Hans Journal of Data Mining 数据挖掘, 2023, 13(2), 143-153
Published Online April 2023 in Hans. https://www.hanspub.org/journal/hjdm
https://doi.org/10.12677/hjdm.2023.132014

Hans汉斯

融合空间特征的债券图表数据文本检测方法研究

李桂钢，胡金蓉，帅梓涵，郎子鑫，罗月梅

成都信息工程大学计算机学院，四川 成都

收稿日期：2023年3月13日；录用日期：2023年4月13日；发布日期：2023年4月21日

摘　要

随着国家明确了金融业发展和改革的重点方向，我国金融数据信息化有了显著的发展和进步。基于债券图表数据的特定情况，人工处理债券图表数据存在效率低、成本高、安全性低等问题，用人工智能的方法来检测债券图表数据逐渐成为了当下的热门研究方向。由于债券图表数据在长时间存放、人为损坏等主客观因素下，会存在模糊、被污染等特点。对此本文使用了Swin-Transformer作为主干网络，它的特征提取能力较CNN (卷积神经网络)更为强大。并对模糊、污染的区域设计了方向感知模块，使其对文本区域的识别正确率更高。实验结果表明，该网络比其它文本检测算法在准确率、召回率、F1值上都有明显提升。

关键词

债券图表数据，文本检测，Swin-Transformer，方向感知模块

Text Detection Method for Bond Chart Data Fusing Spatial Features

Guigang Li, Jinrong Hu, Zihan Shuai, Zixin Lang, Yuemei Luo

School of Computer Science, Chengdu University of Information Technology, Chengdu Sichuan

Received: Mar. 13th, 2023; accepted: Apr. 13th, 2023; published: Apr. 21st, 2023

Abstract

With the clear direction of financial development and reform, China's financial data information technology has made significant progress and development. Based on the specific situation of bond chart data, manual processing of bond chart data is inefficient, high cost, low security, and so

文章引用: 李桂钢, 胡金蓉, 帅梓涵, 郎子鑫, 罗月梅. 融合空间特征的债券图表数据文本检测方法研究[J]. 数据挖掘, 2023, 13(2): 143-153. DOI: 10.12677/hjdm.2023.132014

on. Using artificial intelligence to detect bond chart data has gradually become a hot research direction. Because the bond chart data is stored for a long time and damaged artificially, it will be fuzzy and polluted. In this paper, Swin-Transformer is used as the backbone network, and its feature extraction ability is stronger than that of CNN (convolution network). Direction perception module is designed for blurred and contaminated areas, which makes the recognition of text areas more accurate. The experimental results show that the network improves the accuracy, recall and F1 value significantly compared with other text detection algorithms.

Keywords

Bond Chart Data, Text Detection, Swin-Transformer, Direction-Aware Module

1. 引言

文本作为人类智慧的结晶，是人类文化、思想传承的一种基本信息载体[1]。在电子设备日益发展的今天，智能手机、数码相机等被人们广泛使用，伴随的是不计其数的文本以图像或者视频的数据形式被保存下来。同时，移动互联网技术的快速发展使得人们希望利用计算机对图像或者视频中的文本进行检测，来提高各种应用场景的生产效率，比如债券数据、文档数据等业务需求。

针对文本检测这个任务，早期的传统文本检测方法主要分为两大类，一是基于滑动窗口的检测；二是基于连通区域分析的检测。近年来，由于深度学习具有以下特征而被越来越多的研究人员使用，首先是自动化程度高，能够自动地捕获文本图像的高级视觉特征；其次是检测性能高效，基于深度学习的文本检测算法相比于基于手工设计特征的传统算法而言，识别性能更好；最后是泛化性能优越，可以很容易地应用到相似的视觉任务。但图像中的文字区域存在模糊、污染这种特征量较少的区域，它们可能就会产生错误的判断，因为它们无法分析和理解全局图像语义。

检测模糊和污染区域需要使用全局图像语义才能得到更好的效果，为此本文提出了方向感知模块来分析图像上下文信息，通过周围的信息来识别文本区域可以得到更加准确的结果。以图 1 中的区域 A 为例，将其与区域 B 和 C 进行比较，区域 B 会比区域 C 更有力地表明 A 是文本区域。因此在检测模糊、污染的文本区域时，不同方向上的上下文信息将提供不同的帮助。

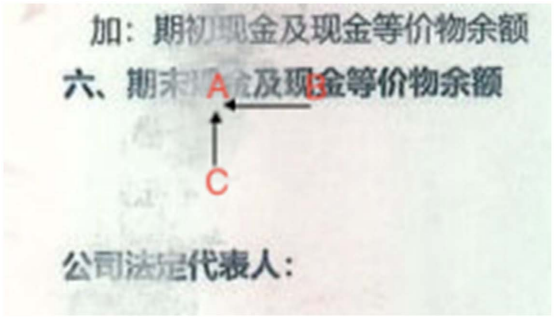


Figure 1. The picture is blurred in some areas
图 1. 图片局部区域模糊

本文使用了特征提取能力更强的 Swin-Transformer [2]作为主干网络，它的特征是通过移动窗口的方式学来的，由于自注意力是在窗口内计算的，所以降低了序列的长度并带来了更大的效率，同时通过移动的操作可以使相邻的两个窗口之间进行交互，从而有了全局建模的能力。为了获取图像不同方向上的信息，本文设计了方向感知模块，在该模块中，首先通过在四个主要方向上采用长短期记忆网络(LSTM)来聚合全局空间上下文信息，LSTM 是一种特殊的空间递归神经网络(RNN) [3]，然后在 LSTM 中制定一个方向感知注意力机制[4]来学习每个方向的注意力权重。因此，本文使用了方向感知模块来获取空间上下文信息，并将多个方向感知模块嵌入到主干网络中，以了解不同层(尺度)中的特征。本文工作的主要贡献如下：

1) 本文使用了特征提取能力更强的 Swin-Transformer 作为主干网络。

2) 本文在 LSTM [5]后加上了一种新颖的注意力机制，并构造了方向感知模块来学习空间上下文信息。

3) 本文在自建数据集上评估我们的方法，并将其与最新方法进行比较，实验结果表明，我们的网络在准确率、召回率、F1 值上都有不错的表现。

2. 相关工作

2.1. 传统的文本检测方法

传统的文本检测方法主要分成两大类，一是基于滑动窗口的检测；二是基于连通区域分析的检测。

基于滑动窗口的文本检测方法，首先对输入图像构建图像金字塔。然后在金字塔各层进行滑动扫描，对每个滑动窗口采样的位置都提取特征并进行文本和非文本区域分类。在文本分类中人工设计的特征通常包括空间差分[6]、像素强度[7]和方向梯度直方图[8]等。随着卷积神经网络[9]的兴起，也有学者提出用 CNN 提取的特征取代上述特征表达以增强滑窗分类的性能[10] [11]。

基于连通区域分析的文本检测方法，首先对输入图像提取候选文本连通区域；然后使用人工设计的规则或特征对这些连通区域进行文本和非文本分类；典型方法有基于笔画宽度变换的方法[12]和基于最大稳定极值区域[13]的方法。

2.2. 基于深度学习的文本检测方法

目前，根据检测文本对象的不同可以将基于深度学习的方法划分为基于回归的方法和基于分割的方法。

基于回归的文本检测方法通过回归对应的边界盒来对目标文本进行定位。Tian 等人[14]提出了一种水平文本的检测方法(CTPN)。该方法将基于锚点(anchoring)的通用目标检测算法与循环神经网络(recurrent neural network)相结合，从而使得检测模型具有挖掘文本序列上下文特征的能力。Zhou 等人[15]提出了一种快速文本检测方法(EAST)。在 EAST 算法中，来自主干网络中不同层级的特征图被上采样至原图大小的 1/4 后融合，以获取不同尺度感受野(receptive field)下的特征图。Ma 等人[16]通过对 FasterR-CNN 中的候选区域生成网络进行改进，提出了一种基于旋转候选框生成网络(RRPN)的任意朝向文本检测网络。

基于分割的文本检测方法通常对像素进行预测，再通过后处理得到文本框输出。Zhang 等人[17]首先采用全连接网络(FCN)预测分割图来区分文本或非文本区域。对于每一个文本区域，采用最大稳定极值区域算法(MSER)来提取字符候选框。Deng 等人[18]提出了一种基于实例分割思想的文本检测器(PixelLink)。该方法首先学习预测输入图片中的每一个像素是否属于文本区域，再通过加入连接预测(link prediction)的方式，学习该像素点与其相邻的 8 个像素点是否属于同一个文本实例，以此来将正样本组成不同的文本实例输出。Liao 等人[19]提出了可微分二值化(differentiable binarization, DB)的方法。DB 将二值化操作

嵌入网络，除了学习文本区域的概率图之外，还会预测对应的阀值图，通过两者结合生成最后的结果。

3. 本文方法

在本节中，将要介绍的是文本检测网络，图 2 展示了文本检测网络的整体结构，该网络使用 Swin-Transformer 提取不同尺度的特征[20]，再使用多个方向感知模块来学习不同尺度的空间上下文特征。该网络将整个图像作为输入，并以端到端的方式输出文本区域。

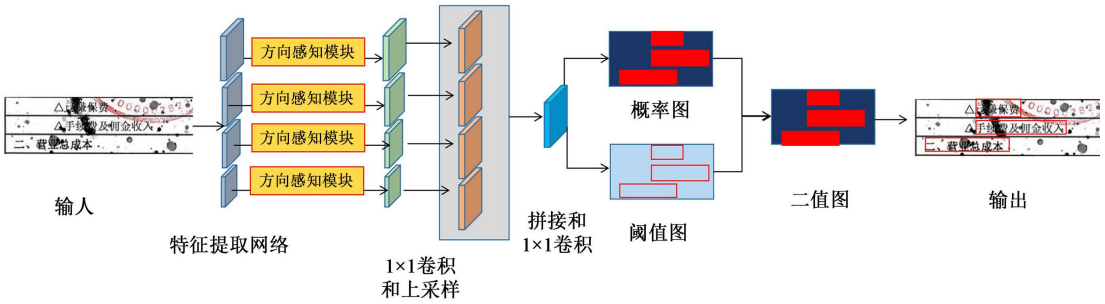


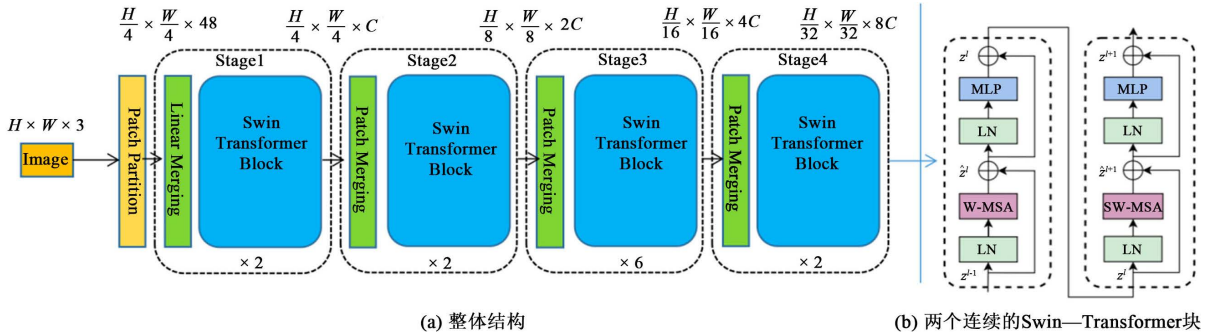
Figure 2. Diagram of the entire text detection network
图 2. 整个文本检测网络的示意图

首先，该网络使用 Swin-Transformer 提取不同尺度的分层特征图。较浅层的特征图有助于保留文本边界，而较深的特征图则具有更多的全局语义，有助于识别文本和非文本区域。其次，对于每一层，都使用了方向感知模块获取空间上下文特征。第三，我们将空间上下文特征与相应主干网络提取的特征图连接起来，并将连接后的特征图上采样到输入的大小。第四，为了利用不同层上特征图的互补优势，我们将上采样的特征图连接起来，并使用 1 × 1 卷积层来生成多尺度特征融合图。此外，根据多尺度特征融合图计算出概率图和阈值图。最后，可以由概率图和阈值图得到二值图。

在以下小节中，首先会介绍我们的特征提取网络。之后，将介绍生成空间上下文特征的方向感知模块，最后介绍该网络的损失函数。

3.1. 特征提取网络

本网络选取了 Swin-Transformer 作为特征提取网络，图 3 展示了 Swin-Transformer 的整体结构。在本小节中，会描述 Swin-Transformer 的特点，它如何使用基于窗口的自注意力计算减少计算量(第 2.1.1 节)，然后再基于移动窗口的自注意力计算进行全局建模，进而增强它的特征提取能力(第 2.1.2 节)。



(a) 整体结构　　(b) 两个连续的Swin—Transformer块

Figure 3. Structure drawing of Swin-Transformer
图 3. Swin-Transformer 的结构图

Swin-Transformer 采用的是分层的思想，分为 Stage 1、Stage 2、Stage 3 和 Stage 4。当输入的图片尺寸为 $H \times W$ 时，设置每个 patch 的尺寸为 4 × 4，每个 patch 有 16 个像素，得到 patches 数量为 $\frac{(H \times W)}{16}$，因为 RGB 图像有 3 个通道，所以得到一个 patch 的特征维度为 16 × 3 = 48，再经过一层线性层投影到 C 维度，$\frac{H}{4} \times \frac{W}{4} \times C$ 就是 Stage 1 的输出。PatchMerging 层将 4 个相邻的 patch 特征进行拼接得到 $4C$ 维度的特征，然后接一个线性层得到 $2C$ 维度特征，再通过 Swin-TransformerBlock 进行特征变换，最终得到 Stage 2 的输出为 $\frac{H}{8} \times \frac{W}{8} \times 2C$。以此类推，Stage 3 的输出为 $\frac{H}{16} \times \frac{W}{16} \times 4C$，Stage 4 的输出为 $\frac{H}{32} \times \frac{W}{32} \times 8C$。

3.1.1. 基于窗口的自注意力计算

在进行自注意力计算时，以前会基于全局进行计算，当进行视觉里的下游任务尤其是图片尺寸很大时，基于全局计算自注意力的复杂度会非常的高，达到平方倍，所以 Swin-Transformer 采用了基于窗口计算自注意力来减少计算量。

如图 4 所示，将原来的图片分成不重叠的多个窗口，窗口内的 patch 图像块是最小的计算单元，每个窗口都有 m × m 个 patch，一般 m 的值默认为 7，此时每个窗口里就有 49 个 patch，自注意力计算都是分别在窗口内完成的，所以减少了计算量。

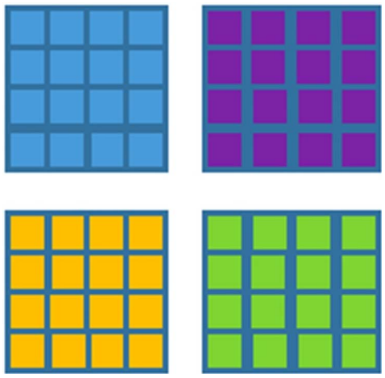


Figure 4. A window-based self-attention calculation method in Swin-Transformer
图 4. Swin-Transformer 中基于窗口的自注意力计算方法

3.1.2. 基于移动窗口的自注意力计算

上面的方法有效的解决了计算量的问题，但是窗口与窗口之间缺少了联系，没法做到全局建模，这就减弱了模型的拟合能力。Swin-TransformerBlock 中先进行一次窗口的自注意力计算，再进行移动窗口自注意力计算，这样就实现了窗口之间的通信，从而达到了全局建模的效果。

图 5 是将窗口进行移位之后的示意图，可以看到移位后的窗口包含了图 4 相邻窗口的元素，这样就达到了窗口与窗口之间的相互通信。但这也引入了一个新问题，即窗口的个数变成了 9 个，大小也不一样，这就导致计算难度增加了。

Swin-Transformer 使用了循环移位和掩码操作的方式，即保证了移动窗口后窗口的数量保持不变，也保证了每个窗口内的 patch 数量不变。

通过循环移位，图片重新分成了 4 个窗口，但循环移位后仍然存在一些问题。A、B、C 三个区域是从很远的地方移位来到这三个窗口的，与这三个窗口原本的元素之间是没有什么联系的，所以我们不需要对二者进行注意力计算，针对这个问题，Swin-Transformer 提出了掩码操作。增加一层 mask，对于我

们需要计算的区域，mask 的值设为 0，这样对结果就不会有任何影响，对于不需要计算的区域，mask 的值设为−100，这样在后面的计算中就可以忽略掉该值(图 6)。

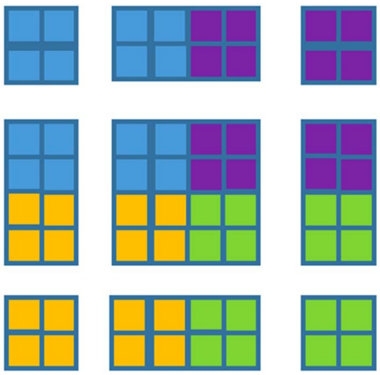


Figure 5. A self-attention calculation method based on moving Windows in Swin-Transformer
图 5. Swin-Transformer 中基于移动窗口的自注意力计算方法

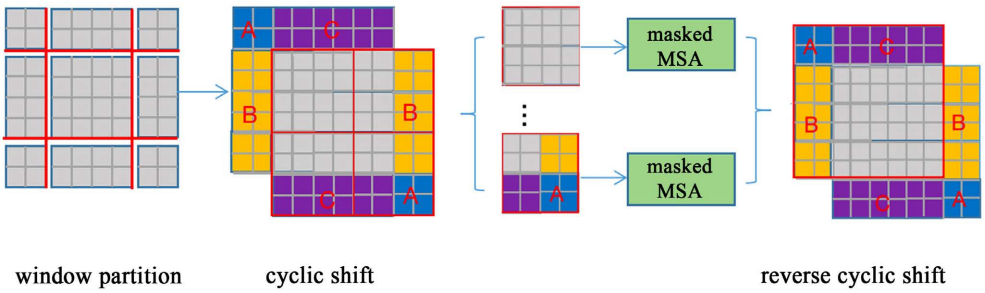


Figure 6. Cyclic shift and mask operation
图 6. 循环移位与掩码操作

本节首先介绍了 Swin-Transformer 的整体结构，然后介绍了它是如何进行全局建模和减少计算量的。

3.2. 方向感知模块

针对图像中有污染、模糊的文本区域，该区域特征量较少，直接判断可能会产生错误。在特征提取网络之后加上方向感知模块，得到空间上下文特征，会有利于我们对文本区域的判断。图 7 展示了方向感知模块，在本小节中，我们首先描述如何使用 LSTM 获取空间上下文特征(第 2.2.1 节)，然后说明如何在 LSTM 中制定方向感知注意力机制，以学习注意力权重并生成空间上下文特征(第 2.2.2 节)。

该模块使用 LSTM 通过两轮在四个主要方向上计算空间上下文特征。第一轮将特征图输入 LSTM 之后，特征图中的任何一个像素都会得到来自它所在的行和列的像素信息，第二轮重复上述操作，特征图中的每一个信息就会得到来自所有像素的信息，这样的特征图是包含丰富的全局上下文信息。并制定注意力机制以生成注意力权重图，以组合不同方向的空间上下文特征。在两轮递归转换中，网络共享权重。

3.2.1. 空间上下文特征

LSTM 相比于原始的 RNN 增加了一个细胞状态，下图 8 是向量 Xt 输入 LSTM 的结构图。LSTM 在处理有序数据时有三个输入：细胞状态 Ct-1，隐藏状态 ht-1，第 t 个位置的向量 Xt，输入经过遗忘门、

更新门层、输出门层得到输出，而输出有两个：细胞状态 Ct，隐层状态 ht。其中 ht 还作为 t 时刻的最终计算结果。

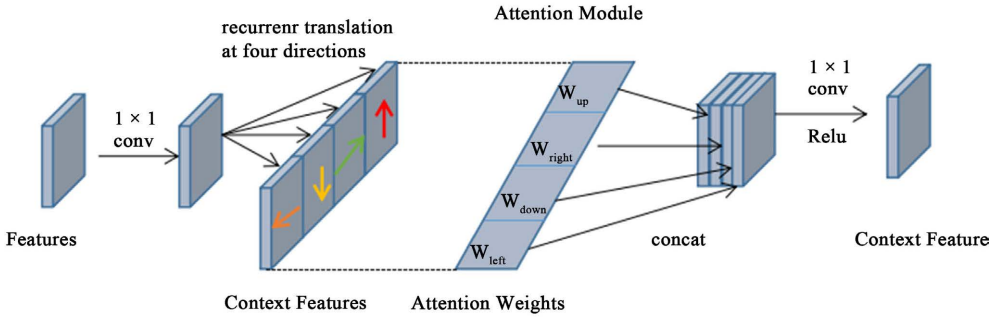


Figure 7. Direction sensing module
图 7. 方向感知模块

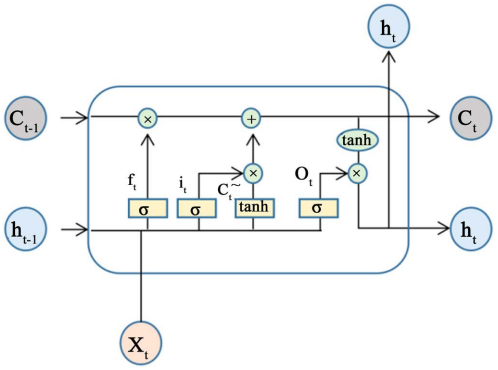


Figure 8. LSTM structure diagram
图 8. LSTM 结构图

遗忘门计算得到的输出 f_t 在 0 到 1 之间，会和上一时刻 C_{t-1} 的元素相乘，当 f_t 某部分为 0 时，C_{t-1} 该部分的信息就会被舍弃，f_t 的值在 0 到 1 时，会保留 C_{t-1} 的部分信息，只有当 f_t 为 1 时，才会完整保留 C_{t-1} 的信息。

$$f_t = \sigma\left(W_f\left[h_{t-1}, X_t\right] + b_f\right) \tag{1}$$

其中，W_f 是权重矩阵，b_f 为偏置量，σ 表示 sigmoid 函数，它的输出是在 0 到 1 之间的。

更新门层有两个部分，一个是 $C_{\tilde{t}}$，这个可以看作是新的输入 X_t 带来的信息，另一部分是 it，这部分的结构和遗忘门是一样的，它的作用可以看作是新的信息保留哪些部分。

$$i_t = \sigma\left(W_i\left[h_{t-1}, X_t\right] + b_i\right) \tag{2}$$

$$C_{\tilde{t}} = \tanh\left(W_c\left[h_{t-1}, X_t\right] + b_c\right) \tag{3}$$

其中，tanh 是双曲正切函数，它的输出在−1 到 1 之间。

有了以上两步操作，就可以对其中的一个输出 C_t 进行更新，公式左边表示对过去的信息有选择的遗忘或保留，公式右边表示表示对新的信息有选择的遗忘或保留，最后把这两部分加起来，就是新的细胞状态 C_t 了。

$$C_t = f_t C_{t-1} + i_t C_{\tilde{t}} \tag{4}$$

输出门层控制另一个输出，就是 LSTM 在 t 位置的最终计算结果 h_t。这里的 o_t 结构和遗忘门一样，表示对输出的内容进行选择，再和经过 tanh 函数缩放的 C_t 相乘得到最后的结果。

$$o_t = \sigma\left(W_o\left[h_{t-1}, X_t\right] + b_o\right) \tag{5}$$

$$h_t = o_t \tanh\left(C_t\right) \tag{6}$$

通过 LSTM，我们得到特征图四个方向上的空间上下文信息，为我们后面判断文本区域和非文本区域提供了有力的帮助。

3.2.2. 方向感知注意力机制

为了有效的以方向感知的方式学习空间上下文信息，本模块进一步在 LSTM 后制定方向感知的注意力机制，以学习注意力权重并生成方向感知的空间上下文特征。该设计和 LSTM 形成了图 7 的方向感知模块。

方向感知注意力机制通过学习有选择的利用聚集在不同方向上的空间上下文信息。首先，我们使用卷积层(具有 1 × 1 内核)以生成四个通道的 W。然后将 W 分成四个注意力权重图，分别表示为 Wleft，Wdown，Wright 和 Wup。再次参见图 7 所示的方向感知模块。四个权重图与相应方向上的空间上下文信息相乘。因此，在训练网络之后，网络学习到适当的注意力权重，以选择性地利用 LSTM 中的空间上下文特征。

方向感知模块应用 LSTM 对特征图沿四个主要方向(左，右，上，下)得到空间上下文信息，应用注意力机制得到四个方向的权重图，将相同方向的空间上下文信息和权重图相乘，再将得到的四个结果连接起来，并使用 1 × 1 卷积将特征维数减少到原来的四分之一。重复上述操作，就可以得到具有丰富全局信息的空间上下文特征。

3.3. 损失函数

损失函数 L 由三部分组成，L_s 表示的是概率图 loss，L_b 表示的是二值图 loss，L_t 表示的是阈值图 loss。其中 α 和 β 分别设为 1.0 和 10。

$$L = L_s + \alpha \times L_b + \beta \times L_t \tag{7}$$

L_s 和 L_b 使用的是二值交叉熵损失函数。L_t 使用的是平均绝对误差损失函数，又称 L1 loss。其中 x_i 是网络的预测值，y_i 是真值。

$$L_s = L_b = \sum_{i \in S_l} y_i \log x_i + \left(1 - y_i\right) \log\left(1 - x_i\right) \tag{8}$$

$$L_t = \sum_{i \in R_d} \left| y_i^* - x_i^* \right| \tag{9}$$

损失函数是衡量网络模型预测的好坏，也就是用来表现预测值和实际值的差距程度。

4. 实验

本次实验编程语言使用 Python3.7 和 Pytorch1.7 框架，在操作系统为 Ubuntu16、内存为 12G、显卡型号为 GeForce RTX 2080 的计算机上进行实验。

4.1. 数据集和评价指标

本文的债券图表数据集的数据源为中国债券信息网。数据集包含了模糊、污染的图片。共计图片 1500 张，图 9 其中包括 1000 张训练图像和 500 张测试图像，使用软件 Labelme 进行标注。

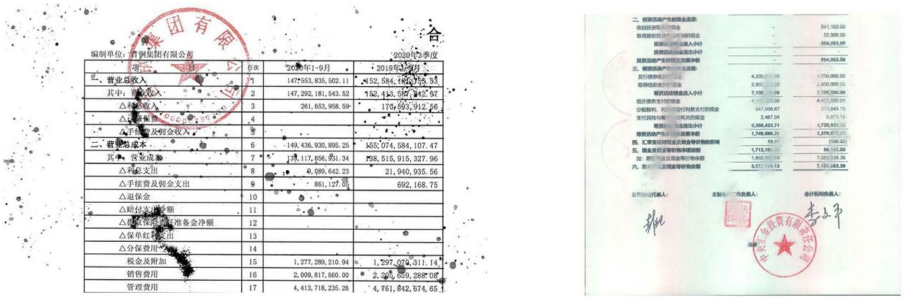


Figure 9. Bond chart data sample diagram
图 9. 债券图表数据示例图

本文使准确率(precision)、召回率(recall)和 F1 作为评价指标，其计算公式如下所示。

$$\text{Precition} = \frac{\text{TruePositive}}{\text{TruePositive} + \text{FalsePositive}} \tag{10}$$

$$\text{Recall} = \frac{\text{TruePositive}}{\text{TruePositive} + \text{FalseNegative}} \tag{11}$$

$$\text{F1} = \frac{2}{1/\text{Precition} + 1/\text{Recall}} \tag{12}$$

其中 TruePositive 表示真实结果和预测结果都为正例，FalseNegative 表示真实结果为正例，预测结果为反例， FalsePositive 表示真实结果为反例，预测结果为正例。正例表示该区域为文本区域，反例表示该区域为非文本区域。

4.2. 实验结果及分析

为了评估本文提出的方法性能，将它和 CTPN、EAST、DB 这三种具有代表性的文本检测方法进行比对。其中 CTPN 是基于回归的深度学习方法，EAST 和 DB 是基于分割的深度学习方法。

Table 1. The method of this paper and the performance of CTPN, EAST and DB on the data set
表 1. 本文的方法和 CTPN、EAST、DB 在数据集上的表现

Method	Precision	Recall	F1
CTPN	86.3	75.8	80.7
EAST	91.8	83.2	87.3
DB	94.7	86.6	90.5
Ours	**98.0**	**91.2**	**94.5**

从表 1 可以看出，本文的方法在准确率(Precision)、召回率(Recall)、F1 都要优与其对比方法。在使用了更强的特征提取网络后，还增加了方向感知模块，能够有效检测到特征量较少的区域，使其在性能上有不错的提升。

为了评估方向感知模块的有效性，本文做了以下消融实验，第一基线(表示为“basic”)是将整个方向感知模块从网络中删除。第二个基线(表示为“basic + context”)考虑了空间上下文，但忽略了方向感知注意力机制。本实验证明了方向感知模块的有效性(表 2)。

Table 2. Ablation results
表 2. 消融实验结果

Method	Precision	Recall	F1
basic	95.7	85.1	90.1
Basic + context	97.0	87.6	92.1
方向感知模块	98.0	91.2	94.5

将债券图片输入模型后，检测得到的结果如图 10 所示。

Figure 10. Bond chart data detection results chart
图 10. 债券图表数据检测结果图

5. 结语

本文针对图像中有模糊、污染的文本区域，提出了一种新的文本检测方法。首先使用了特征提取能力更强、计算量更小的 Swin-Transformer 作为主干网络，又设计了方向感知模块，该模块使用 LSTM 获取不同方向上的空间上下文信息，再和注意力权重图相乘，得到空间上下文特征。最后我们在债券图表数据集上进行了实验，实验结果均优于其它方法。

在未来，我们可以对文本区域破损程度进行研究，如果文本区域特征量损失严重，文字识别的方法已经不能识别该文本，在检测阶段就直接舍弃，减少工作量。

参考文献

[1] 董小君, 宋玉茹. 加快推进我国金融数据治理现代化建设研究[J]. 行政与法, 2022(8): 11-21.

[2] Liu, Z., Lin, Y., Cao, Y., *et al.* (2021) Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows. *IEEE International Conference on Computer Vision* (*ICCV*), Montreal, 10-17 October 2021, 10012-10022. https://doi.org/10.1109/ICCV48922.2021.00986

[3] Elman, J.L. (1990) Finding Structure in Time. *Cognitive Science*, **14**, 179-211. https://doi.org/10.1207/s15516709cog1402_1

[4] Vaswani, A., Shazeer, N., Parmar, N., *et al.* (2017) Attention Is All You Need. 31*st Conference on Neural Information Processing Systems*, Long Beach, December 2017, 6000-6010.

[5] Hochreiter, S. and Schmidhuber, J. (1997) Long Short-Term Memory. *Neural Computation*, **9**, 1735-1780. https://doi.org/10.1162/neco.1997.9.8.1735

[6] Zhong, Y., Karu, K. and Jain, A.K. (1995) Locating Text in Complex Color Images. *Pattern Recognition*, **28**, 1523-1535. https://doi.org/10.1016/0031-3203(95)00030-4

[7] Kim, K.I., Jung, K. and Kim, J.H. (2003) Texture-Based Approach for Text Detection in Images Using Support Vector

Machines and Continuously Adaptive Mean Shift Algorithm. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **25**, 1631-1639. https://doi.org/10.1109/TPAMI.2003.1251157

[8] Minetto, R., Thome, N., Cord, M., Leite, N.J. and Stolfi, J. (2013) T-HOG: An Effective Gradient-Based Descriptor for Single Line Text Regions. *Pattern Recognition*, **46**, 1078-1090. https://doi.org/10.1016/j.patcog.2012.10.009

[9] Kim, Y. (2014) Convolutional Neural Networks for Sentence Classification. *Proceedings of the* 2014 *Conference on Empirical Methods in Natural Language Processing* (*EMNLP*), Doha, October 2014, 1746-1751. https://doi.org/10.3115/v1/D14-1181

[10] Wang, T., Wu, D.J., Coates, A. and Ng, A.Y. (2012) End-to-End Text Recognition with Convolutional Neural Networks. *Proceedings of the* 21*st International Conference on Pattern Recognition*, Tsukuba, 11-15 November 2012, 3304-3308.

[11] Jaderberg, M., Vedaldi, A. and Zisserman, A. (2014) Deep Features for Text Spotting. *Proceedings of the* 13*th European Conference on Computer Vision*, Zurich, 6-12 September 2014, 512-528. https://doi.org/10.1007/978-3-319-10593-2_34

[12] Epshtein, B., Ofek, E. and Wexler, Y. (2010) Detecting Text in Natural Scenes with Stroke width Transform. *Proceedings of* 2010 *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, San Francisco, 13-18 June 2010, 2963-2970. https://doi.org/10.1109/CVPR.2010.5540041

[13] Matas, J., Chum, O., Urban, M. and Pajdla, T. (2004) Robust Wide-Baseline Stereo from Maximally Stable Extremal Regions. *Image and Vision Computing*, **22**, 761-767. https://doi.org/10.1016/j.imavis.2004.02.006

[14] Tian, Z., Huang, W.L., He, T., He, P. and Qiao, Y. (2016) Detecting Text in Natural Image with Connectionist Text Proposal Network. *Proceedings of the* 14*th European Conference on Computer Vision*, Amsterdam, 11-14 October 2016, 56-72. https://doi.org/10.1007/978-3-319-46484-8_4

[15] Zhou, X.Y., Yao, C., Wen, H., *et al*. (2017) EAST: An Efficient and Accurate Scene Text Detector. *Proceedings of* 2017 *IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, 21-26 July 2017, 5551-5560. https://doi.org/10.1109/CVPR.2017.283

[16] Ma, J.Q., Shao, W.Y., Ye, H., *et al*. (2018) Arbitrary-Oriented Scene Text Detection via Rotation Proposals. *IEEE Transactions on Multimedia*, **20**, 3111-3122. https://doi.org/10.1109/TMM.2018.2818020

[17] Zhang, Z., Zhang, C.Q., Shen, W., *et al*. (2016) Multi-Oriented Text Detection with Fully Convolutional Networks. *Proceedings of* 2016 *IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, 27-30 June 2016, 4159-4167. https://doi.org/10.1109/CVPR.2016.451

[18] Deng, D., Liu, H.F., Li, X.L. and Cai, D. (2018) Pixellink: Detecting Scene Text via Instance Segmentation. *Proceedings of the* 32*nd AAAI Conference on Artificial Intelligence*, (*AAAI*-18), *the* 30*th Innovative Applications of Artificial Intelligence* (*IAAI*-18), *and the* 8*th AAAI Symposium on Educational Advances in Artificial Intelligence* (*EAAI*-18), New Orleans, 2-7 February 2018, 6773-6780. https://doi.org/10.1609/aaai.v32i1.12269

[19] Liao, M.H., Wan, Z.Y., Yao, C., Chen, K. and Bai, X. (2020) Real-Time Scene Text Detection with Differentiable Binarization. *Proceedings of the* 34*th AAAI Conference on Artificial Intelligence*, *AAAI* 2020, *the* 32*nd Innovative Applications of Artificial Intelligence Conference*, *IAAI* 2020, *the* 10*th AAAI Symposium on Educational Advances in Artificial Intelligence*, *EAAI* 2020, New York, 7-12 February 2020, 11474-11481. https://doi.org/10.1609/aaai.v34i07.6812

[20] Lint, Y., Dollar, P., Girshick, R., *et al*. (2017) Feature Pyramid Networks for Object Detection. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, 21-26 July 2017, 2117-2125. https://doi.org/10.1109/CVPR.2017.106