同文本图像进行预训练任务训练后得到,所述预训练任务包括基于图像位置的文本预测;

所述训练模型通过输入多个标注有提取信息的位置的合同文本图像进行训练任务训练后得到,所述训练任务包括利用所述预训练模型进行信息提取;

将待检测合同文本图像的输入训练完成的所述训练模型,得到所述训练模型的预定义的提取信息的所在位置及文字。

2. 根据权利要求1所述的系统,其特征在于,所述预训练模型包括:图像预处理模块一、文本和图像嵌入层、注意力层和损失函数层一;

所述图像预处理模块一通过OCR工具对输入的合同文本图像进行文字识别,并获取每个文字的位置;

所述文本和图像嵌入层根据每个文字的位置将其嵌入到文本和图像嵌入向量;

所述注意力层包括多层,以所述文本和图像嵌入层作为第一层的输入,每层通过注意力机制运算后输出到下一层;

所述损失函数层通过损失函数对所述预训练模型的参数进行计算更新。

3. 根据权利要求2所述的系统,其特征在于,所述训练模型包括:图像预处理模块二、预训练层和损失函数层二;

所述图像预处理模块二通过OCR工具对输入的标注有提取信息的位置的合同文本图像进行文字识别,并获取每个文字的位置;

所述预训练层包括训练完成的所述预训练模型中的文本和图像嵌入层和注意力层,对所述标注有提取信息的位置的合同文本图像进行预训练;

所述损失函数层二的输入为所述训练模型进行信息提取的预测标签和训练集中的真实标签,根据所述预测标签和真实标签的比较结果对所述训练模型的参数进行更新。

4. 根据权利要求3所述的系统,其特征在于,所述图像预处理模块一和图像预处理模块二获取每个文字的位置的过程包括:

获取每个文字最小图像区块左上角和右下角的水平坐标和垂直坐标,并根据所述水平坐标和垂直坐标的大小顺序将各个文字排列成一行。

5. 根据权利要求3所述的系统,其特征在于,所述文本和图像嵌入向量包括文本嵌入层和2-D坐标嵌入层,所述2-D坐标嵌入层包括:文字最小图像区块左上角水平坐标嵌入层、文字最小图像区块左上角垂直坐标嵌入层、文字最小图像区块右下角水平坐标嵌入层和文字最小图像区块右下角垂直坐标嵌入层。

6. 根据权利要求2所述的系统,其特征在于,所述注意力机制采用多头自注意力机制或普通自注意力机制;

所述注意力机制采用普通自注意力机制时的计算形式为:

$$\text{Attention}(Q,K,V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V, \text{其中} Q、K、V \text{均为张量,} d_k \text{表示张量} K \text{最后一个}$$

维度,T表示转置运算。

7. 根据权利要求2所述的系统,其特征在于,所述损失函数的计算方式为:给定所述文

字最小图像区块的2-D坐标,基于BERT的MLM任务计算损失。

8.一种基于联合预训练的合同文本图像关键信息提取方法,其特征在于,所述方法包括:

步骤1,定义预训练模型及基于图像位置的文本预测的预训练任务,输入多个合同文本图像至所述预训练模型,依据所述预训练任务计算目标函数后更新所述预训练模型的参数;

步骤2,定义训练模型及利用所述预训练模型进行信息提取的训练任务,输入多个标注有提取信息的位置的合同文本图像至所述训练模型,依据所述训练任务计算目标函数后更新所述训练模型的参数;

步骤3,将待检测合同文本图像的输入训练完成的所述训练模型,得到所述训练模型的预定义的提取信息的所在位置及文字。

9.根据权利要求8所述的方法,其特征在于,所述步骤1对所述预训练模型进行训练的过程包括:

通过OCR工具对输入的合同文本图像进行文字识别,并获取每个文字的位置;

根据每个文字的位置将其嵌入到文本和图像嵌入向量;

设计多层结构的注意力层,以所述文本和图像嵌入层作为所述注意力层的第一层的输入,每层通过注意力机制运算后输出到下一层;

通过损失函数对所述预训练模型的参数进行计算更新。

10.根据权利要求9所述的方法,其特征在于,所述步骤2对所述训练模型进行训练的过程包括:

通过OCR工具对输入的标注有提取信息的位置的合同文本图像进行文字识别,并获取每个文字的位置;

根据预训练层对所述标注有提取信息的位置的合同文本图像进行预训练;所述预训练层为训练完成的所述预训练模型中的文本和图像嵌入层和注意力层;

根据所述预测标签和真实标签的比较结果对所述训练模型的参数进行更新。

基于联合预训练的合同文本图像关键信息提取系统和方法

技术领域

[0001] 本发明涉及文本图像信息抽取领域,尤其涉及一种基于联合预训练的合同文本图像关键信息提取系统和方法。

背景技术

[0002] 合同文本图像的关键信息提取指的是,利用某些方法从合同扫描件或合同图像中提取用户感兴趣的、需要提取的关键信息,比如合同签约的“甲方”、“乙方”,合同的签约时间,合同的契约金额等实体。

[0003] 如今,仍有很多公司使用传统的方法,通过人力从商业合同中一页一页的提取,逐个地查找合同中包含“甲方,乙方,契约时间,契约金额”等所有需要提取的实体,这种方法不仅消耗时间而且人工的成本也很大。

[0004] 另一方面,也有很多公司尝试使用自动化提取的方法从合同文本提取关键信息。

[0005] OCR即光学字符识别技术,目前被广泛利用在手写文字识别、基于银行卡身份证等照片的关键信息识别以及合同文本图像的文字识别等诸多领域。同时,随着深度学习算法的快速发展,卷积神经网络(CNN)、循环神经网络(RNN)以及注意力机制(Attention mechanism)被成功的运用到OCR诸多应用中。

[0006] 比如公开号为CN110458162A的中国发明专利申请使用的是基于深度学习(卷积神经网络)的端到端的自动化提取文字信息的一套方法,这种方法通过标注大量的训练数据,直接基于深度学习算法训练一个模型,完成后模型即可用于实际的识别工作中。

[0007] 然而不管是基于CNN、Faster R-CNN、Mask R-CNN还是GCN的深度学习算法,这类方法在自动提取合同关键信息时有两个明显的缺陷:

[0008] 1、他们需要人工标注大量的标签训练数据,也即需要对每份合同标注“甲方,乙方,契约时间,契约金额”等实体的确切位置,昂贵而且导致很长的项目周期。

[0009] 2、模型花费大量时间学习合同文本的结构、布局、位置关系等信息,标注的数据没有发挥出最大的监督学习作用,最后导致一个不太理想的识别效果。

发明内容

[0010] 本发明针对现有技术中存在的技术问题,提供一种基于联合预训练的合同文本图像关键信息提取系统和方法,解决现有技术中问题。

[0011] 本发明解决上述技术问题的技术方案如下:一种基于联合预训练的合同文本图像关键信息提取系统,包括:预训练模型和训练模型;

[0012] 所述预训练模型通过输入多个合同文本图像进行预训练任务训练后得到,所述预训练任务包括基于图像位置的文本预测;

[0013] 所述训练模型通过输入多个标注有提取信息的位置的合同文本图像进行训练任务训练后得到,所述训练任务包括利用所述预训练模型进行信息提取;

[0014] 将待检测合同文本图像的输入训练完成的所述训练模型,得到所述训练模型的预

定义的提取信息的所在位置及文字。

[0015] 一种基于联合预训练的合同文本图像关键信息提取方法,包括:

[0016] 步骤1,定义预训练模型及基于图像位置的文本预测的预训练任务,输入多个合同文本图像至所述预训练模型,依据所述预训练任务计算目标函数后更新所述预训练模型的参数;

[0017] 步骤2,定义训练模型及利用所述预训练模型进行信息提取的训练任务,输入多个标注有提取信息的位置的合同文本图像至所述训练模型,依据所述训练任务计算目标函数后更新所述训练模型的参数;

[0018] 步骤3,将待检测合同文本图像的输入训练完成的所述训练模型,得到所述训练模型的预定义的提取信息的所在位置及文字。

[0019] 本发明的有益效果是:在现有技术基于深度学习算法进行信息提取方法上新增了预训练任务,该预训练任务不仅融合了图像特征,还融合了文字预测任务,这不仅使得模型学习到更多的先验知识而且由于预训练阶段无需标注数据所以节省了大量的人力;基于图像和文本联合预训练方式相较于单一的基于文本或图像预训练方式的效果要好,信息抽取的正确率更高。

[0020] 在上述技术方案的基础上,本发明还可以做如下改进。

[0021] 进一步,所述预训练模型包括:图像预处理模块一、文本和图像嵌入层、注意力层和损失函数层一;

[0022] 所述图像预处理模块一通过OCR工具对输入的合同文本图像进行文字识别,并获取每个文字的位置;

[0023] 所述文本和图像嵌入层根据每个文字的位置将其嵌入到文本和图像嵌入向量;

[0024] 所述注意力层包括多层,以所述文本和图像嵌入层作为第一层的输入,每层通过注意力机制运算后输出到下一层;

[0025] 所述损失函数层通过损失函数对所述预训练模型的参数进行计算更新。

[0026] 进一步,所述训练模型包括:图像预处理模块二、预训练层和损失函数层二;

[0027] 所述图像预处理模块二通过OCR工具对输入的标注有提取信息的位置的合同文本图像进行文字识别,并获取每个文字的位置;

[0028] 所述预训练层包括训练完成的所述预训练模型中的文本和图像嵌入层和注意力层,对所述标注有提取信息的位置的合同文本图像进行预训练;

[0029] 所述损失函数层二的输入为所述训练模型进行信息提取的预测标签和训练集中的真实标签,根据所述预测标签和真实标签的比较结果对所述训练模型的参数进行更新。

[0030] 进一步,所述图像预处理模块一和图像预处理模块二获取每个文字的位置的过程包括:

[0031] 获取每个文字最小图像区块左上角和右下角的水平坐标和垂直坐标,并根据所述水平坐标和垂直坐标的大小顺序将各个文字排列成一行。

[0032] 进一步,所述文本和图像嵌入向量包括文本嵌入层和2-D坐标嵌入层,所述2-D坐标嵌入层包括:文字最小图像区块左上角水平坐标嵌入层、文字最小图像区块左上角垂直坐标嵌入层、文字最小图像区块右下角水平坐标嵌入层和文字最小图像区块右下角垂直坐标嵌入层。

[0033] 进一步,所述注意力机制采用多头自注意力机制或普通自注意力机制;

[0034] 所述注意力机制采用普通自注意力机制时的计算形式为:

[0035] $\text{Attention}(Q,K,V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$,其中Q、K、V均为张量, d_k 表示张量K最后一个维度,T表示转置运算。

[0036] 进一步,所述损失函数的计算方式为:给定所述文字最小图像区块的2-D坐标,基于BERT的MLM任务计算损失。

[0037] 进一步,所述步骤1对所述预训练模型进行训练的过程包括:

[0038] 通过OCR工具对输入的合同文本图像进行文字识别,并获取每个文字的位置;

[0039] 根据每个文字的位置将其嵌入到文本和图像嵌入向量;

[0040] 设计多层结构的注意力层,以所述文本和图像嵌入层作为所述注意力层的第一层的输入,每层通过注意力机制运算后输出到下一层;

[0041] 通过损失函数对所述预训练模型的参数进行计算更新。

[0042] 进一步,所述步骤2对所述训练模型进行训练的过程包括:

[0043] 通过OCR工具对输入的标注有提取信息的位置的合同文本图像进行文字识别,并获取每个文字的位置;

[0044] 根据预训练层对所述标注有提取信息的位置的合同文本图像进行预训练;所述预训练层为训练完成的所述预训练模型中的文本和图像嵌入层和注意力层;

[0045] 根据所述预测标签和真实标签的比较结果对所述训练模型的参数进行更新。

[0046] 采用上述进一步方案的有益效果是:预训练任务让预训练模型和训练模型学习语言特征和单词含义的同时还让预训练模型和训练模型学习到合同这种特殊的排版布局、视觉信息等规律特征,经过预训练的模型在下阶段的关键信息提取任务上表现明显更优。

附图说明

[0047] 图1为本发明提供的一种基于联合预训练的合同文本图像关键信息提取系统的结构框图;

[0048] 图2为本发明提供的一种合同文本图像关键信息提取系统的预训练模型的实施例的结构示意图;

[0049] 图3为本发明提供的一种合同文本图像关键信息提取系统的训练模型的实施例的结构示意图;

[0050] 图4为本发明提供的一种基于联合预训练的合同文本图像关键信息提取方法的实施例的流程图。

具体实施方式

[0051] 以下结合附图对本发明的原理和特征进行描述,所举实例只用于解释本发明,并非用于限定本发明的范围。

[0052] 本发明提供一种基于联合预训练的合同文本图像关键信息提取系统,如图1所示为本发明提供的一种基于联合预训练的合同文本图像关键信息提取系统的结构框图,由图1可知,该系统包括:预训练模型和训练模型。

[0053] 预训练模型通过输入多个合同文本图像进行预训练任务训练后得到,预训练任务包括基于图像位置的文本预测。

[0054] 训练模型通过输入多个标注有提取信息的位置的合同文本图像进行训练任务训练后得到,训练任务包括利用预训练模型进行信息提取。

[0055] 将待检测合同文本图像的输入训练完成的训练模型,得到训练模型的预定义的提取信息的所在位置及文字。

[0056] 具体的,该预定义的提取信息可以为根据需求设置的文字或字符。

[0057] 不同于基于深度学习算法有且只有一个训练阶段的OCR技术来识别合同文本图像的关键信息,本发明不仅使用了深度学习算法,而且最重要的是使用了基于图像和文本的联合预训练的技术手段。

[0058] 当前很多的解决方案中会使用到一种预训练模型,比如基于自然语言预训练的BERT模型,还有比如基于图像预训练的Mask R-CNN模型。

[0059] 以自然语言预训练模型为例,预训练阶段无需大量人工标注的数据,利用纯文本即可完成训练,这一阶段模型能够学习到语言的规律、理解到单词的语义,经过这个学习,模型在接下来的抽取任务会有较好效果。

[0060] 同样地,在图像预训练模型领域,通过对图像模型进行预训练,模型掌握了常见图形的规律和视觉特征,进而在接下来的抽取任务上的表现会有较好效果。

[0061] 本发明使用一种图像和语言文本联合的预训练技术,旨在让模型学习语言特征和单词含义的同时还让模型学习到合同这种特殊的排版布局,视觉信息等规律特征,经过预训练的模型在下阶段的关键信息提取任务上表现明显更优。

[0062] 以合同信息提取为例,给定合同中一个关键字“契约金额”,那么相对应的契约金额值很有可能在“契约金额”四个字的右边或者在其下方而不太可能在其左方和上方,这种重要的规律是可以通过模型在预训练阶段学习到的。

[0063] 另外,在如合同这类特殊的文本中,文字的视觉特征也暗含着信息,比如文字是否为粗体、是否有下划线以及是否为斜体,这对于信息抽取任务也是较为重要。

[0064] 本发明不仅在现有技术基于深度学习算法进行信息提取方法上新增了预训练任务,该预训练任务不仅融合了图像特征,还融合了文字预测任务,这不仅使得模型学习到更多的先验知识而且由于预训练阶段无需标注数据所以节省了大量的人力;基于图像和文本联合预训练方式相较于单一的基于文本或图像预训练方式的效果要好,信息抽取的正确率更高。

[0065] 实施例1

[0066] 本发明提供的实施例2位本发明提供的一种基于联合预训练的合同文本图像关键信息提取系统的实施例,该系统的实施例包括:预训练模型和训练模型。

[0067] 预训练模型通过输入多个合同文本图像进行预训练任务训练后得到,预训练任务包括基于图像位置的文本预测。

[0068] 优选的,如图2所示为本发明提供的一种合同文本图像关键信息提取系统的预训练模型的实施例的结构示意图,由图2可知,该预训练模型包括:图像预处理模块一、文本和图像嵌入层、注意力层和损失函数层一。

[0069] 图像预处理模块一通过OCR工具对输入的合同文本图像进行文字识别,并获取每

个文字的位置。

[0070] 文本和图像嵌入层根据每个文字的位置将其嵌入到文本和图像嵌入向量,具体的,该文本和图像嵌入向量是一个高维的数学向量空间。

[0071] 注意力层包括多层,以文本和图像嵌入层作为第一层的输入,每层通过注意力机制运算后输出到下一层。

[0072] 优选的,注意力机制可以采用多种形式,比如常用的多头自注意力机制或普通自注意力机制。

[0073] 注意力机制采用普通自注意力机制时的计算形式为:

[0074] $\text{Attention}(Q,K,V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$,其中Q、K、V均为张量, d_k 表示张量K最后一个维度,T表示转置运算。

一个维度,T表示转置运算。

[0075] 具体的,注意力层的个数N可以选择6或12等,一旦确定,整个过程不会变动。

[0076] 损失函数层LOSS通过损失函数对预训练模型的参数进行计算更新。

[0077] 计算损失函数LOSS的目的在于期望对预训练模型中的参数(如文本和图像嵌入向量)进行更新。

[0078] 优选的,损失函数LOSS的计算方式多种多样,具体可以为:给定文字最小图像区块的2-D坐标,基于BERT(Bidirectional Encoder Representation from Transformers,利用Transformer结构的双向编码表示方法)的MLM(masked language model,掩码语言模型)任务计算损失,BERT的预训练目标包括两个,MLM通过Mask(遮住)单词序列中一定比例的单词,并指导模型能够成功预测对Mask位置的单词。

[0079] 比如,给定某个文字序列,MLM任务将第3个字符MASK,模型需要在该文字图像的特征向量和文字特征向量基础上预测该文字,最终输出预测分布 $P(w_3|w_1, w_2, w_4 \cdots w_n, \text{image_vecs})$;其中 w_1, w_2, \cdots, w_n 代表文字,image_vecs代表n个文字的每个文字最小图像区块的图像编码向量,图像编码模型采取Faster R-CNN。

[0080] 训练模型通过输入多个标注有提取信息的位置的合同文本图像进行训练任务训练后得到,训练任务包括利用预训练模型进行信息提取。

[0081] 优选的,如图3所示为本发明提供的一种合同文本图像关键信息提取系统的训练模型的实施例的结构示意图,由图3可知,该训练模型包括:图像预处理模块二、预训练层和损失函数层二。

[0082] 图像预处理模块二通过OCR工具对输入的标注有提取信息的位置的合同文本图像进行文字识别,并获取每个文字的位置。

[0083] 预训练层包括训练完成的预训练模型中的文本和图像嵌入层和注意力层,对标注有提取信息的位置的合同文本图像进行预训练。

[0084] 损失函数层二的输入为训练模型进行信息提取的预测标签和训练集中的真实标签,根据预测标签和真实标签的比较结果对训练模型的参数进行更新。

[0085] 训练模型将信息提取任务当作序列标注任务,预训练层使用的模型和参数使用的训练完成的预训练模型的网络模型和参数。具体的,该训练集可以通过对所有文字或字符进行BIO标注得到,BIO标注将每个元素标注为“B-X”、“I-X”或者“O”。其中,“B-X”表示此元素所在的片段属于X类型并且此元素在此片段的开头,“I-X”表示此元素所在的片段属于X

类型并且此元素在此片段的中间位置，“0”表示不属于任何类型。

[0086] 具体的，图像预处理模块一和图像预处理模块二获取每个文字的位置的过程包括：

[0087] 获取每个文字最小图像区块左上角和右下角的水平坐标和垂直坐标(参照于固定点的相对坐标)，并根据水平坐标和垂直坐标的大小顺序将各个文字排列成一行。

[0088] 进一步的，由图2可知，文本和图像嵌入向量包括文本嵌入层和2-D坐标嵌入层，2-D坐标嵌入层分为四个，包括：文字最小图像区块左上角水平坐标嵌入层、文字最小图像区块左上角垂直坐标嵌入层、文字最小图像区块右下角水平坐标嵌入层和文字最小图像区块右下角垂直坐标嵌入层。

[0089] 以左上角水平坐标嵌入层为例，当不同文字的左上角水平坐标相同，那么这些文字的左上角水平坐标嵌入向量相同。

[0090] 另外，初始的文本和图像嵌入向量可以随机化(如从 $X \sim N(0, 1)$ 抽样)，并确定统一的向量维度(如768维)；随着预训练过程不断进行，嵌入层的向量持续更新。

[0091] 将待检测合同文本图像的输入训练完成的训练模型，得到训练模型的预定义的提取信息的所在位置及文字。

[0092] 实施例2

[0093] 本发明提供的实施例2位本发明提供的一种基于联合预训练的合同文本图像关键信息提取方法的实施例，如图4所示为本发明提供的一种基于联合预训练的合同文本图像关键信息提取方法的实施例的流程图，由图4可知，该方法的实施例包括：

[0094] 步骤1，定义预训练模型及基于图像位置的文本预测的预训练任务，输入多个合同文本图像至预训练模型，依据预训练任务计算目标函数后更新预训练模型的参数。

[0095] 具体的，准备N张合同文本图像，其中文字信息等无需标注；预训练模型的网络可以使用例如BERT主干网络，随后逐个遍历合同文本图像进行OCR文字识别，最后基于设定的预训练任务进行预训练。

[0096] 优选的，对预训练模型进行训练的过程包括：

[0097] 通过OCR工具对输入的合同文本图像进行文字识别，并获取每个文字的位置。

[0098] 根据每个文字的位置将其嵌入到文本和图像嵌入向量。

[0099] 设计多层结构的注意力层，以文本和图像嵌入层作为注意力层的第一层的输入，每层通过注意力机制运算后输出到下一层。

[0100] 通过损失函数对预训练模型的参数进行计算更新。

[0101] 具体的，预训练过程中，实施者需要将准备的N张合同文本图像依次处理，首先进行OCR文字识别得到文字和文字最小图像区块2-D坐标，进入文本和图像嵌入层后得到文本嵌入向量和图像2-D坐标嵌入向量，随后将向量进行注意力计算，最后依据预训练任务计算目标函数，然后更新整个预训练网络模型的参数。

[0102] 所有N张文本图像遍历一次计为1个Epoch，根据训练效果自定义训练Epoch大小。

[0103] 步骤2，定义训练模型及利用预训练模型进行信息提取的训练任务，输入多个标注有提取信息的位置的合同文本图像至训练模型，依据训练任务计算目标函数后更新训练模型的参数。

[0104] 具体的，标注M张标注有提取信息的位置的合同文本图像，在步骤1中训练完成的

预训练网络基础上新增适应下游信息抽取任务的网络层,随后逐个遍历M张合同文本图像并OCR文字识别正式进行训练。

[0105] 优选的,对训练模型进行训练的过程包括:

[0106] 通过OCR工具对输入的标注有提取信息的位置的合同文本图像进行文字识别,并获取每个文字的位置。

[0107] 根据预训练层对标注有提取信息的位置的合同文本图像进行预训练;预训练层为训练完成的预训练模型中的文本和图像嵌入层和注意力层。

[0108] 根据预测标签和真实标签的比较结果对训练模型的参数进行更新。

[0109] 具体的,训练过程中,实施者需要基于预训练模型(文本和图像嵌入层+注意力层)添加基于下游抽取任务的FC layer(如输出维度为N_CLASS的全连接层网络),随后逐个遍历M张合同文本图像,并依据训练任务计算目标函数,然后更新神经网络的模型参数。

[0110] 所有M张文本图像遍历一次计为1个Epoch,根据训练效果自定义训练Epoch大小。

[0111] 步骤3,将待检测合同文本图像的输入训练完成的训练模型,得到训练模型的预定义的提取信息的所在位置及文字。

[0112] 预训练过程和训练过程均完成后可以正式的进入使用阶段:

[0113] 输入文本图像,OCR算法识别文字,模型会自动提取文中预定义好的信息的所在位置及文字。

[0114] 以上仅为本发明的较佳实施例,并不用以限制本发明,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。

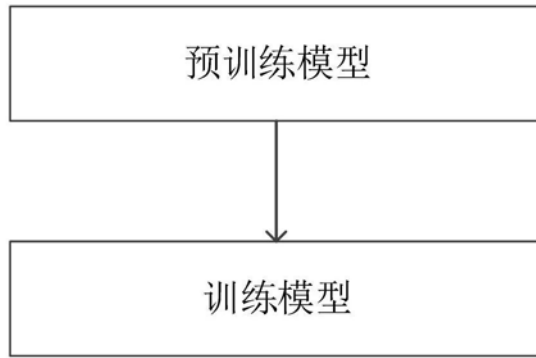


图1

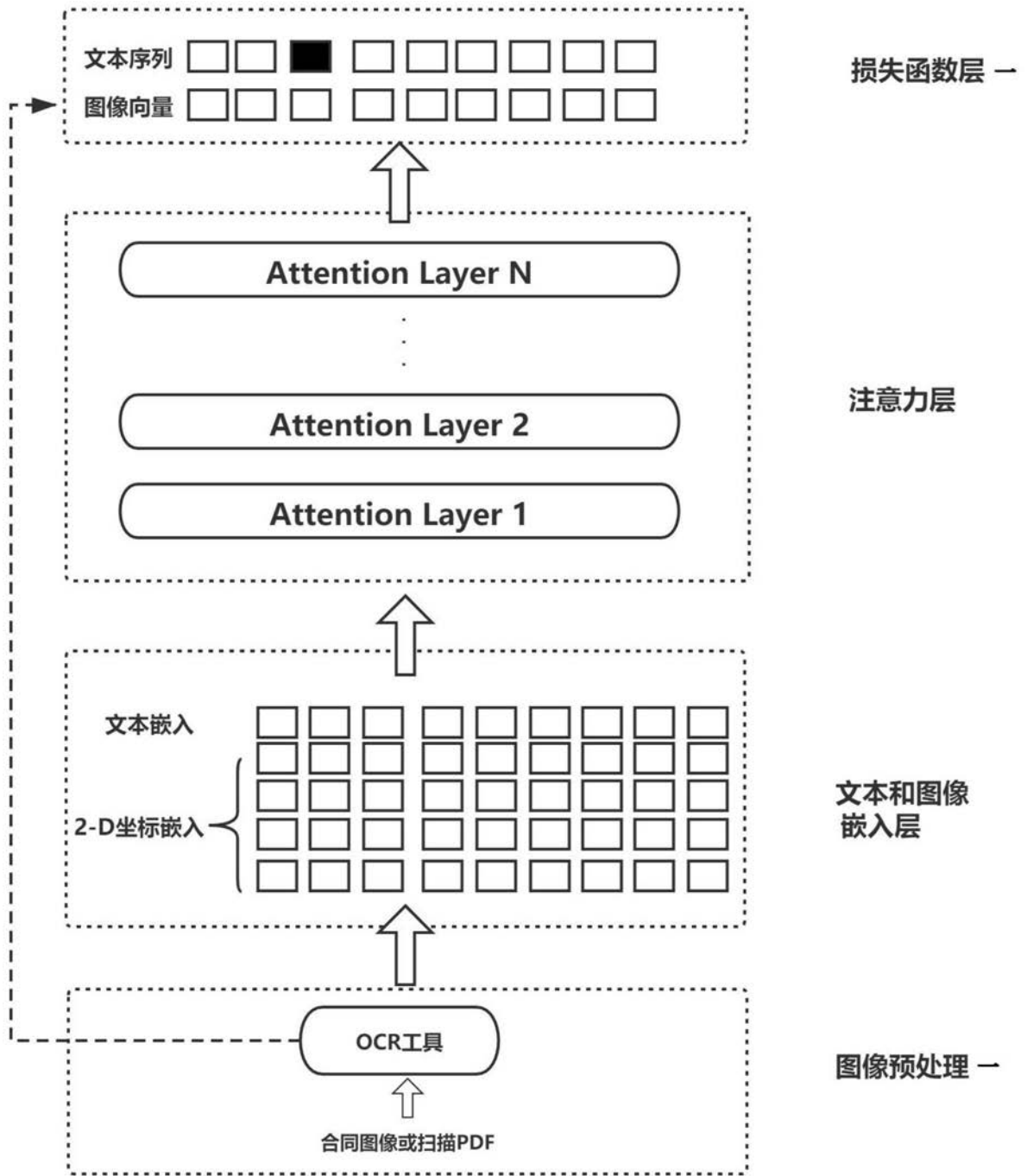


图2

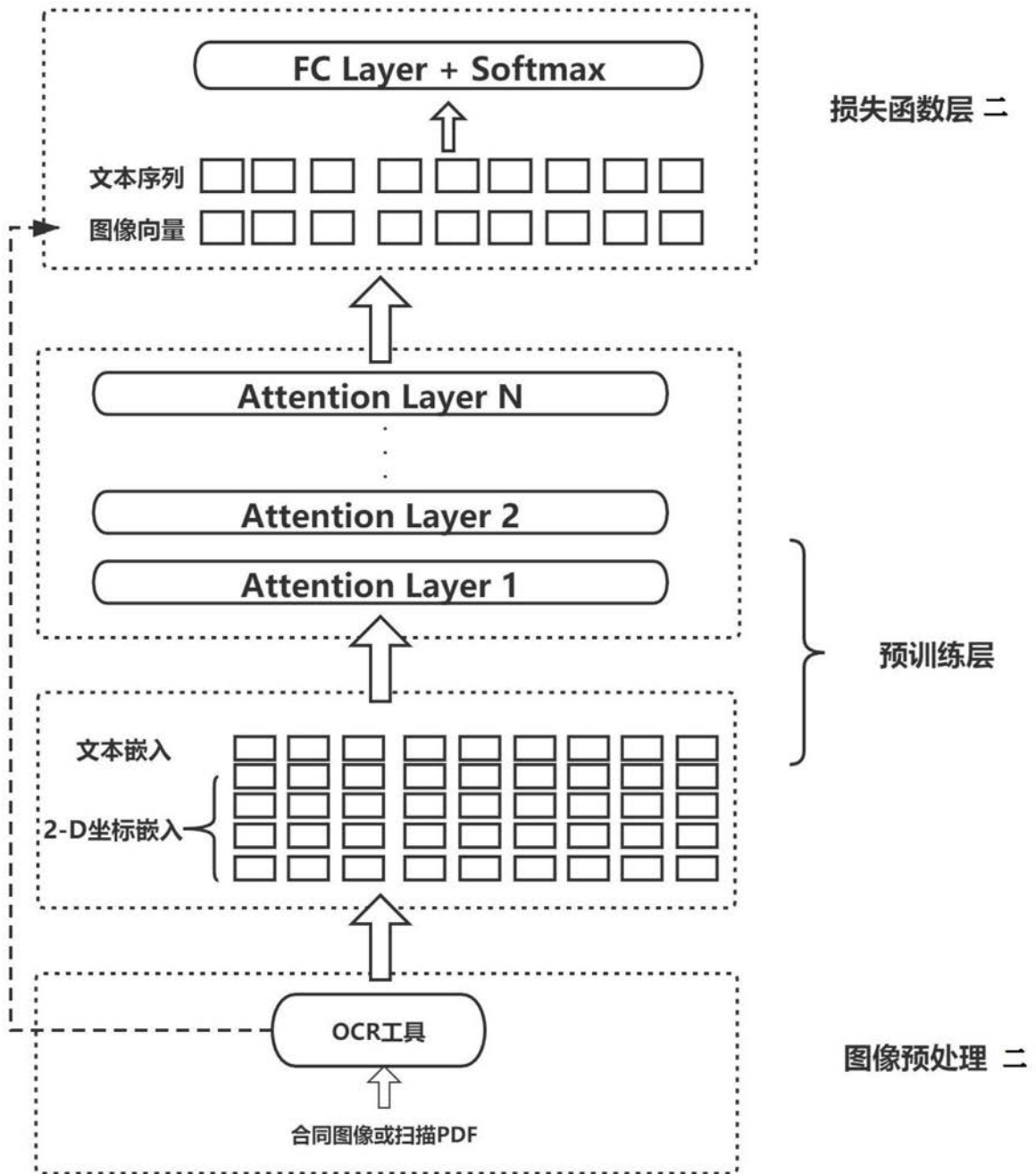


图3

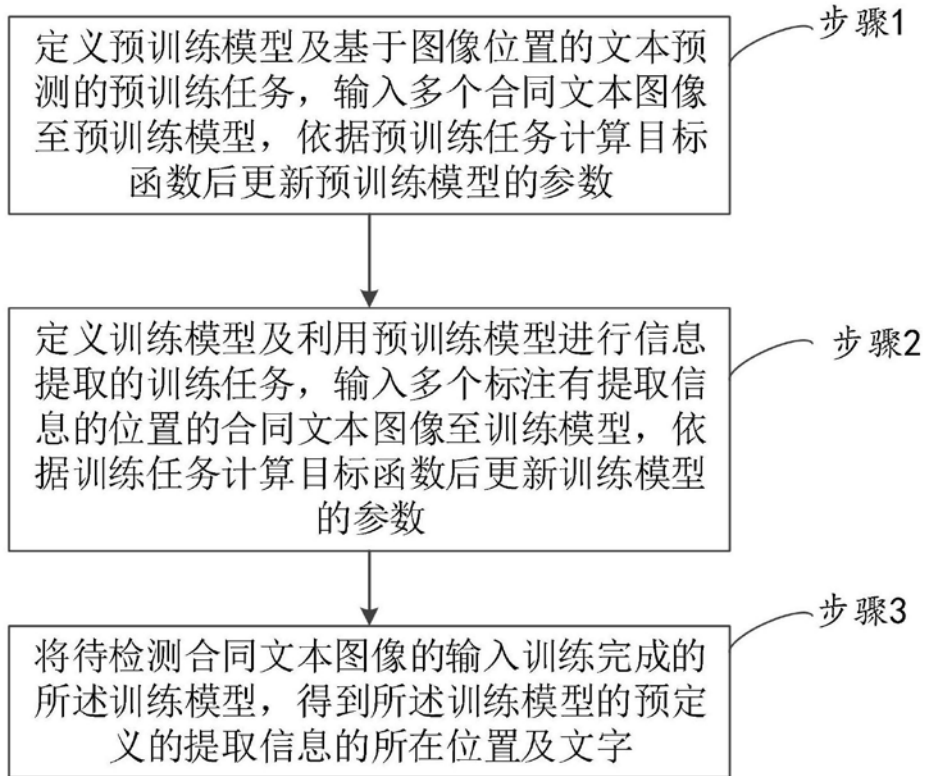


图4