

文章编号: 1004-4353(2022)03-0235-07

基于 HRCenterNet 模型改进的 朝鲜语古籍文字检测方法

郭洪壮, 金小峰

(延边大学 工学院, 吉林 延吉 133002)

摘要: 为减少朝鲜语古籍中的小文字错检和漏检问题,提出了一种基于 HRCenterNet 模型改进的朝鲜语古籍文字检测方法.首先,将 HRCenterNet 中 Bottleneck 模块的 3×3 卷积运算替换为 Involution 算子,即将 Bottleneck 模块替换为 Involution-Bottleneck 模块.其次,通过引入 ECA(efficient channel attention)注意力机制扩展 Involution-Bottleneck 模块,并由此提出了基于 IENeck 模块的 HRCenterNet 改进模型.最后,利用朝鲜语古籍数据集对改进的 HRCenterNet 模型和原模型分别进行了训练,并测试了其在不同 IOU 下的准确率、召回率以及 F1 等指标.实验结果表明,在 $IOU \geq 0.6$ 时,改进的 HRCenterNet 模型在朝鲜语古籍数据集上的准确率、召回率和 F1 指标均优于原模型,且 IOU 值越高模型的检测效果越好.这表明改进的 HRCenterNet 模型显著优于原模型,可应用于朝鲜语古籍文字的检测中.

关键词: 古籍文字检测; 朝鲜语古籍; Involution 算子; ECA 注意力机制

中图分类号: TP391.1

文献标识码: A

Korean ancient books character detection method based on improved HRCenterNet model

GUO Hongzhuang, JIN Xiaofeng

(College of Engineering, Yanbian University, Yanji 133002, China)

Abstract: In order to reduce the misdetection and omission of small characters in Korean ancient books, an improved Korean ancient book character detection method based on HRCenterNet model was proposed. First, replace the 3×3 convolution operation of the Bottleneck module in HRCenterNet with the Involution operator, that is, replace the Bottleneck module with the Involution-Bottleneck module. Second, by introducing the efficient channel attention (ECA) mechanism to extend the Involution-Bottleneck module, an improved HRCenterNet model based on the IENeck module is proposed. Finally, the improved HRCenterNet model and the original model are trained separately using the Korean ancient book dataset, and the precision, recall, and F1 of different models under different IOUs are tested. The experimental results show that when $IOU \geq 0.6$, the precision, recall and F1 index of the improved HRCenterNet model on the Korean ancient book data set are better than the original model, and the higher the IOU value, the better the detection effect of the model. This shows that the improved HRCenterNet model is significantly better than the original model, and can be applied to the detection of Korean ancient books.

Keywords: text detection of ancient books; Korean ancient book; Involution operator; ECA mechanism

收稿日期: 2022-04-12

基金项目: 延边大学外国语言文学世界一流学科建设项目(18YLPY14); 国家社会科学基金重大项目(18ZDA306)

第一作者: 郭洪壮(1998—), 男, 硕士研究生, 研究方向为计算机视觉.

通信作者: 金小峰(1970—), 男, 硕士, 教授, 研究方向为语音信息处理、计算机视觉、机器人技术.

0 引言

古籍数字化是利用和保护古籍文献的一种有效方法^[1],但由于古籍文献受当时印刷技术的限制以及老化、破损等原因,准确检测和切分每个字符存在较大困难^[2].传统的文本检测方法主要分为两种:一种是基于连通域的分析方法^[3-4].该方法主要是通过提取图像中的连通区域来缩小文本的搜索区域,并以此标记和找出图像中的相邻像素,但该方法容易受到图片背景噪声的影响.另一种方法是基于滑动窗口的方法^[5].该方法流程简单,并具有良好的检测召回率,但存在计算量较大的问题(由于该方法会得到大量的候选窗口).近年来,一些学者将基于深度学习的方法应用于古籍文字的检测中,并取得了较好成果^[6-9].2020年,Tang等^[10]提出了一种将无锚框目标检测方法和并行网络相结合的 HRCenterNet 模型,研究显示该模型可有效提高中文古籍文字的检测效果.由于古籍中的汉字和朝鲜文字具有相似的几何形态结构和相同的书写顺序^[11],所以本文在 HRCenterNet 模型的基础上通过引入 Involution 算子^[12]和 ECA(efficient channel attention)注意力机制^[13]提出了一种朝鲜语古籍文字的检测方法,并通过实验验证了该方法的有效性.

1 相关理论

1.1 HRCenterNet

HRCenterNet^[10]网络主要由并行多分辨率卷积和重复多分辨率融合两部分组成,共包含 4 个阶段,如图 1 所示.

Stage 1 首先,利用 3×3 卷积组成的 Stem 模块将输入的原图像分辨率降低到原来的 $1/4$,并将其作为网络的第 1 个分支;然后,利用 4 个 Bottleneck^[14]残差单元提取特征.最后,对第 1 个分支使用步长为 2 的 3×3 卷积进行下采样操作,以此生成一个低分辨率分支.

Stage 2 首先,利用 Basicblock 对 Stage 1 输出的两个分支的特征图进行特征提取. Basicblock 可增加网络深度,进而可提高模型提取字符特征的能力;然后,对两个不同分辨率的分支进行多尺度融合,并通过下采样生成一个低分辨率分支.

Stage 3 再次利用 Basicblock 进行特征提取,以此再次生成一个低分辨率分支,并对其余分支进行多次多尺度融合,由此得到 4 个分支的输出.

Stage 4 首先,在每个分支中再次利用 Basicblock 进行特征提取,并将不同的分支进行多次多尺度融合;然后,将 4 个不同尺度的特征图进行融合,以此得到语义信息丰富的特征图.

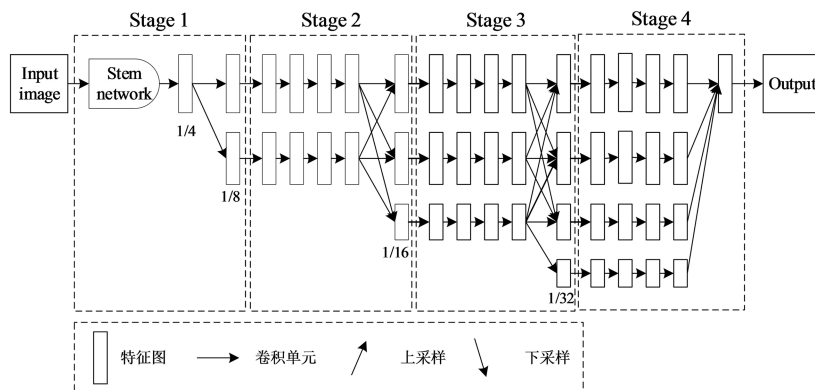


图 1 HRCenterNet 网络结构

由以上 4 个阶段可知,网络经过子网络分支的融合后,模型可以从子网络分支中获得不同感受野的特征,进而可获得高分辨率的特征图,并可最终通过 Stage 4 的输出预测出字符目标的类别、中心点坐标

以及宽高偏移量.

1.2 Involution 算子

Involution 算子^[12]是一种具有空间特异性和通道无关性的算子,其网络结构如图 2 所示. 其中:空间特异性有助于模型通过特征图生成内核参数,进而可提高网络的特征提取;通道无关性可以减少通道数,从而可以降低模型的计算量并防止模型过拟合.

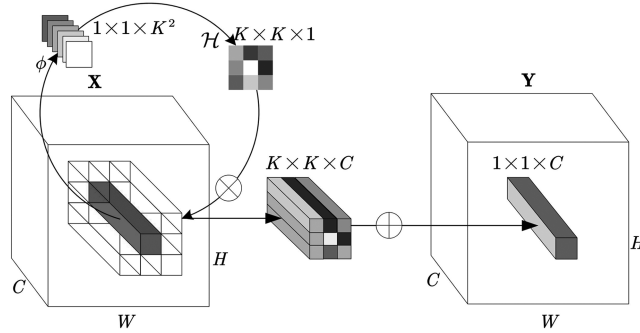


图 2 Involution 算子的网络结构

Involution 算子的核为 $\mathcal{H} \in \mathbf{R}^{H \times W \times K \times K \times G}$ (图 2 中 $G=1$), 其中 K 是 Involution 算子的核大小, H 和 W 分别是特征图的高度和宽度. Involution 算子的定义为:

$$Y_{i,j,k} = \sum_{(u,v) \in \Delta K} \mathcal{H}_{i,j,u+[K/2],v+[K/2],[kG/C]} X_{i+u,j+v,k}. \quad (1)$$

Involution 算子的核形状主要由输入特征图 \mathbf{X} 决定. 假设核生成函数为 ϕ , 且该函数将每个位置 (i,j) 映射为 $\mathcal{H}_{i,j}$, 则 $\mathcal{H}_{i,j}$ 可表示为 $\mathcal{H}_{i,j} = \phi(\mathbf{X}_{\phi_{i,j}})$, 其中 $\phi_{i,j}$ 是以 (i,j) 为中心的邻域像素集合.

为简化核生成函数 ϕ 的形式, 本文将每一个单独的像素 $\mathbf{X}_{i,j}$ 定义为:

$$\begin{aligned} \phi: \mathbf{R}^C &\mapsto \mathbf{R}^{K \times K \times G}, \\ \phi_{i,j} &= \{(i,j)\}, \\ \phi(\phi_{i,j}) &= \phi(\mathbf{X}_{i,j}) = \mathbf{W}_1 \sigma(\mathbf{W}_0 \mathbf{X}_{i,j}). \end{aligned} \quad (2)$$

其中, $\mathbf{W}_0 \in \mathbf{R}^{\frac{C}{r} \times C}$ 和 $\mathbf{W}_1 \in \mathbf{R}^{(K \times K \times G) \times \frac{C}{r}}$ 分别代表两个线性变换, r 是通道缩减比率, σ 是 BN 和 ReLU 操作.

Involution 算子通过核生成函数 ϕ 对特征图中的像素 (i,j) 的特征向量进行运算, 以此生成 $K \times K \times G$ 的特征图, 其中 $K \times K$ 是内核大小, G 是共享内核的数量. Involution 模块的算法为:

输入: 特征图集合为 \mathbf{X}_c , 其空间维度为 $H \times W$, $\mathbf{W}_0 \in \mathbf{R}^{\frac{C}{r} \times C}$, $\mathbf{W}_1 \in \mathbf{R}^{(K \times K \times G) \times \frac{C}{r}}$

输出: 输出 $\mathbf{Y}_{i,j,k}$

Step 1 采用式(2) 计算单个像素, 以此得到 Involution 核;

Step 2 对 Involution 核进行特征重标定: $1 \times 1 \times K^2 \times G \rightarrow K \times K \times G$;

Step 3 采用式(1) 计算输入特征后返回 $\mathbf{Y}_{i,j,k}$.

1.3 ECA 注意力机制

卷积神经网络中的注意力机制是一种提升网络资源分配的方法, 该机制主要包括空间注意力机制和通道注意力机制. 由于 ECA 注意力机制^[13]可以在不降维的条件下比 SE 通道注意力机制^[15]能够获得更好的模型性能, 因此本文在 HRCenterNet 模型中引入 ECA 注意力机制来提高模型对朝鲜语古籍文字的检测精度.

ECA 注意力模块的结构如图 3 所示, 其具体工作流程为: ① 根据全局平均池化获得 C 个特征通道数, 并自适应地确定核大小 k ; ② 对特征通道进行一维卷积, 并通过 sigmoid 函数将特征信息映射为 C 个标量; ③ 将输出的 C 个标量与输入特征加权后的通道信息作为网络注意的特征. ECA 注意力模块采

用生成可变卷积核 k 的方式产生不同的通道权重。
假设 k 和 C 之间存在一种非线性映射 ($C = \phi(k) = 2^{(\gamma \times k - b)}$), 则在每个通道的维数为 C 的情况下, ECA 注意力模块的参数 k 可由式 $k = \psi(C) = \left\lfloor \frac{\log_2 C}{\gamma} + \frac{b}{\gamma} \right\rfloor_{\text{odd}}$ 自适应地确定, 其中 $\lfloor t \rfloor_{\text{odd}}$ 是指与数值 t 最近的奇数, $\gamma = 2, b = 1$.

2 HRCenterNet 模型的改进

2.1 Involution-Bottleneck 模块

在模型中叠加更多的卷积层虽然可以提高目标特征的提取能力, 但由此增加的网络参数会增加计算的复杂度, 且容易使模型出现过拟合现象; 所以, 本文在不增加卷积层的前提下引入 Involution 算子, 将原 HRCenterNet 模型中的 Bottleneck 模块 (见图 4(a)) 中的 3×3 卷积运算替换为 Involution 算子, 即将 Bottleneck 模块替换为 Involution-Bottleneck 模块 (见图 4(b)), 以此通过 Involution 来自适应地调整不同空间位置上的权重系数, 提高朝鲜语古籍文字的全局特征提取能力。

2.2 IENeck 模块

HRCenterNet 模型虽然具有多分辨率并行以及可多尺度段融合操作的特点, 但是该模型缺少注意力机制, 因此本文对 Involution-Bottleneck 模块进行扩展, 即通过引入 ECA 注意力机制建立了 IENeck 模块 (见图 5)。将 ECA 注意力添加在第 2 个 1×1 卷积即可实现原始输入信息的无损传输和自适应地学习全局信息, 进而可提高 HRCenterNet 模型的特征提取能力。

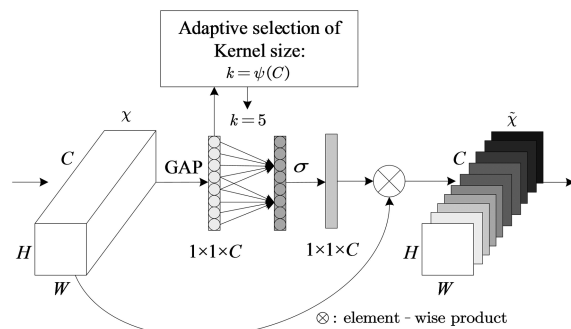
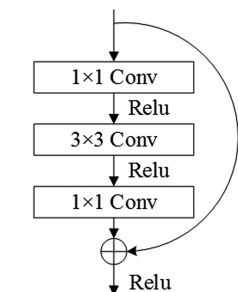
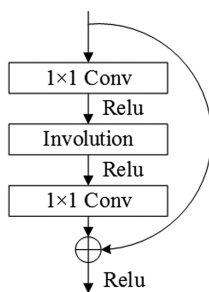


图 3 ECA 注意力的模块结构



(a) Bottleneck 模块



(b) Involution-Bottleneck 模块

图 4 Bottleneck 模块(a) 和 Involution-Bottleneck 模块(b) 的结构

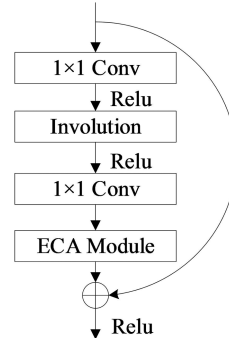


图 5 IENeck 模块的结构

在 IENeck 模块中, ECA 注意力的输入为 C 个大小为 $H \times W$ 的特征图 (记为 $U = [U_1, U_2, \dots, U_C]$)。ECA 的计算过程为:

首先, 对输入特征图 U 进行全局平均池化 (如式 (3) 所示), 以此将 U 转换成 C 个大小为 1×1 的输出特征 Z :

$$Z = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W U_c(i, j). \quad (3)$$

其次, 利用一维卷积运算和 sigmoid 激活函数对 Z 进行计算, 由此得到通道权重系数矩阵 ω :

$$\omega = \sigma(\text{C1D}_k(Z)). \quad (4)$$

其中, $\sigma(\cdot)$ 为 sigmoid 激活函数, $\text{C1D}_k(\cdot)$ 为一维卷积运算, 卷积核大小为 k 。式 (4) 中通道权重系数矩阵 ω 的计算方法即为 ECA。由于式 (4) 在计算过程中仅使用了相邻的 k 个通道, 且 k 的取值是自适应确定的, 因此 ECA 能够实现卷积核尺寸的自适应调整。

3 实验结果与分析

3.1 数据集和实验环境

本文采用人工标注的朝鲜语古籍文本图像数据集进行实验,古籍包括《阐义昭鉴谚解》《谚解胎产集》和《同文类解》.各古籍的文本图像个数分别为 555、160 和 160 张,共计 875 张文本图像.图像中的文本主要为纯朝鲜语文本(图 6(a))和汉朝文字混排的文本(图 6(b)和图 6(c)).其中:《阐义昭鉴谚解》的字符全部为朝鲜文字(一个朝鲜文字的下方可能有一个或多个并列的朝鲜文字),《谚解胎产集》为朝汉文字混合排列(一列或多列汉字后可能出现多列朝鲜文字),《同文类解》在同一列内为朝汉文字混合排列(汉字的字符大于朝鲜文字的字符).

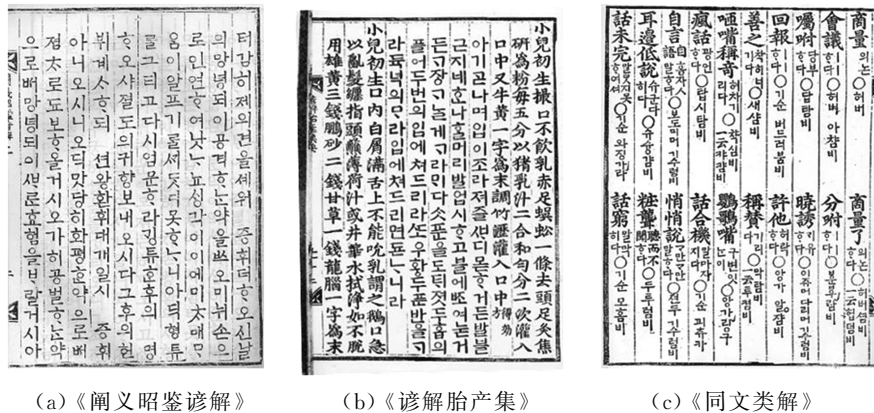


图6 朝鲜语古籍文本图像的示例

实验软件环境为 Python3.8.3 + Pytorch1.2, 硬件环境为 Intel i7-7820X 3.6 GHz CPU + 16 G RAM, 工作站为 NVIDIA RTX2080 GPU(8 GB 显存).将数据集划分为训练集、验证集和测试集,比例为 8 : 1 : 1.模型训练使用 Adam 优化算法,学习率为 1×10^{-6} ,输入图像的大小为 512×512 .在测试时,模型通过载入训练损失值最小的权重对测试集进行测试.评价性能指标采用 IOU(intersection-over-union)、准确率(简记为 P)、召回率(简记为 R)和 F1-score(简记为 F1).本文参考文献[16]的方法划分 IOU. IOU 越大表示古籍文字的判别标准越高,即算法的准确率随 IOU 的增大而减小.

3.2 检测实验结果

为了验证改进 HRCenterNet 模型的有效性,本文将该模型与 HRCenterNet 原模型进行了实验对比.实验时将整张朝鲜语古籍图像作为输入(不进行任何文本行分割),输出为所有单独检测到的字符位置信息.实验结果见表 1.由表 1 可以看出:当 $IOU \geq 0.5$ 时,改进模型的 3 项指标均低于原模型,其中检测错误主要发生在朝汉文字混排的古籍文本图像样本上;当 IOU 在 0.6~0.8 范围内时,改进模型的 3 项指标均优于原模型,且 IOU 越高改进模型的性能提高得越显著.由此表明,改进模型对朝鲜语古籍文字的检测效果更好.

表1 HRCenterNet 模型改进前后的对比实验结果

%

| 模型 | IOU0.5 | | | IOU0.6 | | | IOU0.7 | | | IOU0.8 | | |
|---------------------------------|--------|-------|-------|--------|-------|-------|--------|-------|-------|--------|-------|-------|
| | P | R | F1 | P | R | F1 | P | R | F1 | P | R | F1 |
| HRCenterNet 原模型 ^[10] | 99.34 | 99.04 | 99.19 | 98.04 | 97.74 | 97.89 | 92.62 | 92.33 | 92.47 | 73.44 | 73.21 | 73.32 |
| 改进 HRCenterNet 模型 | 99.22 | 99.02 | 99.12 | 98.11 | 97.91 | 98.01 | 93.71 | 93.52 | 93.61 | 77.91 | 77.75 | 77.83 |

注: IOU0.5 的含义是 $IOU \geq 0.5$, 其他 IOU 值类似.

图 7 和图 8 是对朝汉文字混排且文字大小不同的测试样本进行检测的结果. 对比图 7 和图 8 可以看出,改进模型的检测效果显著优于原模型,由此进一步表明改进模型更适合于检测朝鲜语古籍文字.



图 7 原模型的测试结果



图 8 改进模型的测试结果

3.3 消融实验

为了评估改进方法中 Involution、ECA 和 IENeck 模块的增益效果,本文以 HRCenterNet 模型为基线模型进行了消融实验,结果见表 2.

表 2 各模块的消融实验结果

%

| 模块 | | IOU0.5 | | | IOU0.6 | | | IOU0.7 | | | IOU0.8 | | |
|------------|-----|--------|-------|-------|--------|-------|-------|--------|-------|-------|--------|-------|-------|
| Involution | ECA | P | R | F1 | P | R | F1 | P | R | F1 | P | R | F1 |
| | | 99.34 | 99.04 | 99.19 | 98.04 | 97.74 | 97.89 | 92.62 | 92.33 | 92.47 | 73.44 | 73.21 | 73.32 |
| ✓ | | 99.22 | 98.81 | 99.01 | 97.94 | 97.53 | 97.73 | 93.25 | 92.86 | 93.05 | 75.86 | 75.54 | 75.70 |
| | | -0.12 | -0.23 | -0.18 | -0.10 | -0.21 | -0.16 | 0.63 | 0.53 | 0.58 | 2.42 | 2.33 | 2.38 |
| | ✓ | 99.27 | 99.02 | 99.14 | 98.03 | 97.79 | 97.91 | 93.62 | 93.39 | 93.50 | 76.98 | 76.79 | 76.88 |
| | | -0.07 | -0.02 | -0.05 | -0.01 | 0.05 | 0.02 | 1.00 | 1.06 | 1.03 | 3.54 | 3.58 | 3.56 |
| ✓ | ✓ | 99.22 | 99.02 | 99.12 | 98.11 | 97.91 | 98.01 | 93.71 | 93.52 | 93.61 | 77.91 | 77.75 | 77.83 |
| | | -0.12 | -0.02 | -0.07 | 0.07 | 0.17 | 0.12 | 1.09 | 1.19 | 1.14 | 4.47 | 4.54 | 4.51 |

由表 2 可以看出:测试 Involution 模块时,当 $IOU \geq 0.7$ 时模块的 P 、 R 、 $F1$ 值均优于基线模型,且其增益效果随着 IOU 的提高而提高. 测试 ECA 注意力模块时,当 $IOU \geq 0.6$ 时模块的 P 、 R 、 $F1$ 值均优于基线模型,且其增益效果随着 IOU 的提高而提高. 其原因是 ECA 注意力模块比 Involution 模块能更好地注意到图像中的字符特征信息. 测试 IENeck 模块时,当 $IOU \geq 0.6$ 时模块的 P 、 R 、 $F1$ 值均优于 Involution 模块和 ECA 注意力模块. 其原因是 IENeck 融合了 Involution 和 ECA 注意力模块的正向积极作用.

4 结论

实验结果表明,本文提出的基于 HRCenterNet 模型改进的朝鲜语古籍文字检测方法不仅在空间上能够获取全局信息,而且在通道上可抑制无用信息. 当 $IOU \geq 0.8$ 时,改进模型检测文字的准确率达到了 77.91%,显著优于 HRCenterNet 模型,因此该方法对于检测朝鲜语古籍文本具有很好的应用价值. 在今后的研究中,我们将尝试采用小样本学习的方法对模型进行训练,以期进一步提高朝鲜语古籍文字检测的准确率.

参考文献:

- [1] 姜哲,马少平,夏莹.大型中文古籍《四库全书》自动版面分析系统[J].中文信息学报,2000,14(2):14-20.
- [2] KIM M, OH I. Script-Free text line segmentation using interline space model for printed document images[C]//2011 International Conference on Document Analysis and Recognition, Beijing, China: IEEE, 2011:1354-1358.
- [3] KIM S H, JEONG S, LEE G S, et al. Word segmentation in handwritten Korean text lines based on gap clustering techniques[C]//Proceedings of Sixth International Conference on Document Analysis and Recognition, Seattle, USA: IEEE, 2001:189-193.
- [4] WAHYONO, JO K H. A clustering strategy for touching characters in Korean and English printed text segmentation[C]//International Conference on Ubiquitous Robots and Ambient Intelligence, Daejeon: IEEE, 2012:23-25.
- [5] 靳简明,丁晓青,彭良瑞,等.印刷维吾尔文本切割[J].中文信息学报,2005,18(5):76-83.
- [6] YANG H, JIN L, HUANG W, et al. Dense and tight detection of Chinese characters in historical documents: data-sets and a recognition guided detector[J]. IEEE Access, 2018,6:30174-30183.
- [7] XIE Z, HUANG Y, JIN L, et al. Weakly supervised precise segmentation for historical document images[J]. Neurocomputing, 2019,350:271-281.
- [8] WU S H, WANG J P, MA W H, et al. Precise detection of Chinese characters in historical documents with deep reinforcement learning[J]. Pattern Recognition, 2020,107:107503.
- [9] MA W, ZHANG H, JIN L, et al. Joint layout analysis, character detection and recognition for historical document digitization[C]//2020 17th International Conference on Frontiers in Handwriting Recognition (ICFHR). Dortmund, Germany: IEEE, 2020:31-36.
- [10] TANG C W, LIU C L, CHIU P S. HRCenterNet: An anchorless approach to Chinese character segmentation in historical documents[C]//2020 IEEE International Conference on Big Data, Atlanta, GA: IEEE, 2020:1924-1930.
- [11] 薛春寒,金小峰.基于迁移学习的少样本朝鲜语古籍文字的识别方法[J].延边大学学报(自然科学版),2021,47(4):350-355.
- [12] LI D, HU J, WANG C, et al. Involution: Inverting the inherence of convolution for visual recognition[J]. arXivpreprint, arXiv:2103.06255, 2021.
- [13] WANG Q, WU B, ZHU P, et al. ECA-Net: Efficient channel attention for deep convolutional neural networks [C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, WA, USA: IEEE, 2020:11531-11539.
- [14] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[J]. IEEE, 2016:770-778.
- [15] HU J, SHEN L, ALBANIE S, et al. Squeeze-and-Excitation networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017:214-232.
- [16] 杨海林.基于深度学习的古籍文字识别和检测[D].广州:华南理工大学,2019.