

基于SCA-DBNet的西夏古籍文字检测研究

郑廷帅, 白科, 史伟

(宁夏大学信息工程学院, 宁夏银川 750021)

摘要: 针对西夏文字的笔画复杂、结构相似性高以及西夏文古籍存在缺字、褪色导致检测精度不理想、文字漏检和错检等问题, 综合分析现有的主流文字检测方法, 提出一种基于SCA-DBNet的西夏文检测方法。首先, 引入空间与通道重构卷积模块(ScConv)替换MobileNetV3特征提取骨干网络中的传统卷积模块(Conv), 从而减少计算冗余, 提升模型的特征表达能力。同时, 在特征融合阶段引入自适应尺度融合模块(ASF), 以解决特征融合过程中不同层特征的尺度不一致问题, 并设计混合注意力模块(CFAM)替换原有空间注意力模块(SAM)来增强对多尺度变化特征的刻画, 从而提高西夏古籍文字检测的准确性。在自建西夏文数据集上进行对比实验, 该模型相比基础模型的准确率与召回率分别提升了3.2%和2.9%, 相比其他文字检测模型的准确率和召回率也有一定提升, 能够准确地检测西夏文。

关键词: 文字检测; 空间和通道重构卷积; 自适应尺度融合; 注意力机制; 西夏古籍

DOI: 10.11907/rjtk.251308

扫描二维码阅读全文:

中图分类号: TP391.1

文献标识码: A

文章编号: 1672-7800(2026)007-0119-08



Research on Text Detection in Tangut Ancient Books Based on SCA-DBNet

ZHENG Tingshuai, BAI Ke, SHI Wei

(School of Information Engineering, Ningxia University, Yinchuan 750021, China)

Abstract: Aiming at the problems such as the complex strokes and high structural similarity of Tangut characters, as well as the existence of missing characters and fading in Tangut ancient books, resulting in unsatisfactory detection accuracy, missed detection and incorrect detection of characters, a Tangut character detection method based on SCA-DBNet is proposed by comprehensively analyzing the existing mainstream character detection methods. This method first introduces the Spatial and Channel Reconstruction Convolution module (ScConv) to replace the original convolution module (Conv) and combines it with the MobileNetV3 feature extraction network to reduce computational redundancy and improve the feature expression ability of the model. Meanwhile, in the feature fusion stage, the Adaptive Scale Fusion module (ASF) is introduced to solve the problem of inconsistent scales of features in different layers during the feature fusion process, and the Hybrid Attention module (CFAM) is designed to replace the original Spatial attention module (SAM) to enhance the characterization of multi-scale varying features, thereby improving the accuracy of text detection in Tangut ancient books. Compared with the basic model on the self-built Tangut text dataset, the accuracy rate and recall rate have increased by 3.2% and 2.9% respectively. Compared with other text detection models, there have been certain improvements in both accuracy rate and recall rate, and it can accurately detect Tangut text.

Key Words: text detection; spatial and channel reconfiguration convolution; adaptive scale fusion; attention mechanism; Tangut ancient books

0 引言

西夏文字是记录西夏历史、社会发展的重要载体^[1]。西夏古籍文字检测是西夏古籍OCR(Optical Character Rec-

ognition)中的重要一项。西夏古籍多为纸质或木质材料, 由于长期受到自然与人为因素影响, 出现了潮湿、虫蛀、酸化等问题, 导致背景与文字分界不清晰、残破缺失、纸页发黄等, 无法得到有效的保护与利用^[2]。因此, 对于西夏古籍的数字化保护迫在眉睫。目前基于深度学习的文字检

收稿日期: 2025-05-23

基金项目: 国家自然科学基金项目(62166030)

作者简介: 郑廷帅(2000-), 男, 宁夏大学信息工程学院硕士研究生, 研究方向为文物数字化; 白科(1999-), 男, 宁夏大学信息工程学院硕士研究生, 研究方向为文物数字化; 史伟(1967-), 女, 博士, 宁夏大学信息工程学院教授、硕士生导师, 研究方向为模式识别与智能系统、文物数字化。本文通讯作者: 史伟。

测方法可分为两种类型:一种是基于回归的文字检测方法,一种是基于分割的文字检测方法^[3]。

基于回归的文字检测方法与目标检测算法比较类似,图像中的文本视为待检测目标,其余部分视为背景。Tian等^[4]充分考虑文本检测长度不固定的问题,提出了CTPN(Connectionist Text Proposal Network)。该网络使用VGG(Visual Geometry Group)模型进行特征提取,并使用垂直锚点机制,有效提升了检测精度^[5]。Liu等^[6]提出一个端到端、可训练的多方向快速定向文本定位网络FOTS(Fast Oriented Text Spotting),在检测与识别任务中共享了卷积特征层。Liao等^[7]基于SSD(Single Shot Multibox Detector)模型提出了TextBoxes网络。该网络对目标检测器SSD进行了调整,将其默认文本框更改为适应文本方向和宽高比规格的四边形。但是该网络仅能检测水平方向的文字,并且字符间距过大,检测效率较低^[8]。因此,Liao等^[9]作了进一步改进,提出TextBoxes++网络。该网络在SSD网络的输出层后面通过预测回归,以四边形或倾斜矩形框出任意方向的文本。

基于分割的文字检测方法主要依赖于像素级别的分割技术,通过识别图像中的文本像素来检测文本位置和形状^[10]。Wang等^[11]考虑到文本长度和文本间距等问题,提出了PSENet(Position Sensitive Embedding Network)网络。

该网络采用渐进尺度扩展方法,提升了文本间距过近时的检测精度,但检测耗时较长。Liao等^[12]提出DBNet(Differentiable Binarization Network)网络,该网络将可微分二值化模块DB(Differentiable Binarization)引入分割模型,使模型能够通过自适应的阈值图进行二值化。

上文提到的文字检测模型在一般的文字检测任务中表现了较好的效果,但在西夏文古籍检测领域的应用较少,且检测效果不佳。针对上述问题,本文提出一种基于SCA-DBNet网络的西夏古籍文字检测方法。首先,使用MobileNetV3网络模型构建特征提取主干网络,并引入空间和通道重构卷积模块替换原有卷积层,可减少计算冗余,同时提升了模型的特征表达能力。其次,利用自适应尺度特征融合模块集成低级空间特征和高级语义信息,从而提高西夏古籍文字检测的准确性。

1 DBNet算法

DBNet是一种基于分割的文字检测算法,其采用一种新颖的自适应二值化方法。DBNet网络结构如图1所示。该方法通过网络学习来确定每个像素点的二值化阈值,将二值化过程完全融入到网络训练中,从而有效提高了最终输出图像对阈值变化的稳定性。

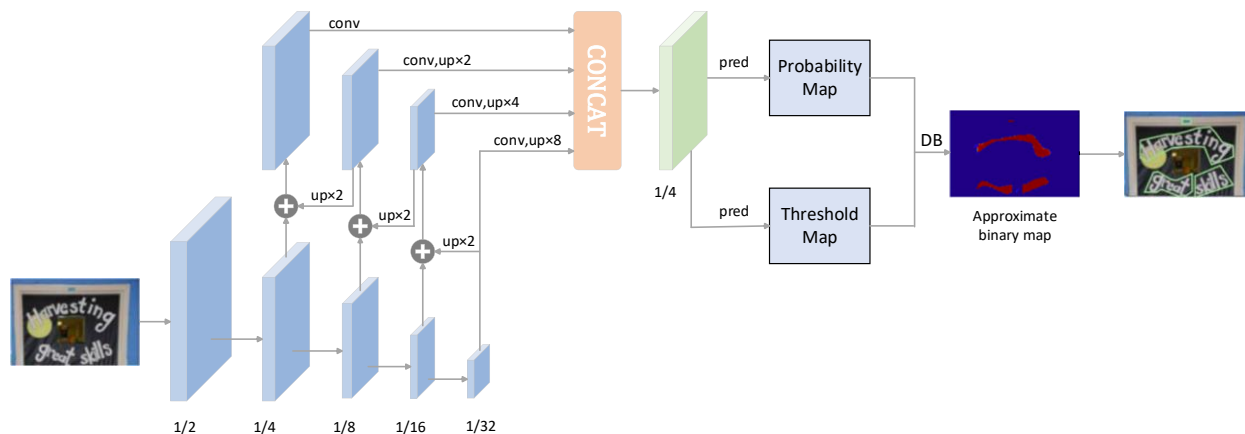


Fig. 1 DBNet network structure

图1 DBNet网络结构

1.1 特征金字塔网络

特征金字塔网络(Feature Pyramid Network, FPN)结构常被用于提升目标检测和分割模型性能^[13]。在DBNet模型中,FPN结构可以同时利用深层语义信息与浅层细节信息,通过融合不同层的特征达到预测的效果。具体过程为:先将输入的原始图片进行一次卷积计算,得到原图比例的1/2大小,再进行4次下采样,依次得到原图比例1/4、1/8、1/16、1/32大小的特征图。之后自顶向下对特征图进行两倍上采样,并与自底向上对应的相同大小特征图相加。最后进行3×3的卷积并对每层结果进行相应的上采样,最终得到原图1/4大小的特征图。

1.2 可微分二值化

DBNet算法采用一种使用近似阶跃函数的二值化方法实现了可微分二值化,其核心是DB(Differentiable Binarization)模块^[14]。可微分二值化公式如式(1)所示。

$$\hat{B}_{i,j} = \frac{1}{1 + e^{-k(P_{i,j} - T_{i,j})}} \quad (1)$$

式中, $\hat{B}$ 表示近似的二值图, P 表示概率图, T 表示从网络中学习的自适应阈值图, k 表示放大因子。可微分二值化在训练期间可以与分割网络一起进行优化,使其能区分文本与背景,并能分离紧密连接的文本实例。

1.3 损失函数

DBNet网络的损失函数如式(2)所示。

$$L = L_s + \alpha \times L_b + \beta \times L_t \quad (2)$$

式中, L 表示总损失, L_s 表示概率图的损失, L_b 表示二值图的损失, L_t 表示阈值图的损失, α 和 β 表示权重系数。根据损失的数值, 将 α, β 设置为 1.0 和 10。 L_s 和 L_b 都应用了二元交叉熵 (Binary Cross-Entropy, BCE) 损失, 如式 (3) 所示。

$$L_s = L_b = \sum_{i \in S_i} y_i \log x_i + (1 - y_i) \log (1 - x_i) \quad (3)$$

式中, S_i 表示正样本与负样本比例为 1:3 的采样集, x_i 表示预测为文字区域的概率值, y_i 则表示实际的标签值。 L_t 为阈值图损失, 如式 (4) 所示。

$$L_t = \sum_{i \in R_d} |y_i^* - x_i^*| \quad (4)$$

式中, R_d 表示扩张多边形内像素的一组索引, y_i^* 表示

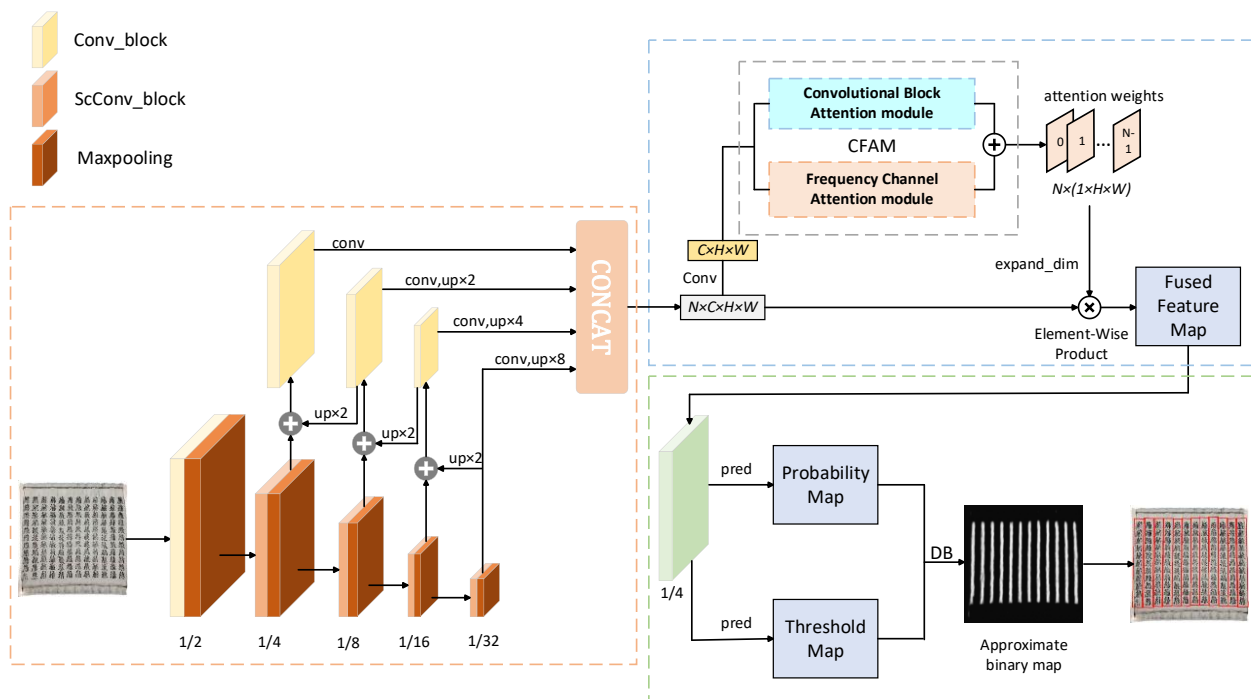


Fig. 2 SCA-DBNet network structure

图2 SCA-DBNet网络结构

2.1 ScConv 模块

为了提升主干网络对文字区域细节特征的提取能力, 并减少背景中冗余信息对文字特征的影响, 提高模型在西夏文检测任务中的准确率, 本文引入 ScConv 模块替换原有的 Conv 模块^[15]。ScConv 由空间重构单元 (Spatial Reconstruction Unit, SRU) 和信道重构单元 (Channel Reconstruction Unit, CRU) 组成, 模块结构如图 3 所示。SRU 能通过组归一化和门控机制分离重要特征与背景噪声, 有效抑制纸张纹理、狐斑等无关信息。CRU 采用分割—变换—融合策略, 压缩低效通道, 提升文字区域关键通道的权重。输入特征首先通过 SRU, 获得空间细化特征 X^w , 然后利用 CRU 运算获得信道细化特征。给定一个特征映射 $X \in \mathbb{R}^{N \times C \times H \times W}$, 减去平均值 μ 并除以标准差 σ 来标准化输

入特征 X , 如式 (5) 所示。

$$X_{out} = GN(X) = \gamma \frac{X - \mu}{\sqrt{\sigma^2 + \varepsilon}} + \beta \quad (5)$$

其中, μ, σ 是 X 的均值和标准差, ε 是为了避免分母为零时造成异常而加入的一个小的正常数, γ 和 β 是可训练参数。GN 层中的可训练参数 $\gamma \in \mathbb{R}^C$ 用于计算每个批次中各通道的空间像素方差。归一化相关权重 $W_\gamma \in \mathbb{R}^C$ 表示不同特征映射的重要性, 如式 (6) 所示。

$$W_\gamma = \{w_i\} = \frac{\gamma_i}{\sum_{j=1}^C \gamma_j}, i, j = 1, 2, \dots, C \quad (6)$$

然后将 W_γ 重新加权的特征映射权值通过 sigmoid 函数映射到 (0, 1) 范围内, 并通过阈值进行门控。这里将阈值以上的权重设置为 1, 得到信息权重 w_1 ; 将阈值以下的

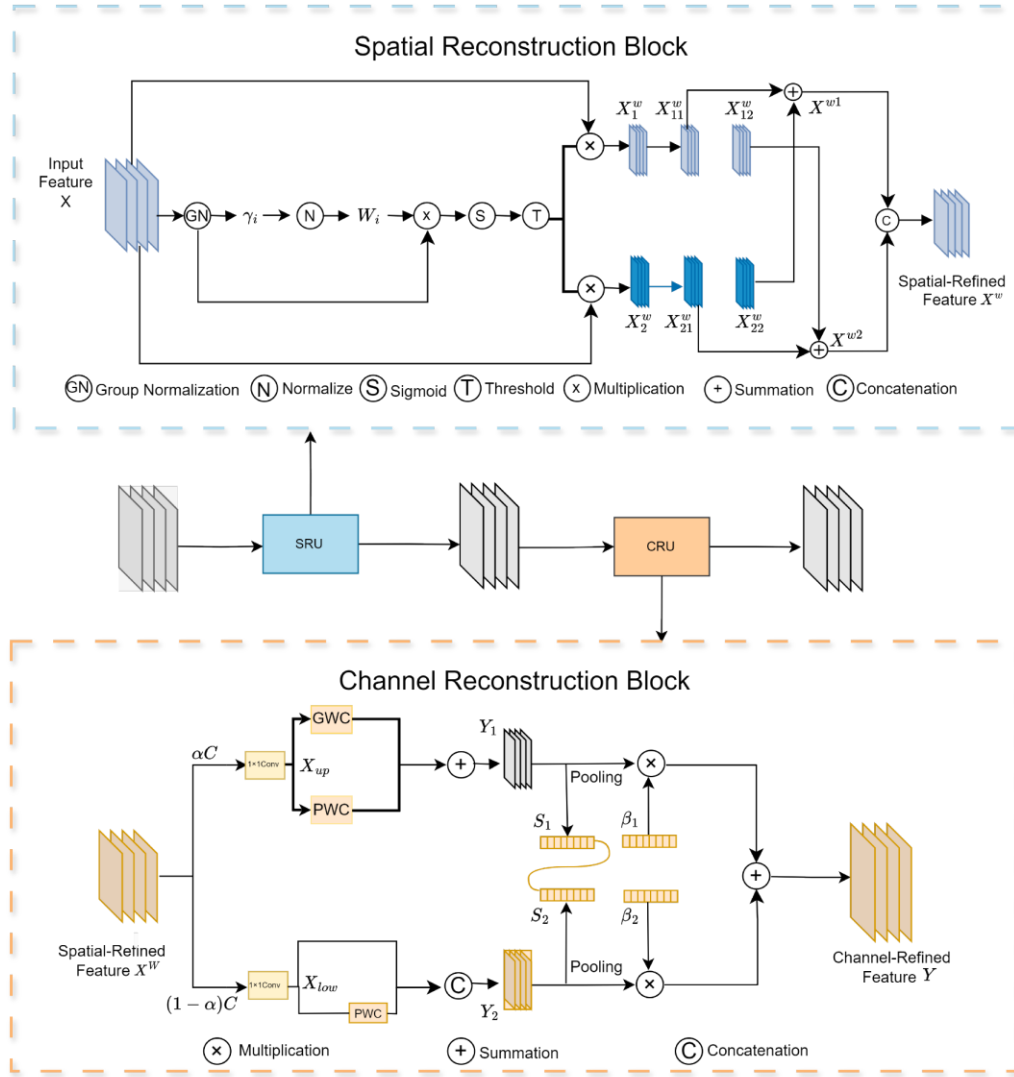


Fig. 3 ScConv module structure

图3 ScConv模块结构

权重设置为0,得到非信息权重 W (实验中阈值设置为0.5)。获取 W 的整个过程如式(7)所示。

$$W = \text{Gate}\left(\text{Sigmoid}\left(W_{\gamma}(\text{GN}(X))\right)\right) \quad (7)$$

将输入特征 X 分别乘以 W_1 和 W_2 ,得到两个加权特征:信息量较大的特征 X_1^w 和信息量较小的特征 X_2^w 。重构操作中将信息丰富的特征与信息较少的特征相加,生成信息更丰富的特征,从而节省空间。采用交叉重构运算,将加权后两个不同的信息特征充分结合,加强其之间的信息流。之后将交叉重构的特征 X^{w1} 和 X^{w2} 在通道维度上进行拼接操作,得到空间精细特征映射 X^w 。具体过程如式(8)所示。

$$\begin{cases} X_1^w = W_1 \otimes X \\ X_2^w = W_2 \otimes X \\ X_1^w \oplus X_2^w = X^{w1} \\ X_{21}^w \oplus X_{12}^w = X^{w2} \\ X^{w1} \cup X^{w2} = X^w \end{cases} \quad (8)$$

其中, $\otimes$ 是逐元素的乘法, $\oplus$ 是逐元素的求和, $\cup$ 是Concatenation操作。将SRU应用于中间输入特征 X 后,将

信息较少的特征分离并再次重构,增强其判别性特征,抑制空间维度上的冗余特征。然而,空间精细特征映射 X^w 在通道维度上仍然是冗余的。

最后,在CRU单元中将输入的空间细化特征 X^w 分割成两部分:一部分为通道数 αC 的特征,另一部分为通道数 $(1-\alpha)C$ 的特征。之后利用 1×1 卷积对两部分特征的所有通道进行压缩处理,分别得到 X_{up} 和 X_{low} 。在转换操作中,将输入的 X_{up} 作为“富特征提取”的输入,分别进行GWC (Groupwise Convolutional Filter) 和PWC (Pointwise Convolutional Filter),相加得到输出 Y_1 。将输入 X_{low} 作为“富特征提取”的补充进行PWC,并将得到的结果和原来的输入取并集得到 Y_2 。在融合操作中,采用一种简化的SKNet实现对特征图 Y_1 和 Y_2 的融合。通过全局平均池化操作,将其中的空间信息与通道统计信息进行整合,得到池化后的特征表示 S_1 和 S_2 。再对其进行Softmax函数操作,以此生成特征权重向量 β_1 和 β_2 。最终,依据这些特征权重向量产生输出 $Y = \beta_1 Y_1 + \beta_2 Y_2$, Y 即为通道提炼的

特征。

2.2 CFAM 模块

注意力机制可以提升模型对输入信息中关键信息的捕捉能力,并减少干扰信息的影响。本文将通道—空间注意力模块 CBAM (Convolutional Block Attention Module) 与多谱通道注意力模块 FcaNet (Frequency Channel Attention Network) 进行并联相加,设计了一种混合注意力模块 CFAM,如图 4 所示,用其替换自适应尺度融合模块 (ASF) 中的 SAM 注意力模块^[16]。

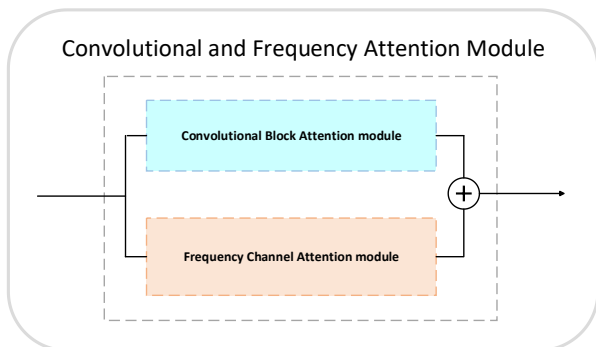


Fig. 4 CFAM module structure

图 4 CFAM 模块结构

CBAM 模块结构如图 5 所示,该注意力机制融合了通道注意力模块 (CAM) 和空间注意力模块 (SAM)。

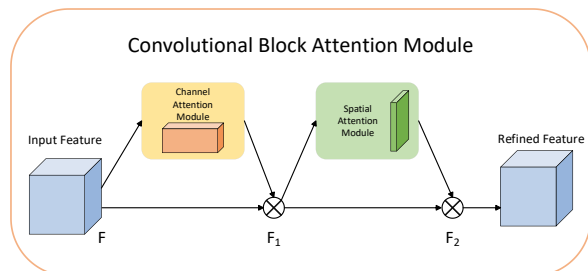


Fig. 5 CBAM module structure

图 5 CBAM 模块结构

通道注意力模块 (CAM) 先对输入特征图进行最大值池化和平均值池化,得到特征向量。之后将特征向量输入到共享感知器,得到最大特征向量和平均特征向量。将两个特征向量逐元素相加,经过 Sigmoid 函数进行归一化,再与输入特征图相乘,得到输出特征图。

空间注意力模块 (SAM) 以通道注意力模块 (CAM) 的输出特征图作为输入,并对输入进行最大值池化和平均值池化操作,得到两个特征图。随后将得到的特征图沿着通道维度进行拼接,再使用 7×7 的卷积核进行降维。最后通过 Sigmoid 函数进行归一化,再与特征图相乘,得到输出特征图。

FcaNet 注意力机制融合了频域信息与空域信息,其通过对输入特征图进行频域变换操作,以及通过低通滤波器和高通滤波器对空间域特征进行分解,得到低频和高频成分,并分别捕捉全局语义信息与局部细节信息^[17]。对这两部分特征分别应用通道注意力机制生成对应的注意力权

重,从而增强关键通道的特征响应。将加权后的低频和高频特征进行傅里叶逆变换,还原到空间域并与原始特征融合,从而提升模型对多尺度信息的表达能力。

CBAM 注意力机制可动态优化西夏古籍图像的重要区域,将关注点聚焦于重要通道和位置,以增强对文字区域的敏感性。FcaNet 可以自动学习和强调更重要的频率成分,从而更有效地聚焦于对检测任务有益的频率特征。CFAM 将两者提取的特征相加并融合,能够更有效地捕捉到图像西夏文字区域的细节和纹理信息,明确文字与背景分界的划分,避免错检、漏检。同时,减少纸质古籍狐斑发黄背景对于缺色文字的影响。

2.3 ASF 模块

丰富的特征信息是准确检测文字区域的基础,主要包括低层和高层特征。为了解决特征融合过程中不同层特征的尺度不一致问题,本文引入自适应尺度融合模块 (ASF),使其能够在空间维度学习不同尺度和不同空间位置的权重。ASF 模块主要由卷积层、空间注意力模块 (SAM) 和元素加权相乘操作组成,引入该模块后使得模型在特征融合过程中能更关注文本区域的相关信息,提高特征表示的鲁棒性^[18]。为了进一步提升 ASF 模块性能,对 ASF 模块进行改进,引入混合注意力模块 CFAM 替换原有的 SAM 注意力模块。改进后的 ASF 模块结构如图 6 所示,其可以有效捕获多尺度特征,减少信息损失。

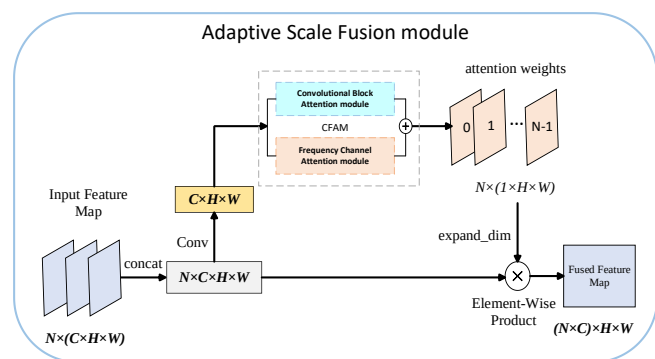


Fig. 6 ASF module structure

图 6 ASF 模块结构

将输入的 N 张特征图进行拼接后,经过一个 3×3 的卷积层获得中间特征图,并将得到的特征图输入到一个 CFAM 混合注意力模块得到注意力权重,再将注意力权重沿通道维度拆分为 N 部分,与相应调整后的特征进行加权相乘,得到 ASF 的输出。

3 实验与结果分析

3.1 数据集

本文使用数据集的图像数据来源于宁夏大学西夏研究院扫描整理的西夏文《孙子兵法》,共收集了 6 000 张原始数据图像,图片格式为 PNG,每张页面大约包含 200 个西夏文字符,标注方法采用手动标注。将训练集、验证集

和测试集按照8:1:1的比例进行划分,数据集图像与标注图像如图7所示。

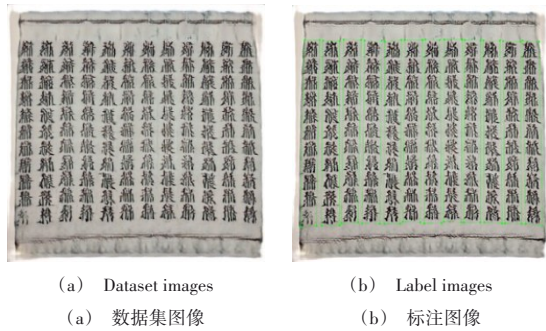


Fig. 7 Dataset images and label images

图7 数据集图像与标注图像

3.2 实验设置

本文的实验环境如表1所示。训练时使用具有自适应能力的Adam优化器作为模型优化器,并采用反向传播方法动态调整模型参数。设置初始学习率为0.001,批处理大小为16,模型训练迭代次数为300,损失函数为二分类交叉熵函数。通过网络学习和分析不断迭代更新权重最小化输出与真实值之间的误差,从而获得最优文字检测模型来识别图像中的文字区域。

Table 1 Experimental environment

表1 实验环境

配置	参数
操作系统	Windows11
GPU	NVIDIA RTX A6000
CPU	Intel(R)Core(TM)i9-13900K@3.00GHz
内存	64 G
显存	48 G

为了客观地评价西夏文字检测模型的有效性,本文使用精确率(Precision)、召回率(Recall)以及调和平均数(F1-score)作为评价指标。计算公式如式(9)一式(11)所示。

$$\text{Precision} = \frac{TP}{TP + FP} \quad (9)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (10)$$

$$\text{F1_score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (11)$$

式中,TP表示被正确预测为西夏文的像素,TN表示被正确预测为背景的像素,FP表示将背景错误预测为西夏文的像素,FN表示将西夏文错误预测为背景的像素。

3.3 消融实验

为了验证DBNet模型所用的改进策略对性能的影响,在西夏文数据集上开展了消融实验。实验结果如表2所示。

从表2可以看出,加入ScConv模块使模型的准确率、召回率与F值分别提升了0.6%、1.5%和1.3%。该模块通过减少计算冗余并提升模型的特征表达能力,提高了检测

准确率。ASF模块使模型的准确率、召回率与F值分别提升了0.4%、1.1%和1.0%。之后,通过引入CFAM模块进一步改进ASF模块,相较于基础模型的准确率、召回率和F值分别提升了1.6%、2.3%和2.2%。实验结果表明,ASF模块能够自适应地捕获多尺度特征,通过在该模块中引入CFAM模块进一步提升了效果,加强了对多尺度变化特征的融合能力。最后,通过整合ScConv模块和ASF模块,模型的准确率、召回率和F值分别提升了3.2%、2.9%和3.3%,证明了通过结合两个模块,使整体网络能够有效地提高特征表示,减少信息损失,进一步提升西夏古籍文字检测效果,3个性能指标相比原模型均有显著提升。

Table 2 Ablation experiment results

表2 消融实验结果

(%)

算法	Precision	Recall	F-measure
DBNet	93.6	87.1	90.0
DBNet+SC	94.2	88.6	91.3
DBNet+ASF	94.0	88.2	91.0
DBNet+ASF-CF	95.2	89.4	92.2
DBNet+SC+ASF-CF	96.8	90.0	93.3

3.4 消融实验可视化

图8为改进后的DBNet模型和增加不同模块的DBNet模型对西夏文古籍的检测效果。由图8可见,相比之下基础模型对于缺字、褪色的文字区域检测效果较差,在增加ASF模块后对于缺色文字的检测效果有一定提升。在分别增加ScConv模块和改进ASF模块后的可视化检测结果与本文模型检测结果无明显差别。结果表明,本文模型、增加ScConv模块后的模型与增加改进ASF模块后的模型均能较好地检测出缺字、褪色的文字区域,三者仅在定量分析结果上有所差别。

3.5 实验结果对比分析

为了验证改进后的DBNet网络模型相比目前主流网络模型在检测效果方面的显著优势,本文将改进后的模型与EAST^[19]、PSENet、FCENet^[20]、DBNet、PAN^[21]和TextFuseNet^[22]文字检测网络模型在西夏文数据集上进行对比实验,结果如表3所示。

从表3可以看出,原DBNet模型和EAST模型相比,在准确率、召回率和F值上有明显提升,但和其他较新的基于分割的文字检测模型相比,准确率、召回率和F值差距不大,且略低于TextFuseNet。本文模型相比原DBNet模型和其他较新的文字检测模型,在准确率、召回率和F值上均取得了更好效果,表明改进后的网络模型更加稳定,减少了文字漏检和背景误判现象。

3.6 对比实验可视化

图9为改进后的DBNet模型和其他主流模型对西夏文古籍的检测效果。由图9可见,EAST、PSENet、PAN对于颜色较浅的褪色区域检测效果较差,FCENet、DBNet和TextFuseNet能够检测出褪色区域,但对缺字区域的检测效果较差。本文改进后的模型在清晰检测出文字区域的同时,

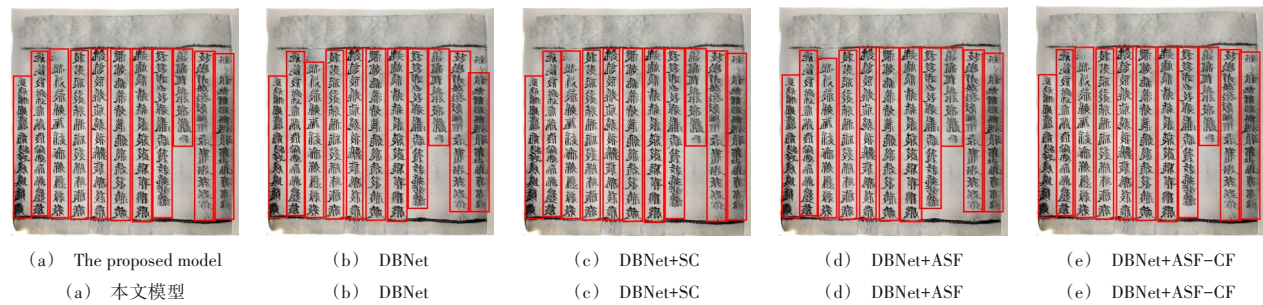


Fig. 8 Comparison of Tangut script detection results
图 8 西夏文检测结果比较

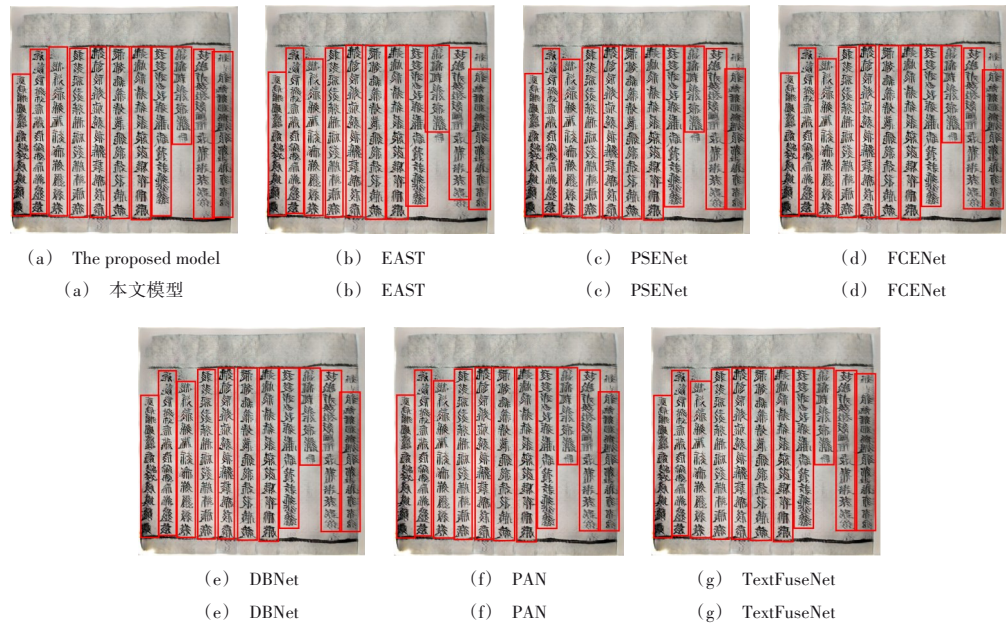


Fig. 9 Tangut script detection results of different methods
图 9 不同方法的西夏文检测结果

Table 3 Experimental results of different detection algorithms

表 3 不同检测算法的实验结果 (%)			
算法	Precision	Recall	F-measure
EAST	85.6	74.5	79.7
PSENet	92.4	86.1	89.1
FCENet	92.7	86.0	89.2
DBNet	93.6	87.1	90.0
PAN	91.2	84.6	87.8
TextFuseNet	94.2	88.2	91.1
DBNet+SC+ASF-CF	96.8	90.0	93.3

对于缺字、褪色的文字区域也具有较好的检测效果。由此可以看出,本文改进的模型对于褪色的西夏文也可以进行有效检测,并能清晰地划分出文字区域,相比其他主流模型效果明显提升。

4 结语

针对西夏文古籍文字检测精度不高、文字漏检和错检等问题,本文以 DBNet 模型为基础,通过改进 FPN 网络引入 ScConv 模块,以减少计算冗余,提升计算效率,并增强模

型的特征表达能力。同时,引入 ASF 模块,并用 CFAM 混合注意力模块替换原有的 SAM 模块,加强了多尺度变化特征的融合,降低误检率和漏检率。实验结果表明,本文算法取得了 96.8% 的准确率、90.0% 的召回率和 93.3% 的 F 值,均优于其他算法。本文为提高西夏文古籍检测精度提供了一种有效方法,具有一定的理论价值与较好的实际应用前景。后续研究将继续优化模型结构,进一步提高西夏文古籍检测精度和效率。

参考文献:

[1] FENG J H. A brief discussion on the Tangut script calligraphy in the Chinese character cultural circle[J]. Journal of Chuxiong Normal University, 2013, 28(2): 89-93.
冯继红. 浅谈汉字文化圈中的西夏文书法[J]. 楚雄师范学院学报, 2013, 28(2): 89-93.
[2] MENG Y F. Research on several problems of digital informatization of Xixia characters[D]. Beijing: Beijing Jiaotong University, 2019.
孟一飞. 西夏文字数字信息化若干问题研究[D]. 北京: 北京交通大学, 2019.
[3] LI Y H, CHEN Y Y. A review of deep learning-based scene text detection

- methods[J]. Computer Engineering and Applications, 2021, 57(6): 42-48.
- 李益红,陈袁宇. 深度学习场景文本检测方法综述[J]. 计算机工程与应用, 2021, 57(6): 42-48.
- [4] TIAN Z, HUANG W, HE T, et al. Detecting text in natural image with connectionist text proposal network[C]// Proceedings of the European Conference on Computer Vision, 2016: 56-72.
- [5] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [DB/OL]. <https://arxiv.org/pdf/1409.1556v2>.
- [6] LIU X, LIANG D, YAN S, et al. FOTS: fast oriented text spotting with a unified network[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 5676-5685.
- [7] LIAO M, SHI B, BAI X, et al. TextBoxes: a fast text detector with a single deep neural network[C]// Proceedings of the 31st AAAI Conference on Artificial Intelligence, 2017: 4161-4167.
- [8] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector[C]// Proceedings of the European Conference on Computer Vision, 2016: 21-37.
- [9] LIAO M, SHI B, BAI X. TextBoxes++: a single-shot oriented scene text detector[J]. IEEE Transactions on Image Processing, 2018, 27(8): 3676-3690.
- [10] FU X R. Research on oracle bone inscription detection method based on deep learning[D]. Changchun: Jilin University, 2024.
- 付新然. 基于深度学习的甲骨文字检测研究方法研究[D]. 长春: 吉林大学, 2024.
- [11] WANG W, XIE E, LI X, et al. Shape robust text detection with progressive scale expansion network[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 9336-9345.
- [12] LIAO M, WAN Z, YAO C, et al. Real-time scene text detection with differentiable binarization[C]// Proceedings of the AAAI Conference on Artificial Intelligence, 2020: 11474-11481.
- [13] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 2117-2125.
- [14] WANG H X, WU J L, CHEN D S. Binarization method for document images with complex backgrounds[J]. Journal of Image and Graphics, 2023, 28(7): 2011-2025.
- 王红霞,武甲礼,陈德山. 分离复杂背景下的文档图像二值化方法[J]. 中国图象图形学报, 2023, 28(7): 2011-2025.
- [15] LI J, WEN Y, HE L. SCConv: spatial and channel reconstruction convolution for feature redundancy[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 6153-6162.
- [16] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module[C]// Proceedings of the European Conference on Computer Vision, 2018: 3-19.
- [17] QIN Z, ZHANG P, WU F, et al. FcaNet: frequency channel attention networks[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021: 783-792.
- [18] LIAO M, ZOU Z, WAN Z, et al. Real-time scene text detection with differentiable binarization and adaptive scale fusion[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 45(1): 919-931.
- [19] ZHOU X, YAO C, WEN H, et al. East: an efficient and accurate scene text detector[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 5551-5560.
- [20] ZHU Y, CHEN J, LIANG L, et al. Fourier contour embedding for arbitrary-shaped text detection[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 3123-3131.
- [21] WANG W, XIE E, SONG X, et al. Efficient and accurate arbitrary-shaped text detection with pixel aggregation network [DB/OL]. <https://arxiv.org/abs/1908.05900>.
- [22] YE J, CHEN Z, LIU J, et al. TextFuseNet: scene text detection with richer fused features[C]// Proceedings of the Twenty-ninth International Joint Conference on Artificial Intelligence, 2020: 516-522.

(责任编辑:黄健)