

面向文本识别的小样本阴影消除方法

王筋辉, 彭光灵, 段亮, 袁国武, 岳昆

引用本文

王筋辉, 彭光灵, 段亮, 袁国武, 岳昆. 面向文本识别的小样本阴影消除方法[J]. 计算机科学, 2024, 51(9): 147-154.

WANG Jiahui, PENG Guangling, DUAN Liang, YUAN Guowu, YUE Kun. [Few-shot Shadow Removal Method for Text Recognition](#) [J]. Computer Science, 2024, 51(9): 147-154.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[重参数化增强的双模态实时目标检测模型](#)

Re-parameterization Enhanced Dual-modal Realtime Object Detection Model
计算机科学, 2024, 51(9): 162-172. <https://doi.org/10.11896/jsjcx.230700106>

[基于原型修正的小样本半监督语义图像翻译算法](#)

Few-shot Semi-supervised Semantic Image Translation Algorithm Based on Prototype Correction
计算机科学, 2024, 51(8): 224-231. <https://doi.org/10.11896/jsjcx.230500038>

[基于高深约束与边缘融合的单目3D目标检测](#)

Monocular 3D Object Detection Based on Height-Depth Constraint and Edge Fusion
计算机科学, 2024, 51(8): 192-199. <https://doi.org/10.11896/jsjcx.230500071>

[任务感知的多尺度小样本SAR图像分类方法](#)

Task-aware Few-shot SAR Image Classification Method Based on Multi-scale Attention Mechanism
计算机科学, 2024, 51(8): 160-167. <https://doi.org/10.11896/jsjcx.230500171>

[基于伪标签依赖增强与噪声干扰消减的小样本图像分类](#)

Few-shot Image Classification Based on Pseudo-label Dependence Enhancement and Noise Interference Reduction
计算机科学, 2024, 51(8): 152-159. <https://doi.org/10.11896/jsjcx.230500066>

面向文本识别的小样本阴影消除方法

王笏辉 彭光灵 段亮 袁国武 岳昆

云南大学信息学院 昆明 650500

云南省智能系统与计算重点实验室 昆明 650500

(wjh@ynu.edu.cn)

摘要 阴影消除是计算机视觉领域中面对阴影场景的重要任务,旨在检测和消除图像中的阴影区域。由于图像编辑技术受到阴影图像质量的制约,现有方法利用其他任务中的知识和阴影特性来获得更加有效的特征向量,从而实现阴影消除。在带有文本内容的阴影图像中,由于文本颜色和形状等特征不同于前景和背景,因此可能将文本错误地检测为阴影的一部分进而导致错误的阴影消除结果。针对该问题,提出了一种面向文本识别的小样本阴影消除方法。在小样本目标检测基础框架模型中,利用被错误识别为阴影的文本特征生成基类数据和新类数据,增强对该类文本的特征学习;在部分检测框合并算法中,利用文本本身长宽比多样化、变化大的特性,以多个约束为前提合并结构相关性较强的检测框,实现对目标的正确框定。建立在真实数据与合成数据上的实验结果验证了所提方法的有效性。

关键词: 文本识别;阴影消除;阴影检测;小样本学习;目标检测

中图分类号 TP391

Few-shot Shadow Removal Method for Text Recognition

WANG Jiahui, PENG Guangling, DUAN Liang, YUAN Guowu and YUE Kun

School of Information Science and Engineering, Yunnan University, Kunming 650500, China

Yunnan Key Laboratory of Intelligent Systems and Computing, Kunming 650500, China

Abstract Shadow removal is an important task in the field of computer vision, with the goal of detecting and removing shaded regions from shadow regions in images. As image editing techniques are constrained by the quality of shaded images, existing methods exploit the knowledge from other tasks and the properties of shadows to obtain more effective feature vectors for shadow removal. Since the color and shape features of the text differ from the foreground and background in the shaded images, the text may be incorrectly detected as part of the shadows to generate incorrect results. To address this problem, a few-shot shadow removal method for text recognition is proposed. First, the features of the text incorrectly identified as shadows are used to produce base class data and new class data to enhance feature learning of such text in the infrastructure part of the few-shot target detection model. Second, the text itself is used to merge structurally relevant detection frames with multiple constraints to fix the objects correctly in the enhancement part of the detection frame merging algorithm. Experimental results validate the effectiveness of the proposed method on real and synthetic datasets.

Keywords Text recognition, Shadow removal, Shadow detection, Few-shot learning, Object detection

1 引言

阴影作为一种光源被物体遮挡所产生的现象,存在于许多不同的场景中,随着图像数据规模的迅速增长,阴影场景愈加趋向复杂化和多样化。由此带来的图像光照不均,导致图像的局部表面颜色和亮度发生变化、图像质量降低,给计算机

视觉领域中的许多图像处理算法带来了干扰。例如,光学字符识别(Optical Character Recognition, OCR)技术在阴影场景下由于光照不均导致阴影区域的文本往往与非阴影区域的文本存在光照亮度、纹理、色彩等方面的差异,给基于OCR技术的图像文本检测和识别带来了较大的挑战。如何正确消除阴影成为了计算机视觉领域的重要任务和研究热点。

到稿日期:2023-07-31 返修日期:2023-11-23

基金项目:国家自然科学基金(62002311, U23A20298);云南省重点实验室项目(202205AG070003);云南省重大科技专项计划项目(202202AD080001);云南省基础研究计划项目(202201AT070394)

This work was supported by the National Natural Science Foundation of China(62002311, U23A20298), Foundation of Key Laboratory of Yunnan Province (202205AG070003), Major Science and Technology Special Foundation of Yunnan Province(202202AD080001) and Basic Research Project of Yunnan Province(202201AT070394).

通信作者:段亮(duanl@ynu.edu.cn)

带有文本内容的阴影场景通常分为文档阴影场景和自然阴影场景,如图 1 所示。针对文档阴影场景中的阴影,Bako 等^[1]提出了传统文档图像阴影消除方法,Lin 等^[2]提出了基于深度学习的文档图像阴影消除方法。面对具有文本内容的自然阴影场景,常用的方法是利用卷积神经网络(Convolutional Neural Network, CNN)消除图像中的阴影。然而,这些方法往往会因为图像中文本的颜色和几何表征不同于前景和背景的其他物体,容易将文本错误地识别为阴影的一部分,导致无法正确地消除阴影。如图 2 所示,左侧为带有文本的阴影图像原图,中间为基于 CNN 得到的错误检索结果,右侧为应得到的正确结果,其将文本误识别为阴影,同时文本的颜色和形态被破坏。因此,建立适用于该类场景的阴影消除方法、准确检测并消除图像中的阴影亟待研究。

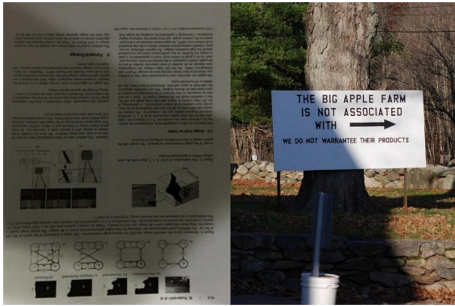


图 1 文档阴影场景示例与自然阴影场景示例

Fig. 1 Examples of document shaded and natural shaded scenes



图 2 基于 CNN 的阴影检测方法示例

Fig. 2 Example of shadow detection based on CNN

基于上述分析,本文提出了一种面向文本识别的小样本阴影消除方法,旨在正确地检测和消除图像中的错误阴影,并得到更完整的检测结果。首先,针对文本被误识别为阴影的样本数量少、类型新(图像中文本结构、颜色和形态等特征被破坏而不同于普通文本)等特点,引入基于小样本学习的目标检测模型;接着,针对文本长宽比变化大、普通检测框无法正确框选的问题,给出目标检测框合并算法,通过对目标物体的准确框选来实现目标检测和阴影消除。

在目标检测公开数据集 Pascal VOC07^[3]、阴影检测公开数据集 ISTD^[4]和 SBU^[5]、当前主流阴影检测模型 DSC^[6]检测自然阴影场景文本图像得到的错误结果以及合成数据集上,将所提出的方法与原小样本检测模型进行比较。在 Pascal VOC07 上,指标 $NAP50$ 从 34.60% 提升到 41.8%;在真实数据集和合成数据集上,指标 BER 分别从 9.931 和 7.821 下降到 5.718 和 5.531。实验结果表明,检测框的合并算法

有效提升了原小样本检测模型的性能,通过将阴影检测方法 DSC 与本文提出的小样本阴影消除方法相结合,可明显改善带有文本内容的自然阴影场景图像的阴影检测效果。

总的来说,本文的主要贡献包括:

1)提出了一种检测框的合并算法,使用 3 个约束准确地判断检测框之间是否存在结构性特征的关联,基于 FSFOD (Frustratingly Simple Few-Shot Object Detection)^[7]构建基于检测框合并的小样本目标检测模型,从而正确框选长宽比多样化、变化大的目标物体。

2)利用基于检测框合并的小样本阴影消除模型,结合 DSC 阴影检测方法,消除由于将不同于背景和前景的文本错误识别为阴影而造成的干扰,在带有文本内容的阴影场景下正确消除图像中的阴影。

3)建立在真实数据与合成数据上的实验结果验证了本文方法的有效性,针对长宽比多样化、变化大的目标物体,本文方法都可有效优化现有 FSFOD 和 DSC 模型。

2 相关工作

下文分别从文档阴影场景和自然阴影场景这两个方面阐述相关工作,并分析国内外的研究现状。

1) 文档阴影场景的阴影消除方法

早期方法基于阴影文档图像的统一背景色来消除文档阴影。Yang 等^[8-10]提出双边滤波方法,通过减少具有类似反射成分的区域亮度对比来消除文档图像中的阴影。Jung 等^[11]提出了“水填充”算法,首先识别文档图像中的阴影区域,然后使用直方图均衡化、色彩校正和对比度增强等技术来消除阴影。该方法效率高,对硬件配置要求低,实用性强。上述方法容易产生颜色偏移,导致结果的亮度高于原图像。Kligler 等^[12]提出了一种可见性检测算法,将文档图像表示为三维点云,能够检测和纠正其模糊、噪音和低对比度等可见性问题,不足之处在于阴影边缘残留、阴影消除结果不佳。Oliveira 等^[13]确定文档图像存在的恒定背景颜色,将直接插值法用于文档图像的阴影消除,通过计算图像中每个小块区域全局和背景颜色局部间的比例来获得阴影图,然后检测背景区域并在剩余部分进行插值,当文档大部分区域被阴影遮挡时,该方法会产生错误的背景区域识别结果。

Lin 等^[2]提出了基于深度神经网络的阴影消除方法 BEDSR-Net,提供了一个由阴影图像、对应无阴影图像、阴影 mask 组成的数据集,并验证了基于合成数据集训练得到的模型在真实数据上的有效性,提出了背景估计网络、阴影去除网络和用于估计文档的全局背景颜色和阴影的注意力网络,实现了图像的阴影消除。该方法使用的参数少且通用性强,产生的结果质量较高,但计算量大,无法处理分辨率高度退化的图像。

2) 自然阴影场景的阴影消除方法

针对包含文本内容的自然阴影场景,现有方法通常将具有文本内容的自然阴影场景图像作为统一的自然阴影场景进行处理。早期的方法基于光照不变原理建立物理模型,以消除阴影。例如,Finlayson 等^[14-15]通过照明模型和颜色信息的物理模型来消除输入图像的阴影,但往往只对宽动态范围

图像有效。另一种方法基于带标签阴影图像的颜色、纹理和边缘等手工特征^[16-19],利用特征描述图像区域并将其分为阴影和非阴影区域,但该方法消除阴影的能力有限,在复杂情况下的处理效果不佳。

鉴于 CNN 在消除阴影方面具有良好的表现,Khan 等^[20]提出了基于 CNN 的阴影检测方法,通过训练两个 CNN 来分别检测阴影区域和阴影边缘,并将输出特征输入到条件随机场(Conditional Random Field)以消除图像中的阴影。Shen 等^[21]提出了一种深度结构化阴影边缘检测器,采用结构化标签来提高预测阴影图的局部一致性。Vicente 等^[22]通过最小化图像簇误差平方和来实现注释恢复,使用带有嘈杂注释的大型数据集训练两个 CNN 以消除阴影。文献^[23]提出了基于生成对抗网络(Generative Adversarial Network, GAN)的阴影消除网络模型,先从全局上生成一个粗糙的阴影消除结果,再利用与阴影相关的残差信息对粗糙的结果在颜色和细节上进行局部优化,从而获得更加真实自然的无阴影图像。Hu 等^[24]提出了方向感知上下文的注意力机制,利用空间循环神经网络(Spatial Recurrent Neural Network)获取二维空间的上下文信息,以提取更好的全局特征并精准判断阴影区域。Wang 等^[25]通过两个堆叠条件 GAN 的相互促进来提升阴影检测和阴影消除任务的性能。Zhu 等^[26]提出了一种新的双射映射网络(BMNet),将阴影去除和阴影生成的学习过程耦合在一个统一的参数共享框架中。通过一致的双向约束和两个过程的同步优化,BMNet 可以在前向阴影去除过程中有效地恢复底层背景内容。Zhu 等^[27]将阴影去除任务重新表述为变分优化问题,设计了一种迭代算法并将其展开到深度神经网络中,增加了模型的可解释性。Luo 等^[28]提出了一种深度交互式学习模型,该模型交互式地优化阴影去除器和阴影检测器,以提高模型的整体能力。

综上所述,虽然基于深度神经网络模型可大大提高阴影消除的精度,但现有方法仍容易将一些深色且区别于背景及前景的目标内容误识别为阴影。现有模型训练数据以一般的自然阴影场景为主,在面对自然阴影场景中的文本时可能会产生错误,将不同于背景和前景的文本错误识别为阴影,使原本非阴影部分的文本区域获得错误的亮度改变,图像质量降低,无法获得正确的阴影修复图像,从而影响图像编辑技术等下游任务的执行。因此,本文针对以文本为主要阴影干扰的真实数据量及类型较少、不同于普通文本结构形态等特点,研究面向文本识别的小样本阴影消除方法,旨在提升模型在自然阴影文本场景的阴影检测准确率;针对阴影中的干扰长宽比变化大、普通目标检测框难以被正确框选的情况,提出了一种目标检测框的合并算法,以实现对长宽比变化大、多样化的目标物体的正确检测。

3 本文方法

本章给出了基于检测框合并的小样本目标检测模型,实现对长宽比变化大的目标物体的正确检测。下文概述小样本目标检测问题,给出了模型框架和目标检测框合并算法。

3.1 小样本目标检测概述

基于小样本学习的目标检测方法只需要提供少量标注

信息,就能够检测出对应的目标对象,利用少量目标域数据集将知识迁移到目标域任务,可有效解决传统目标检测方法进行大量图像标注时耗时耗力的问题。本文考虑到以文本为主要阴影干扰的真实数据量及类型较少的特点,以 FSFOD 模型为基础构建面向文本识别的小样本阴影消除方法。图 3 给出了小样本检测模型 FSFOD 的网络结构^[7]。

FSFOD 采用二级目标检测方法 Faster-RCNN^[29]作为基础检测模型,训练过程使用两阶段微调方法。第一阶段通过基类数据集对特征提取网络和分类器进行联合训练,获得普适性的特征参数,使得模型从基类获取的特征能转移到新类;第二阶段固定整个特征提取网络,将基类数据和新类数据组成平衡数据集,并对检测模型的最后一层(框预测分类器和回归器)进行微调,引入如下余弦相似性函数替代普通全连接方法,以减小类内方差,提高对新类数据的检测精度。

$$S_{i,j} = \frac{\infty F(x)_i^T \omega_j}{\|F(x)_i\| \|\omega_j\|} \quad (1)$$

其中, $S_{i,j}$ 是输入 x 的第 i 个目标建议检测框和类 j 的权重向量之间的相似性得分, ∞ 为比例因子(一般固定为 20)。

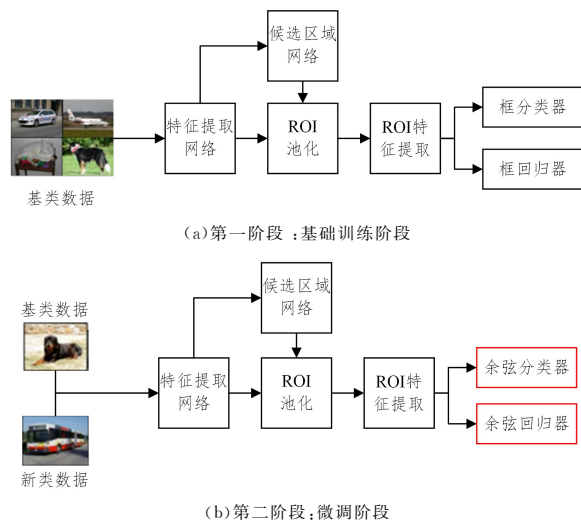


图 3 FSFOD 小样本目标检测网络的结构图

Fig. 3 Network structure of FSFOD few shot object detection

3.2 模型框架

本节给出了基于检测框合并的小样本阴影消除模型框架,如图 4 所示。该模型以小样本目标检测模型 FSFOD 为基础目标检测器,提出了目标检测框的合并算法,正确框定并消除长宽比多样化、类似文本的阴影干扰。该方法分两个阶段执行:

第一阶段为基础训练阶段。为消除基础训练阶段和微调阶段之间的鸿沟,并降低特征偏移对新类别实例检测可能产生的影响,使用文献^[30]中的方法构建与新类数据集具有较强相关性的基础类别数据集,并基于 Faster-RCNN 模型的思想进行训练。

第二阶段为微调阶段。引入 FSFOD 模型中的余弦相似性,结合设计的检测框合并算法进行微调训练。首先,利用以 CNN 为主体的骨干网络对输入图像进行特征提取,通过候选区域网络(Region Proposal Network, RPN)预选目标检测框。然后,设计目标检测框合并算法,根据检测框的结构特点提出

合并约束,设定相关比例参数,合并具有较强相关性的相邻目标检测框,并重新计算目标物体属于前景的概率,以实现对长宽比变化大、多样化的目标物体的检测。最后,将合并后的目标检测框坐标与卷积特征图输入兴趣区域池化层(Region of Interest Pooling)和基于余弦相似性的分类器,从而消除阴影二值图中的干扰阴影。

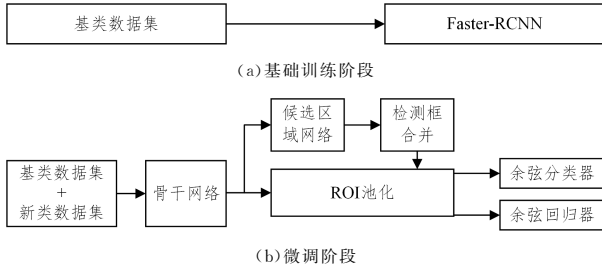


图4 小样本阴影消除模型的框架

Fig. 4 Framework of few-shot shadow removal model

3.3 检测框合并算法

针对长宽比变化大的目标物体,小样本目标检测模型FSFOD无法很好地检测并框定图像中的目标区域,导致漏检或误检。在图5给出的例子中,FSFOD无法准确检测整栋房屋、整列火车等长宽比较大的目标物体,存在框选区域过大或过小、框选不完整和检测准确率低的现。



图5 基于FSFOD模型对长宽比变化大的目标进行检测的示例

Fig. 5 Example of FSFOD based detection of objects with large aspect ratio variations

为解决该问题,本节给出目标检测框合并算法,通过合并相关性强的相邻检测框,完成对此类长宽比变化大的目标物体的完整检测和框定,从而提高准确率。FSFOD中的RPN网络输出一系列候选目标检测框以及相应的检测框坐标值,包括中心坐标 x 和 y ,以及宽 w 、高 h 。因此,对两两相邻的检测框进行合并,每次合并两个检测框(可多次合并)。

假设待合并的两个目标检测框为检测框左框 g_1 和右框 g_2 。下面根据检测框的特性设置如下合并约束。

合并约束1:待合并检测框的IoU值须大于参数 α ,以保证每次尽量选取与标签重合度大的检测框,减少合并次数。

合并约束2:待合并的两个检测框的相隔距离须小于较窄框宽度的一定比例参数 β ,如式(2)所示:

$$\left\{ \begin{array}{l} |x_1 - x_2| - \frac{(w_1 + w_2)}{2} \leq w_{\min} \beta, \\ \quad \text{若 } |x_1 - x_2| \geq \frac{(w_1 + w_2)}{2} \\ |x_1 - x_2| \leq w_{\min} \beta, \\ \quad \text{若 } |x_1 - x_2| < \frac{(w_1 + w_2)}{2} \end{array} \right. \quad (2)$$

其中, x_1, x_2, w_1 和 w_2 分别为检测框 g_1 和 g_2 的横坐标、宽和长,

w_{\min} 为两框中窄框的宽度值。

合并约束3:待合并的两个检测框的纵坐标之差须小于较矮框高度的一定比例参数 γ ,如式(3)所示:

$$|y_1 - y_2| \leq h_{\min} \gamma \quad (3)$$

其中, y_1 和 y_2 分别为检测框 g_1 和 g_2 的纵坐标, h_{\min} 为两框中较矮框的高度值。

约束2和约束3的目的在于避免错误地合并一些相距过远、高度差距较大、相关性不强的相邻检测框。

进一步地,若两个检测框满足以上3个合并约束,则对其进行合并。设定合并后检测框的中心坐标的横坐标和纵坐标为 x_3 和 y_3 ,其表达式分别如式(4)和式(5)所示:

$$x_3 = \left(x_1 - \frac{w_1}{2} \right) + \frac{w_2}{2} \quad (4)$$

$$y_3 = \frac{y_1 + y_2}{2} \quad (5)$$

最后,分别使用式(6)和式(7)设定实现合并后的检测框的宽 w_3 和高 h_3 ,从而得到新的合并检测框 g_3 的中心坐标、宽和高。

$$w_3 = \left(x_2 + \frac{w_2}{2} \right) - \left(x_1 - \frac{w_1}{2} \right) \quad (6)$$

$$h_3 = \frac{h_1 + h_2}{2} \quad (7)$$

上述思想如算法1所示。

算法1 目标检测框合并

输入:输入的候选检测框集合 $G_n = \{g_1, g_2, \dots, g_n\}$;候选框包含的4个坐标值 $g_n = (x_n, y_n, w_n, h_n)$

参数: α, β, γ

输出:合并框4个坐标值 $g_i' = (x_i', y_i', w_i', h_i')$

步骤:

1. List $\leftarrow \emptyset$ //初始化 List,用于存储相邻检测框
2. $i \leftarrow 1$
3. While $i \leq n$ Do
4. $j \leftarrow i + 1$
5. While $j \leq n$ Do
6. If $|y_i - y_j| \leq h_{\min} \gamma$ and $\text{IoU}(g_i, g_j) > \alpha$ and $|x_i - x_j| - \frac{(w_i + w_j)}{2} \leq w_{\min} \beta$ or $(|x_i - x_j| < \frac{(w_i + w_j)}{2})$ and $|x_i - x_{i+1}| \leq w_{\min} \beta$ Then
7. List $\leftarrow \text{List} \cup \{g_j\}$ //若满足约束则入列
8. Else Break
9. End If
10. $j \leftarrow j + 1$
11. End While
12. While List $\neq \emptyset$ Do
13. 利用式(4)~式(7)计算 g_i' 的4个坐标值 x_i', y_i', w_i', h_i'
14. 将已删除合并框且新合并框加入候选框集合
15. $i \leftarrow \min(k; g_k \in G_n \text{ and } k > i)$ //移到下一个框
16. List $\leftarrow \emptyset$ //清空列表
17. End While
18. $i \leftarrow i + 1$
19. End While
20. Return g_i'

算法1的时间复杂度取决于输入数据集的大小,其中 n

表示原始检测框的数量。假设 IoU 和平均坐标、尺寸的计算复杂度为常数时间,则第 1 步初始化列表的时间复杂度为常数;第 3 步执行 n 次;第 5 步最坏情况下执行 $n-i$ 次;第 13 步计算合并框的平均坐标和尺寸,从 G_n 中删除和添加元素的时间复杂度为常数。因此,算法 1 的时间复杂度为 $O(n^2)$ 。

4 实验

本章给出了建立在真实数据与合成数据上的实验结果,分析了本文方法的有效性。

4.1 数据集

本节使用 3 个公开数据集、1 个实验所获数据集及 1 个合成数据集,测试基于检测框合并算法的小样本阴影消除模型的有效性。

数据集 1 为目标检测公开数据集 Pascal VOC07,数据集 2 为阴影检测公开数据集 ISTD 和 SBU,数据集 3 为阴影检测模型 DSC 检测带有文本内容的自然阴影场景图像得到的错误阴影结果数据,数据集 4 为合成测试数据集。由于当前获取到的错误阴影结果数据量较少,因此使用 PIL 库,通过从公开合成文本数据集 SCSD(Synthetic Chinese String Dataset)中随机选取文本图像,并从数据集 ISTD 中选取阴影二值图,将两部分内容进行转二值图、阈值扩散变换、选取文本倍数随机化(分别设置为 0.5 倍、1 倍、2 倍、3 倍)等操作,实现所需测试数据集的合成,得到接近 3 000 个合成测试数据,用于测试本文方法的有效性。

针对数据集 1,根据文献[30]中的方法构建与需检测的干扰阴影相关性强的基类数据集,包括自行车、船、汽车、摩托车、瓶子、椅子、餐桌、盆栽、沙发、电视/显示器、猫、牛、狗、马、羊,将其作为小样本目标检测方法的 15 个基类内容。部分基类数据如图 6 所示。



图 6 部分基类数据示例

Fig. 6 Examples of partial base class data

针对数据集 2,同样根据文献[30]中的方法构建相关性高、长宽比变化大、多样化的新类数据集,随机从数据集 ISTD 和 SBU 中选取公交车、长椅、长屋、长砖、长飞机、长火车、连续树等长宽比变化大、不固定的数据作为小样本目标检测方法的 5 个新类,同时将数据集 3 中得到的错误阴影结果作为新类数据内容之一。部分新类数据如图 7 所示。

同时,分别将数据集 3 和数据集 4 作为测试小样本阴影消除方法有效性的真实数据集和合成数据集。图 8 给出了

数据集 4 中关于数据集 SCSD 中部分随机选取的文本图像示例,图 9 给出了部分合成后的数据集。



图 7 部分新类数据示例

Fig. 7 Examples of partial new class data

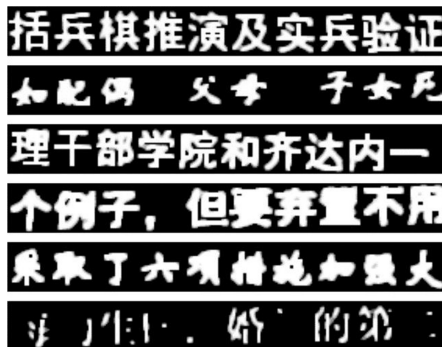


图 8 数据集 SCSD 中随机选取的文本图像示例

Fig. 8 Examples of randomly selected text image from SCSD

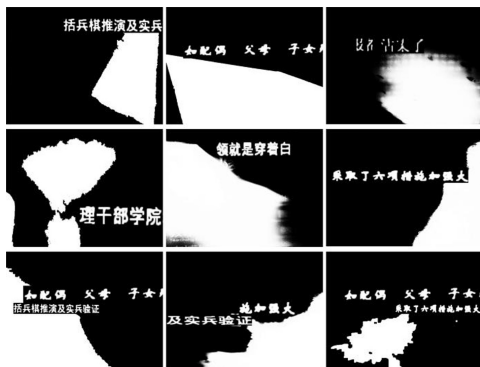


图 9 合成数据集示例

Fig. 9 Examples of synthetic dataset

根据实验中数据的使用情况,表 1 列出了小样本检测模型所需的基类和新类数据集的描述信息,表 2 列出了用于测试模型效果的真实数据集和合成数据集的统计信息。

表 1 小样本数据集统计信息

Table 1 Statistics of few shot datasets

数据集	种类	每类数据数	数据合计
基类	15	10	150
新类	5	10	50

表 2 模型测试数据集统计信息

Table 2 Statistics of test datasets

数据集	数据合计
真实数据	39
合成数据	2996

4.2 实验环境

本文实验环境为 Intel i9-9900K 处理器、16 GB 内存、

Ubuntu 18.04 操作系统,代码用 Python 语言编写,约束参数 α, β 和 γ 分别设置为 0.5, 0.4 和 0.15。

4.3 评价指标

本节从小样本目标检测和阴影检测两个方面对所提方法进行测试。因只关注小样本方法对新类目标的检测准确率,使用 NAP50 (Novel-Class Average Precision 50) 作为小样本测试指标,使用阴影误差率 (Shadow Error Rate, SER)、非阴影误差率 (Non-Shadow Error Rate, NER)、平衡错误率 (Balanced Error Rate, BER) 和准确率 (Accuracy, ACC) 作为阴影检测指标,计算式分别如式(8)~式(12)所示:

$$NAP50 = \frac{TP}{(TP+FP)} \quad (8)$$

$$SER = \left(1 - \frac{TP}{N_p}\right) \times 100 \quad (9)$$

$$NER = \left(1 - \frac{TN}{N_n}\right) \times 100 \quad (10)$$

$$BER = \left(1 - 0.5 \times \left(\frac{TP}{N_p} + \frac{TN}{N_n}\right)\right) \times 100 \quad (11)$$

$$ACC = \frac{TP+TN}{N_p+N_n} \quad (12)$$

其中, TP, FP, TN, N_p 和 N_n 分别表示真阳性、假阳性、真阴性、阴影像素总和、非阴影像素总和。BER 可有效评估类不平衡结果,其值越小表示阴影检测效果越好。

4.4 实验结果

为验证所提出的检测框合并算法的有效性,基于小样本目标检测方法 FSFOD 进行嵌入,实现基于检测框合并的小样本目标检测方法。以阴影检测模型 DSC 为初始模型对带有文本的自然阴影场景图像进行阴影检测,通过提出的小样本阴影消除模型检测并消除 DSC 错误检测的干扰阴影区域。因此,选择 FSFOD 和 DSC 作为本节实验测试的基准算法。基于训练数据构建 FSFOD 和 DSC 模型,基于测试数据集测试 FSFOD 模型的评价指标 NAP50、模型 DSC 的评价指标 SER, NER, BER 和 ACC。

1) 检测框的合并算法对小样本目标检测方法 FSFOD 的有效性

本组实验选取数据集 Pascal VOC07, ISTD 和 SBU 组成的基类数据集和新类数据集作为测试集,采用 10-shot 测试 FSFOD 模型在是否使用检测框合并算法时的 NAP50 的变化情况,如表 3 所列。

表 3 是否使用检测框合并方法的小样本方法 NAP50 对比

Table 3 Comparison of NAP50 for few shot methods with or without test frame merging method (%)	
方法	NAP50
不使用检测框合并	34.60
使用检测框合并	41.80

从表 3 可以看出,使用检测框合并算法之后,FSFOD 的 NAP50 高于不使用合并算法的结果, NAP50 从 34.6% 提升到 41.8%, 整体上升约 7.2%。是否使用检测框合并算法的视觉对比结果如图 10 所示,图 10 中第一行为未使用检测框合并算法的小样本目标检测结果,第二行为使用了检测框合并

算法的小样本目标检测结果。可以看出,使用检测框合并后的方法的准确率得到提升,且框定的目标物体更加完整准确。



图 10 新类数据集上是否使用检测框合并方法的视觉对比
Fig. 10 Visual comparison of whether the detection frame merging method is used on the new class dataset

2) 基于检测框合并的小样本阴影消除模型的有效性

为了测试本文模型的有效性,基于数据集 3 和数据集 4 进行测试。使用小样本阴影消除模型对 DSC 模型进行改进,测试结果如表 4 和表 5 所列。本文模型在真实数据集上的 BER 下降约 4.2, ACC 提升约 3.9; 在大型合成数据集上的 BER 下降约 2.29, ACC 提升约 2.64。是否使用 FSFOD, 以及是否使用 FSFOD 和检测框合并算法的视觉对比如图 11 所示。其中,图 11(a) 给出了仅使用模型 DSC 的检测结果,图 11(b) 给出了使用 FSFOD 的检测结果,图 11(c) 给出了使用 FSFOD 和检测框合并算法的检测结果。可以看出,使用 FSFOD 和检测框合并算法可更加完整地检测并消除错误的阴影文本区域,验证了本文方法的有效性。同时,为了验证模型对模糊图片的阴影消除性能,将图片的模糊度提高 50% 后进行可视化展示。图 11(d) 给出了仅使用模型 DSC 的检测结果,图 11(e) 给出了使用 FSFOD 的检测结果,图 11(f) 给出了使用 FSFOD 和检测框合并算法的检测结果。可以看出,FSFOD 和检测框合并算法具有很强的模糊图像阴影消除能力,能有效消除错误阴影。

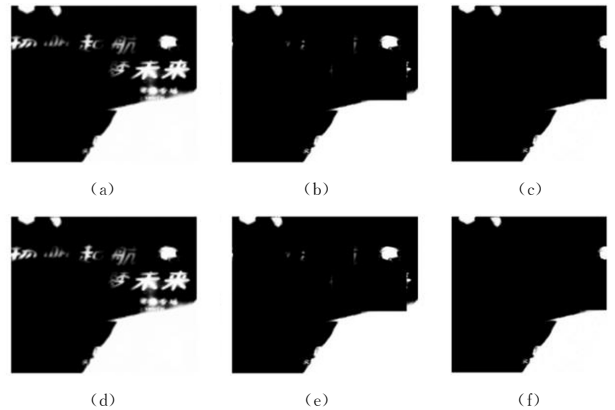


图 11 是否使用模型在数据集上的视觉对比

Fig. 11 Visual comparison with or without the proposed model on dataset

表 4 真实数据集上是否使用 FSFOD 和检测框合并算法的性能对比

Table 4 Performance comparison on real dataset with/without FSFOD and detection frame merging algorithm				
模型	SER	NER	BER	ACC
均不使用	5.585	14.626	9.931	90.187
FSFOD	5.522	10.957	8.239	91.764
FSFOD+检测框合并	5.236	6.201	5.718	94.115

表5 合成数据集上是否使用FSFOD和检测框合并算法的性能对比

Table 5 Performance comparison on synthetic dataset with/without FSFOD and detection frame merging algorithm

模型	SER	NER	BER	ACC
均不使用	7.691	7.922	7.821	91.775
FSFOD	7.364	7.494	7.429	92.231
FSFOD+检测框合并	5.917	5.144	5.531	94.416

4.5 参数分析

为了验证3个合并约束中各参数对小样本阴影消除模型的影响,本节进行了消融实验。

1) IoU值的影响

为验证约束1中IoU值的选取对小样本阴影消除模型的影响,分别将IoU值设置为0.4、0.5、0.6和0.7,并分别在真实数据集和合成数据集上进行测试,结果如表6和表7所列。其中IoU值为0.5时效果最好。

表6 真实数据集上不同IoU值时的性能对比

Table 6 Performance comparison on real dataset with different IoU values

IoU	SER	NER	BER	ACC
0.4	5.554	6.577	6.066	93.791
0.5	5.236	6.201	5.718	94.115
0.6	6.561	7.770	7.166	92.765
0.7	8.068	9.555	8.812	91.230

表7 合成数据集上不同IoU值时的性能对比

Table 7 Performance comparison on synthetic dataset with different IoU values

IoU	SER	NER	BER	ACC
0.4	6.403	5.566	5.984	93.893
0.5	5.917	5.143	5.531	94.416
0.6	6.660	5.789	6.225	93.615
0.7	7.675	6.671	7.173	92.522

2) 检测框相隔距离比例的影响

为验证约束2中相隔距离比例的选取对小样本阴影消除模型的影响,将相隔距离比例分别设置为0、0.05、0.10、0.15和0.20,并分别在真实数据集和合成数据集上进行测试,结果如表8和表9所列,加粗数据为测试最优比例,其中比例为0.15时取得测试最优值。

表8 真实数据集上不同相隔距离比例的性能对比

Table 8 Performance comparison on real dataset with different separation distance scales

比例	SER	NER	BER	ACC
0	6.370	5.537	5.954	93.929
0.05	6.331	5.499	5.915	93.972
0.1	6.125	5.227	5.676	94.249
0.15	5.917	5.144	5.531	94.416
0.2	5.956	5.152	5.554	94.389

表9 合成数据集上不同相隔距离比例的性能对比

Table 9 Performance comparison on synthetic dataset with different separation distance scales

比例	SER	NER	BER	ACC
0	6.119	7.246	6.683	93.216
0.05	6.130	7.259	6.694	93.205
0.1	5.670	6.715	6.192	93.673
0.15	5.236	6.201	5.718	94.115
0.2	5.474	6.483	5.9785	93.872

3) 检测框纵坐标差比例的影响

为验证约束3中检测框纵坐标差比例选取对小样本阴影消除模型的影响,将纵坐标差比例分别设置为0、0.2、0.4、0.6、0.8和1,并分别在真实数据集和合成数据集上进行测试,结果如表10和表11所列。其中,真实数据集上纵坐标差为0.6时性能最优;合成数据集上纵坐标差为0.4时为最优。由于真实数据集数量较少,认为纵坐标差比例取0.4时较优。

表10 真实数据集上不同纵坐标差比例的性能对比

Table 10 Performance comparison on real dataset with different vertical coordinate scales

比例	SER	NER	BER	ACC
0	7.519	8.905	8.212	91.790
0.2	5.548	6.571	6.059	93.797
0.4	5.236	6.201	5.718	94.115
0.6	5.028	5.954	5.491	94.327
0.8	5.618	6.544	6.081	93.825
1	5.853	6.932	6.392	93.486

表11 合成数据集上不同纵坐标差比例的性能对比

Table 11 Performance comparison on synthetic dataset with different vertical coordinate scales

比例	SER	NER	BER	ACC
0	6.559	5.701	6.130	93.724
0.2	6.207	5.391	5.799	94.107
0.4	5.917	5.144	5.531	94.416
0.6	6.083	5.288	5.686	94.237
0.8	6.213	5.400	5.806	94.098
1	6.544	5.727	6.135	93.742

结束语 本文利用目标检测框间的结构相关性及阴影所带来干扰的长宽比特性,基于3个约束合并图像中多个结构特征相关性强的相邻检测框,来提升对此类目标物体检测的准确率。实验结果表明,本文方法在小样本公开数据集、阴影公开数据集和合成数据集上的性能均有提升,并通过实例验证了本文方法的可用性。

由于本文直接根据阴影的结构性特点考虑检测框的合并,当需要合并的目标检测框数量较大时,会出现错误合并。此外,在模型测试的实验中,真实数据集数量少且多为合成数据集。因此,未来可考虑引入RNN、语序特征的连续性、判断是否合并检测框,进一步提高模型的准确率,并尝试基于更多的真实数据集测试模型的有效性。

参考文献

- [1] BAKO S, DARABI S, SHECHTMAN E, et al. Removing shadows from images of documents [C]// The 13th Asian Conference on Computer Vision. 2016:173-183.
- [2] LIN Y H, CHEN W C, CHUANG Y. BEDSR-Net: A Deep Shadow Removal Network from a Single Document Image [C]// IEEE Conference on Computer Vision and Pattern Recognition. 2020:12902-12911.
- [3] EVERINGHAM M, VAN G, WILLIAMS C, et al. The PASCAL Visual Object Classes Challenge 2007 (VOC2007) Results [EB/OL]. <http://www.pascal-network.org/challenges/VOC/voc2007/workshop/index.html>. 2007.

- [4] WANG Y P, ZHAO X, LI Y, et al. Densely cascaded shadow detection network via deeply supervised parallel fusion [C] // International Joint Conference on Artificial Intelligence. 2018: 1007-1013.
- [5] VICENTE T, HOU L, YU C, et al. Large-scale training of shadow detectors with noisily-annotated shadow examples [C] // European Conference on Computer Vision. 2016: 816-832.
- [6] HU X W, FU C W, ZHU L, et al. Direction-Aware Spatial Context Features for Shadow Detection [C] // IEEE Conference on Computer Vision and Pattern Recognition. 2018: 7454-7462.
- [7] WANG X, HUANG T, GONZALEZ J, et al. Frustratingly Simple Few-Shot Object Detection [C] // International Conference on Machine Learning. 2020: 9919-9928.
- [8] YANG Q, TAN K, AHUJA N. Shadow removal using bilateral filtering [J]. IEEE Transactions on Image Processing, 2012, 21(10): 4361-4368.
- [9] ZHANG L, ANDY M, TAN C. Removing shading distortions in camera-based document images using inpainting and surface fitting with radial basis functions [C] // International Conference on Document Analysis and Recognition. 2007: 984-988.
- [10] TSOI Y, BROWN M. Geometric and Shading Correction for Images of Printed Materials a Unified Approach Using Boundary [C] // IEEE Conference on Computer Vision and Pattern Recognition, 2004: 240-246.
- [11] JUNG S, ABUL H, KIM C. Water-filling: An efficient algorithm for digitized document shadow removal [C] // Asian Conference on Computer Vision. 2018: 398-414.
- [12] KLIGLER N, KATZ S, AYLLET T. Document enhancement using visibility detection [C] // IEEE Conference on Computer Vision and Pattern Recognition. 2018: 2374-2382.
- [13] OLIVEIRA D, LINS R, GABRIEL F. Shading removal of illustrated documents [C] // International Conference on Image Analysis and Recognition. 2013: 308-317.
- [14] FINLAYSON D, HORDLEY S D, LU C, et al. On the removal of shadows from images [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2006, 28(1): 59-68.
- [15] TIAN J D, QI X J, QU L Q, et al. New spectrum ratio properties and features for shadow detection [J]. Pattern Recognition, 2016, 51(3): 85-96.
- [16] GUO R Q, DAI Q Y, HOIEM D. Single-image shadow detection and removal using paired regions [C] // IEEE Conference on Computer Vision and Pattern Recognition. 2011: 2033-2040.
- [17] HUANG X, HUA G, TUMBLIN J, et al. What characterizes a shadow boundary under the sun and sky? [C] // IEEE International Conference on Computer Vision. 2011: 898-905.
- [18] VICENTE Y, TOMAS F, HOAI M, et al. Leave-one-out kernel optimization for shadow detection [C] // IEEE International Conference on Computer Vision. 2015: 3388-3396.
- [19] ZHU J J, SAMUEL K, MASOOD S, et al. Learning to recognize shadows in monochromatic natural images [C] // IEEE Conference on Computer Vision and Pattern Recognition. 2010: 223-230.
- [20] KHAN S, BENNAMOUN M, SOHEL F, et al. Automatic feature learning for robust shadow detection [C] // IEEE Conference on Computer Vision and Pattern Recognition. 2014: 1939-1946.
- [21] SHEN L, CHUA T, LEMAN K. Shadow optimization from structured deep edge detection [C] // IEEE Conference on Computer Vision and Pattern Recognition. 2015: 2067-2074.
- [22] VICENTE T, HOU L, YU C, et al. Large-scale training of shadow detectors with noisily-annotated shadow examples [C] // European Conference on Computer Vision. 2016: 816-832.
- [23] SHI H, ZHANG L. Image Shadow Removal Algorithm Based on Generative Adversarial Network [J]. Computer Science, 2021, 48(6): 145-152.
- [24] HU X W, FU C W, ZHU L, et al. Direction-aware spatial context features for shadow detection [C] // IEEE Conference on Computer Vision and Pattern Recognition. 2018: 7454-7462.
- [25] WANG J F, LI X, YANG J. Stacked conditional generative adversarial networks for jointly learning shadow detection and shadow removal [C] // IEEE Conference on Computer Vision and Pattern Recognition. 2018: 1788-1797.
- [26] ZHU Y, HUANG J, FU X, et al. Bijective mapping network for shadow removal [C] // IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022: 5627-5636.
- [27] ZHU Y, XIAO Z, FANG Y, et al. Efficient model-driven network for shadow removal [C] // AAAI Conference on Artificial Intelligence. 2022: 3635-3643.
- [28] LUO W, XIE X, DENG K, et al. Learning Shadow Removal from Unpaired Samples via Reciprocal Learning [J]. IEEE Transactions on Image Processing, 2023, 32: 3455-3464.
- [29] REN S, HE K, ROSS B, et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks [C] // Neural Information Processing Systems. 2015: 91-99.
- [30] HUANG W L, LIN Z, YANG J C, et al. Text localization in natural images using stroke feature transform and text covariance descriptors [C] // IEEE International Conference on Computer Vision, 2013: 1241-1248.



WANG Jiahui, born in 1996, Ph.D. His main research interests include machine learning and knowledge engineering.



DUAN Liang, born in 1986, Ph.D, associate professor, master supervisor. His main research interests include unsupervised machine learning and knowledge engineering.