

中图法分类号: TP18; TP391 文献标识码: A 文章编号: 1006-8961(2026)01-0099-21

论文引用格式: Xue Z Y, Liu Q T, Zhou K N, Wang W and Jiang X H. 2026. Synthetic dataset based on physical properties and forgery detection method for shadow-consistency analysis. Journal of Image and Graphics, 31(1):0099-0119(薛子育, 刘庆同, 周康能, 王伟, 姜秀华. 2026. 面向光影一致性分析的物理属性合成数据集及伪造检测方法. 中国图象图形学报, 31(1):0099-0119)[DOI:10.11834/jig.240705]

面向光影一致性分析的物理属性合成数据集及伪造检测方法

薛子育^{1*}, 刘庆同¹, 周康能², 王伟³, 姜秀华⁴

1. 国家广播电视总局广播电视科学研究院, 北京 100866; 2. 南开大学计算机学院, 天津 300350;
3. 中国科学院自动化研究所, 北京 100190; 4. 中国传媒大学信息与通信工程学院, 北京 100024

摘要: **目的** 图像和视频合成技术在媒体后期处理领域广泛应用, 随着技术门槛的降低, 大量合成素材被发布并迅速传播。然而, 部分合成内容可能含有误导性信息, 威胁视听内容的真实性和安全性。传统合成检测方法主要依赖合成痕迹或画面异常检测, 但随着合成技术的不断进步, 现有方法在检测精度和适应性方面仍存在优化空间, 需要改进以应对日益复杂的合成内容检测需求。**方法** 本文提出一种融合物理与深度学习的合成图像检测方法, 创新性地结合光照和阴影一致性分析。通过特征提取与融合网络, 实现光照图与光照强度的一致性分析, 判断物体采集环境; 利用交比估计检测光照方向一致性, 有效提升了检测精度和适应性。同时构建了具有物理属性的数据集, 为合成图像检测提供数据支持。**结果** 在NIST 16(National Institute of Standards and Technology Database 16)、Coverage和CASIA(Chinese Academy of Sciences Institute of Automation Database)数据集上的实验表明, 本文方法在AUC(area under the curve)指标上分别达到94.2%、93.6%和90.3%, F1分数分别达到80.2%、79.3%和58.1%, 优于对比方法。在噪声攻击实验中, 本文方法对尺寸变化、高斯模糊、高斯噪声和JPEG(Joint Photographic Experts Group)压缩表现出更强的适应性, 平均AUC为84.03%。此外, 本文提出的数据集在训练过程中表现出高可用性, 使用该数据集训练的模型AUC平均提升18.1%。**结论** 本文方法在准确性和鲁棒性方面均优于对比方法, 构建的数据集能够有效支持合成图像检测模型的训练、验证和测试, 为该领域的研究提供了重要参考。数据集下载链接: <https://doi.org/10.57760/sciencedb.j00240.00069>。**关键词:** 合成图像检测; 光照检测; 阴影检测; 检测数据集; 人工智能安全

Synthetic dataset based on physical properties and forgery detection method for shadow-consistency analysis

Xue Ziyu^{1*}, Liu Qingtong¹, Zhou Kangneng², Wang Wei³, Jiang Xiuhua⁴

1. Academy of Broadcasting Science, National Radio and Television Administration, Beijing 100866, China;
2. College of Computer Science, Naikai University, Tianjin 300350, China; 3. Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China; 4. School of Information and Communication Engineering, Communication University of China, Beijing 100024, China

收稿日期: 2024-11-21; 修回日期: 2025-04-03; 预印本日期: 2025-04-10

* 通信作者: 薛子育 xueziyu@abs.ac.cn

基金项目: 中央财政专项项目(102130220490010009001); 国家重点研发计划资助(2021YFC3320100); 国家广播电视总局广播电视科学研究院基本科研业务费项目(JBKY20240020)

Supported by: National Fiscal Expenditure Program of China (102130220490010009001); National Key R&D Program of China (2021YFC3320100); Fundamental Research Funds for the Academy of Broadcasting Science, National Radio and Television Administration (JBKY20240020)

Abstract: Objective Synthesis and forgery of media content such as images and videos have become common methods in media post-production. As the technical barriers gradually decrease, the Internet is disseminated with a large amount of synthetic materials. This situation introduces great challenges to the integrity and security of digital media, which affects the authenticity and safety of audiovisual content. In the realm of synthetic media detection, research has greatly increased in recent years. Most commonly used detection methods focus on physical attributes or directly utilize deep neural networks for detection. Image forgery detection methods based on physical attributes can be broadly divided into those that utilize physical features and those that employ deep neural networks. Methods relying on physical features identify visual inconsistencies in images. These methods leverage the fact that synthetic images are often composed of multiple stitched images, which can result in irregularities in lighting, shadows, depth of field, and object distance. Inconsistency detection involves examining inherent physical properties such as lighting and shadows to assess their coherence. Lighting detection methods primarily analyze factors such as light source position, chromatic deviation, object reflectivity, light source color, and local light sources. Shadow detection methods focus on properties such as planar homology, shadow constraints, vanishing point positions for light and shadow, and consistency in geometric relationships. However, due to the diverse and complex environments in which images are acquired, the high precision required by these physical detection methods causes difficulty in relying solely on a single approach for forgery detection in real-world scenarios. In image forgery detection methods based on deep neural networks, some approaches use long short-term memory (LSTM), dual stream parallel network, pyramid architectures, and Transformers to conduct synthesis detection by identifying anomalies in local regions between image blocks while locating synthesized areas. While these methods have demonstrated certain levels of success in experimental settings, the relentless advancement in synthesis technology has resulted in synthetic images that are increasingly indistinguishable from authentic ones. As a result, detection methods relying solely on visual anomalies or signs of synthesis can no longer meet the growing demands for accuracy in detection. **Method** We present a novel approach called the lighting and shadow fusion network (LSFNet) to image forgery detection based on the consistency of lighting and shadows. It combines highly sensitive physical attribute detection with robust deep neural network techniques, which improves the capability to identify even the most meticulously crafted synthetic images. LSFNet comprises key modules, including illumination maps, object and shadow estimation, illumination analysis, and homogeneity analysis. Its goal is to integrate physical detection methods including illumination and shadow with deep neural network techniques for effective detection of well-crafted synthetic images. The method begins by creating an illumination map that labels the object and shadow areas, which results in a boundary illumination map formed by overlapping the illumination map with the object area. This study employs a feature extraction network to derive features from the boundary illumination map and the original image. As a result, a feature fusion network for consistency analysis of the illumination map and light intensity is established to determine the shooting environment of objects within the image. It estimates distances for objects and shadows using cross-ratio for homogeneous vertex estimation to ensure consistent detection of lighting direction in images. Finally, this study utilizes result fusion methods to comprehensively evaluate lighting analysis results and homogeneity analysis results to predict outcomes for synthetic images. Recognizing that existing datasets for synthetic image detection often lack detailed information regarding physical attributes, we introduce a new dataset designed to assist in detecting synthetic images by incorporating essential characteristics such as lighting conditions and object distances crucial for accurately identifying forgery in the complex digital media landscape at present. **Result** Experiments conducted on the NIST 16 (National Institute of Standards and Technology Database 16), Coverage, and CASIA (Chinese Academy of Sciences Institute of Automation Database) datasets demonstrate that our method achieves AUC (area under the curve) scores of 94.2%, 93.6%, and 90.3%, respectively, and F1 scores of 80.2%, 79.3%, and 58.1%, outperforming the compared methods. In noise attack experiments, our method exhibits stronger adaptability to size variation, Gaussian blur, Gaussian noise, and JPEG (Joint Photographic Experts Group) compression, with an average AUC of 84.03%. **Conclusion** The contributions of this study are summarized as follows: 1) we propose a versatile image forgery detection framework for high-fidelity synthetic images, which is called LSFNet. This framework integrates deep neural networks with physical detection methods and combines boundary lighting map detection and physical attribute consistency analysis. Experimental results show that our method outperforms mainstream techniques in accuracy and demonstrates robustness, which makes it suitable for various synthetic image detection

tasks; 2) We present a shadow extraction and consistency analysis method tailored for image forgery detection. This approach, which utilizes Mask R-CNN as the backbone network for shadow prediction, models the relationship between shadow regions and real object areas. By assessing consistency through intersection over union, it significantly enhances the accuracy of image forgery detection; 3) We introduce a synthetic image detection dataset featuring physical attributes such as light source position, color temperature, camera position, object distance, depth of field, and complex backgrounds. The dataset is called the synthetic image physical properties detection dataset. This dataset comprises 22 030 images——15 012 collected images and 7 191 synthetically generated ones——all meticulously crafted by professionals to ensure high realism. The dataset maintains a training-validation-testing split of 7:2:1. It includes various synthesis methods such as stitching, copy-pasting, and deletion to effectively support training testing and optimization upgrades for image forgery detection models. The dataset is linked at: <https://doi.org/10.57760/sciencedb.j00240.00069>.

Key words: synthetic image detection; illumination detection; shadow detection; detection dataset; artificial intelligence security

0 引言

近年来,数字图像合成技术(Crowson等,2022; Lee等,2020; Liu等,2020; Shi等,2022)逐步趋于成熟,通过替换(Lee等,2020)、删除(Lahiri等,2020)、属性改变(Nillius和Eklundh,2001)等操作即可完成图像合成。该项技术已经在影视后期、节目制作、娱乐产业等多个领域有所应用。然而,部分专业人员利用合成技术制作具有迷惑性的虚假媒体内容并发布至互联网,严重威胁了媒体内容的真实性和安全性。基于此,合成内容检测以及合成区域定位成为重要的研究方向。图像合成大体可以分为拼接、复制粘贴和删除3类。拼接是将一幅图像中的物体进行分割,并将其粘贴在其他图像中。复制粘贴是将图像中的数字对象进行分割,并将其粘贴到原图像中的其他位置。删除是将一幅图像中的某个物体进行删除,并将删除的背景进行还原,如图1所示。

图像合成检测方法主要包括基于物理特征的检测方法和基于深度神经网络的检测方法。基于物理特征的检测方法是对图像中的视觉不一致性或明显的合成痕迹进行检测。由于合成图像的内容大多由几幅图像拼接而成,因此可能出现异常的光照和阴影等,不一致性检测即是对图像中的光照和阴影等物理属性进行检测并开展一致性判断。光照检测方法主要是对光源位置(Liu等,2020; Nillius和Eklundh,2001)、色度偏差(Johnson和Farid,2006)、物体反射率(Johnson和Farid,2007)、光源颜色(Gholap和Bora,2008)、局部光源属性(Riess和Angelopoulou,2010)等进行检测和分析,或是利用图

卷积网络(Pan等,2024)对光照情况进行判断。阴影检测方法是利用平面同源性(Zhang等,2009)、阴影约束(Kee等,2013)、光影消失点位置(Merlin等,2019)以及几何关系一致性(Nasiri和Behrad,2019)等进行合成检测。除此以外,直接对合成过程中可能带来的图像痕迹进行检测也是常用方法,检测属性包括Daubechies小波特征(Zhang等,2016)、边缘强化(Salloum等,2018)、自相关计算(Wu等,2018)、重采样特征(Bunk等,2017; Bappy等,2017)以及JPEG(Joint Photographic Experts Group)压缩特征(Park等,2018)等。然而在实际应用中,图像采集环境多样复杂,物理检测方法要求的精度高,仅依靠单一的检测手段难以应对真实环境下的合成内容进行检测。

在基于深度神经网络的合成图像检测方法中,部分学者采用长短时记忆网络(long short-term memory, LSTM)(Bappy等,2017,2019)、两流并行网络(Zhou等,2018; Han等,2024)、金字塔架构(Hu等,2020)、特征提取器(Hu等,2020)以及Transformer(Atak和Yasar,2024; Hao等,2021)等进行合成检测,并通过在图像区块间判断局部区域异常定位合成区域。Liu等人(2024)分别提出双路径检测框架和多尺度编解码器架构,通过自上而下提取局部与全局特征,结合注意力机制与对比学习实现多尺度合成区域定位。Niloy等人(2023)和Zhou等人(2023)分别利用特征分布差异与非互斥对比学习,通过对比损失学习和动态切换边缘区域角色,解决了合成区域与原始区域的分离问题,并实现了无需预训练的高精度篡改定位。Bai(2024)和Hao等人(2024)分别提出双流结构化编码器和不确定引导策

略,通过融合模块整合原始图像与噪声特征,并结合上下文信息跨层融合,增强细节捕捉能力。Guo 等人(2023)提出的 HiFi-Net(hierarchical fine-grained network)框架,通过层次化特征融合与跨域检测能力,进一步提升了物理属性不一致性检测的普适性。

以上方法采用深度神经网络实现了合成图像的痕迹检测,取得了一定的实验效果。然而,随着合成技术的不断提升,现有方法在实际应用中仍存在一定的局限性。值得注意的是,虽然合成图像的效果不断增强,但是合成过程对于光照、阴影和景深等语义信息的考虑仍有所欠缺。基于此,本文通过将变化感知性强的物理属性检测方法与鲁棒性强的深度神经网络检测方法相结合,提出基于光照和阴影一致性的合成图像检测方法,可以更有效地对制作精良的合成图像进行检测。同时,为了解决现阶段合成图像检测数据集(Dong 等, 2013; NIST, 2017; Novozámský 等, 2021; Wen 等, 2016)均不具备详细物理属性的问题,本文提出具有光照、物距和景深等详细物理属性标注的合成图像检测数据集。

本文的贡献点可归纳如下:1)面向高逼真度的合成图像,提出一个适用性更强的合成图像检测框架,命名为光照与阴影结合的检测网络(lightning and

shadow fusion network, LSFNet),该框架创新性地将深度神经网络与物理检测方法相结合,融合了边界光照图检测和物理属性一致性检测的优势。实验结果证明,本文提出的方法较主流方法的准确率更高,且具有一定的鲁棒性,可以适用于多类合成图像的检测任务。2)提出一种适用于合成图像检测的阴影提取方法和一致性分析方法,方法以 Mask R-CNN(convolutional neural network)为骨干网络进行阴影预测,对阴影区域与实物区域进行关联建模。基于不变性的几何约束,对实物与阴影之间的一致性进行定量比较,有效判断阴影与实物的几何关系,从而显著提升了合成图像检测的准确率。3)提出一个包含光源位置、光源色温、机位、物距、景深和复杂背景等物理属性的合成图像检测数据集,命名为合成图像物理属性检测(synthetic image physical properties detection, SIPD)数据集, SIPD 数据集包含 22 203 幅图像,其中,采集图像 15 012 幅,合成图像 7 191 幅,所有的合成图像均由专业人员手工制作,逼真度高。数据集的训练、验证和测试数据比例为 7:2:1,包含拼接、复制粘贴以及删除等多种合成方法,能够有效支撑合成图像检测模型的训练、测试与优化升级。

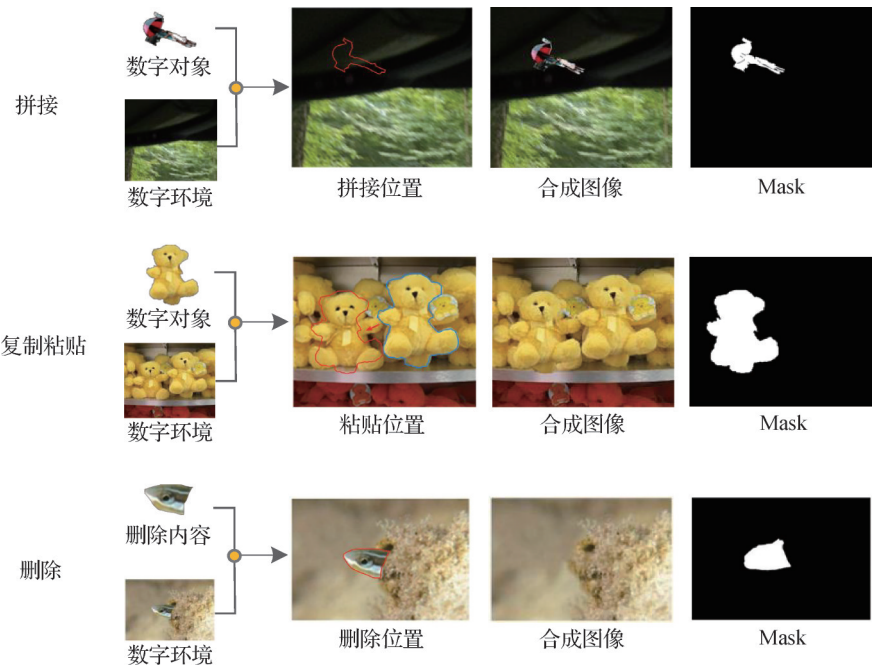


图1 图像合成示意图(拼接、复制粘贴和删除)

Fig. 1 Schematic diagram of image synthesis (stitching, copy-paste, and deletion)

1 相关工作

1.1 合成图像检测方法

图像合成过程中,色差、压缩等图像固有属性可能发生改变,传统物理属性检测方法主要针对色差、相机响应函数、双相干性以及PEG量化等合成痕迹展开。随着合成技术的优化,合成痕迹的隐蔽性显著增强,仅依赖痕迹检测已难以满足需求。值得注意的是,合成过程往往难以完全保留物理属性的一致性,尤其是合成物体的光照、阴影、景深等属性易与背景环境失配,因此基于物理属性不一致性的检测方法更具普适性。

在光照一致性检测方面,主要用于判断图像中的光照异常。早期的检测方法包括对光源位置(Liu等, 2020; Nilius 和 Eklundh, 2001)、色度偏差(Johnson 和 Farid, 2006)、物体反射率(Johnson 和 Farid, 2007)、光源颜色(Gholap 和 Bora, 2008)、局部光源(Riess 和 Angelopoulou, 2010)以及光照图(Riess 和 Angelopoulou, 2010)进行光源分析。Riess 和 Angelopoulou(2010)提出基于颜色分块的局部光源估计方法,通过计算图像块间的像素距离判断合成区域。Zhu 等人(2021)提出双流拼接图像定位网络,光照流提取灰度边缘与逆强度色度特征,图像流捕获全局合成痕迹。Pan 等人(2024)提出Cas-HGCN(cascade hierarchical graph convolutional network),通过融入光照一致性检测的深度学习方法,利用图结构建模进行光照一致性检测。此外,学者还探索了统计模型(Sekhar 和 Shankar, 2021)、图像去雾(Chen 等, 2022)、尺度变化度量(Hadwiger 和 Riess, 2022)、噪声对比(Bai, 2024)以及边缘增强(Hao 等, 2024)等策略,进一步提升了检测精度。

在阴影检测方面,通过建模物理约束识别合成图像,早期方法包括平面同源性(Zhang 等, 2009)、交叉约束(Kee 等, 2013)、几何约束(Crowson 等, 2022)、光影消失点估计(Merlin 等, 2019)以及阴影几何估计(Nasiri 和 Behrad, 2019)。然而,此类方法依赖特定场景假设,例如 Riess 和 Angelopoulou(2010)提出的方法需要两个以上的物体及其阴影才能建立几何关系,因此无法对单个物体的图像进行合成检测。针对这一局限性,Han 等(2024)提出的HDF-Net创新性地融合RGB与SRM(steganalysis

rich model)双流特征,结合多尺度视觉变换器提取全局噪声模式,解决了同源合成中局部噪声一致性失效的难题,在NIST 16数据集上实现了对模糊后处理操作的强鲁棒性检测。

现阶段,基于深度神经网络的检测方法也取得了一定效果(叶国升 等, 2023),Palmer 等人(2022)通过3D渲染模拟光照条件,对比不同角度反射光差异实现合成检测。Guo 等人(2023)将层次化特征融合与跨域检测能力,补充至物理属性不一致性检测的普适性分析中。Zhou 等人(2023)提出非互斥对比学习框架,通过动态切换边缘区域的正负样本角色并引入枢纽一致性损失,解决了图像合成定位任务中数据不足和边缘区域非互斥性的问题。Qu 等人(2024)提出结合多模块自适应感知的APSC-Net模型,显著提升了复杂场景下的定位合成区域的效果。

尽管以上方法在光照、阴影等单一属性检测中取得了显著进展,但仍存在一定的局限性。现有方法多针对特定场景或单一属性进行优化,对复杂多变的合成场景适用性弱,尤其在跨域、多属性联合检测方面仍有提升空间。通过综合研判光照、阴影、噪声和边缘等多种属性,构建多属性联合检测框架,可以更好地适应复杂场景的需求,进一步提升检测的鲁棒性和普适性。

1.2 物体和阴影分割方法

物体与阴影分割是计算机视觉的重要研究方向,主要包括基于传统机器学习与基于深度学习两类方法。

传统方法依赖纹理、颜色和边缘等人工特征,并利用分类算法对特征的阴影进行估计。例如对分割区域外观的光照预测(Guo 等, 2011)、对物体纹理和颜色特征的分类(Vicente 等, 2018)以及对颜色和边缘特征的阴影估计(Lalonde 等, 2010; Huang 等, 2011)等。尽管部分研究尝试多特征融合方法(Huang 等, 2011; Lalonde 等, 2010; Vicente 等, 2018),但是人工提取特征具有一定的局限性导致其适用场景受限。

在早期的利用深度学习算法开展阴影分割的方法中,阴影衰减网络(Le 等, 2018)、干扰感知阴影检测网络(Zheng 等, 2019)以及循环生成式对抗网络(Ding 等, 2019)均有所使用。在近期的研究中,Hu 等人(2021)提出一个带有细节增强模块的快速阴影检测网络,通过捕获阴影细节,实现阴影区域的还

原。Wang 等人(2022)通过端到端的方式进行阴影检测,提出双向关系的学习模块和可变的感兴趣区域识别模块,用于确定阴影之间的关系。Xing 等人(2024)提出一种用于跟踪视频帧中物体阴影的方法,方法以周期序列的方式进行阴影估计。尽管深度模型精度较高,但在复杂光照与动态场景下的鲁棒性仍有提升空间。

1.3 合成图像检测数据集

现阶段,主流的合成图像检测数据集包括 NIST 16(National Institute of Standards and Technology Database 16)、Columbia、CASIA (Chinese Academy of Sciences Institute of Automation Database)、Coverage 和 IMD 2020 (image manipulation dataset 2020) 等。NIST 16 数据集包含拼接、复制粘贴和删除 3 类合成图像。Columbia 数据集(Ng 和 Chang, 2004)专注于图像拼接方法,使用 Photoshop 作为合成工具,包含 1 845 幅图像,真实图像与合成图像的数量大致相等。CASIA 数据集(Dong 等, 2013)包含拼接和复制粘贴两类合成方法,提供原图像和掩码图像。此外, CASIA 数据集还加入了过滤和模糊等后处理操作,使合成效果更加逼真。2016 年,美国伊利诺伊大学香槟分校的研究人员提出 Coverage 数据集(Wen 等, 2016),包含直接复制、形态变化等 6 类复制粘贴模式,并提出相似度量测评方法。2020 年,捷克的研究人员提出 IMD 2020 数据集(Novozámský 等, 2021),该数据集包含 2 000 幅网络合成图像,研究人员下载

了这些图像并获取了原图像,制作了掩码图。与其他数据集相比,IMD 2020 更贴近现实生活。

总体而言,基于物理属性的检测数据集对采集环境有一定要求,且属性设置较为复杂。现有的数据集通常不支持物理属性的研究工作。然而,通过判断物体与物理环境的匹配程度来检测内容合成,已成为重要的合成图像检测途径。因此,开展基于物理属性的检测数据集研究是必要的。

2 方 法

2.1 模型架构

根据前期论文的研究成果,本文将光照和阴影作为重点研究对象,尝试将物理属性与深度神经网络相结合,提出一种新的合成图像检测架构,命名为光照与阴影结合的检测网络(LSFNet),如图 2 所示。LSFNet 主要包括光照图、目标和阴影估计、光照分析和同源分析等主要模块,旨在将光照和阴影等物理检测方法与深度神经网络检测方法相结合,能够对制作精良的合成图像进行有效检测。本文方法首先建立光照图,并对物体区域与阴影区域进行标注,通过对光照图与物体区域进行重叠建立边界光照图。本文利用特征提取网络对边界光照图进行特征提取,建立特征融合网络实现对光照图和光照强度的一致性分析,判断图像中物体所处的拍摄环境。本文对物体和阴影进行距离分析,利用交

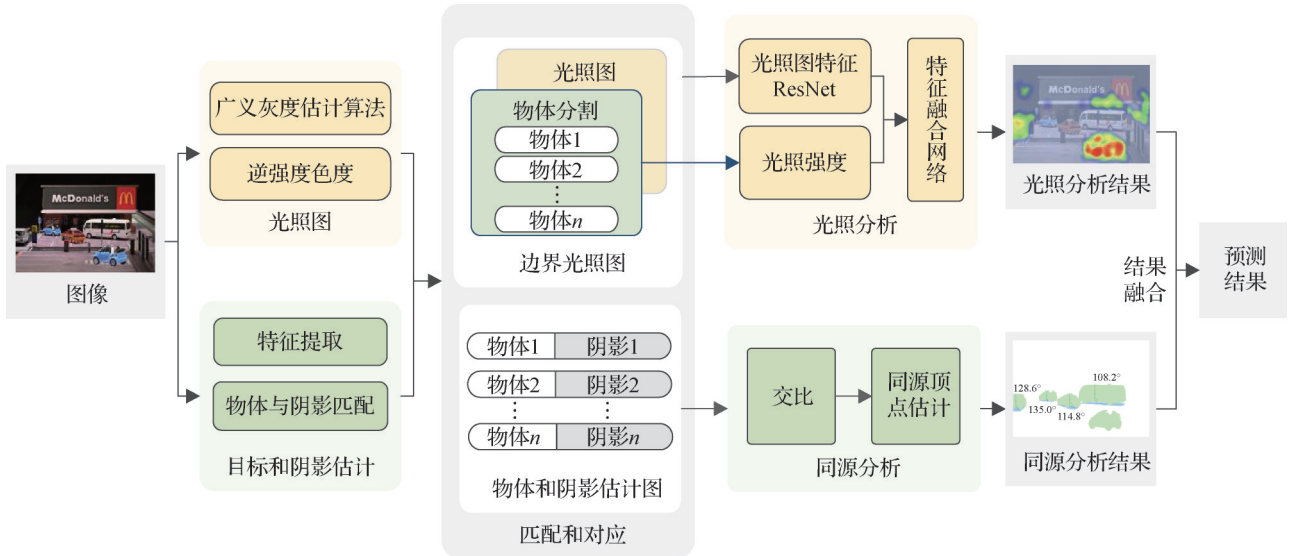


图2 光照与阴影结合的检测网络(LSFNet)架构

Fig. 2 Architecture of the lighting and shadow fusion detection network (LSFNet)

比进行同源顶点估计,实现图像光照方向的一致性检测。最后利用结果融合方法对光照分析结果和同源分析结果进行综合判断,实现图像合成结果的预测。

2.2 特征提取

2.2.1 光照图

光照图用于开展图像的光照分析,检测图像中光照属性不一致的区域。在早期的研究中,Riess和Angelopoulou(2010)提出获取光照图的方法,通过将图像的相似颜色区域进行细分建立超像素块。利用超像素块内的局部像素估计光源颜色,并对相似颜色的超像素块进行重新上色,获得光照图。本文参照Pomari等人(2018)提出的思路,首先使用广义灰度估计算法(van de Weijer等,2007),根据像素点对光源 e 的灰度进行估计,具体为

$$\lambda e^{n,p,\sigma} = \left(\int \left| \frac{\alpha^n \Gamma^\sigma(x)}{\alpha x^n} \right|^p dx \right)^{\frac{1}{p}} \quad (1)$$

式中, x 表示像素坐标, λ 是比例因子, $|\cdot|$ 是绝对值, α 是微分算子, $\Gamma^\sigma(x)$ 为 x 位置观测到的强度, σ 是平滑高斯核, p 是闵可夫斯基范数, n 为导数的阶。

同时,本文使用逆强度色度空间(Tan等,2004)对方法进行扩展,在 x 处的强度 $\Gamma_c(x)$ 和色度 $\chi_c(x)$ 将按照下式的方法进行转换,具体为

$$\chi_c(x) = m(x) \frac{1}{\sum_{i \in \{R,G,B\}} \Gamma_i(x)} + Y_c \quad (2)$$

式中, Y_c 表示通道 c 中光源的色度, $m(x)$ 用于捕捉光源位置、照明方向和相机位置等的几何关系,以此获得光照图。

2.2.2 目标和阴影估计

本文利用特征金字塔网络(feature pyramid network, FPN)(Lin等,2017)获取物体和阴影实例。方

法借助区域提议网络(region proposal network, RPN)(Ren等,2017)对物体和阴影区域进行框取,利用Mask R-CNN(He等,2017)对每个候选框进行特征提取,特征提取器采用ResNeXt-101-FPN(Xie等,2017)。本文将边界框预测模块和掩码预测模块进行最小化实现结果预测(Wang等,2020),并对训练中的损失函数进行调整。边界框预测模块中的参数调整过程,需要计算检测到的阴影区域与真实的阴影区域之间的最小化损失,如图3所示。在完成阴影和物体区域的提取后,本文参照Wang等人(2020)提出的方法对阴影和对应的物体进行关联。首先计算各个阴影和物体区域的中心点,并获取阴影中心 $(x_{mid}^{sh}, y_{mid}^{sh})$ 和物体中心 $(x_{mid}^{ob}, y_{mid}^{ob})$ 之间的最短距离。如果该距离小于阈值,则判定为关联区域。本文将阈值设定为阴影的高度。在完成目标和阴影的匹配后,将关联区域中物体和阴影表示为线段。首先将物体包围框的最高点标记为 $(x_{high}^{ob}, y_{high}^{ob})$,同时将线段 $(x_{high}^{ob}, y_{high}^{ob})(x_{mid}^{ob}, y_{mid}^{ob})$ 所在的直线 $line^{ob}$ 进行标记,如图4(a)所示。同样,将阴影边缘点标记为 $(x_{edge}^{sh}, y_{edge}^{sh})$ 。线段 $(x_{mid}^{sh}, y_{mid}^{sh})(x_{edge}^{sh}, y_{edge}^{sh})$ 所在的直线记为 $line^{sh}$ 。将直线 $line^{ob}$ 和 $line^{sh}$ 的交点记为 o ,如图4(b)所示。建立线段 l^{ob} 和 l^{sh} 分别对物体和阴影进行标记,如图4(c)所示。由于图像的合成区域大多为物体,因此本文将目标分割结果与光照图进行结合,以辅助分类器进行分类。在获取到光照图和物体分割图以后,本文将光照图与物体边界进行重叠组合,形成边界光照图。

2.3 一致性分析

2.3.1 光照分析

本文分别利用ResNet-50(He等,2016)对边界光照图 $IMB \in \mathbb{I}^{3 \times H \times W}$ 和目标分割图 $OS \in \mathbb{I}^{3 \times H \times W}$ 进行特征提取,并利用特征融合网络对提取结果进行

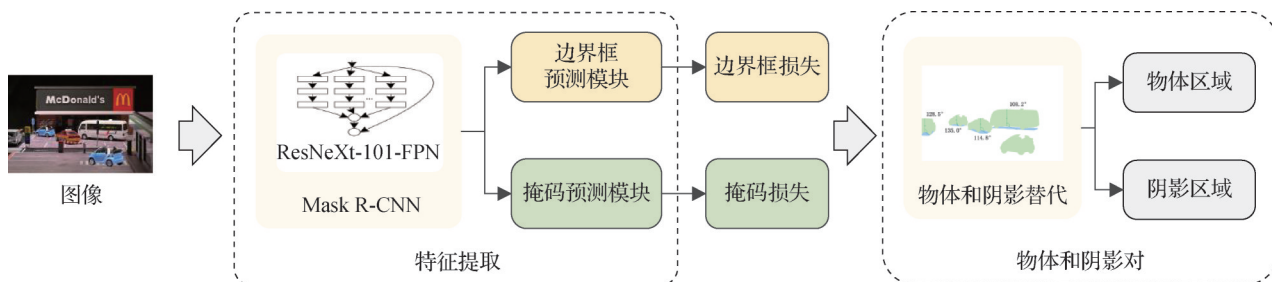


图3 物体和阴影分割过程

Fig. 3 Process of object and shadow segmentation

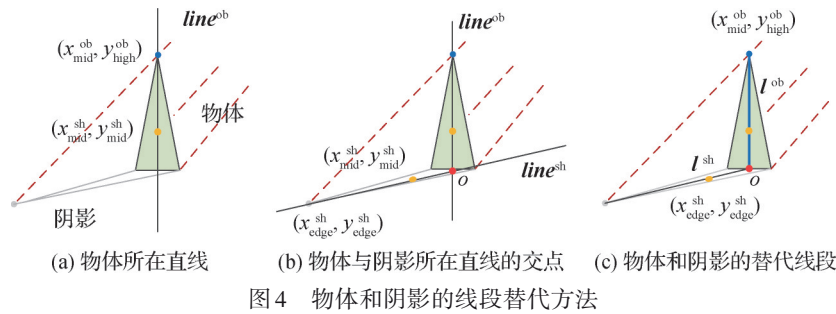


Fig. 4 Line segment substitution method for objects and shadows ((a) line on which the object lies; (b) intersection point of the lines on which the object and its shadow lie; (c) substitute line segment for the object and shadow)

分类。目标分割图 OS 的特征主要用于开展光照强度分析,特征融合网络将对光照图特征和光照强度特征进行融合分析,获得分类结果,特征融合过程如图5所示。

1)特征提取。在分类模块中主要对边界光照图和光照强度进行特征提取,两个特征提取器均采用 ResNet-50 作为骨干网络,选取 ResNet-50 第4阶段的输出作为最终输出(Chen等,2021a),以此获取到边界光照图的特征 $F_{IMB} \in \mathbb{R}^{C \times H_{IM} \times W_{IM}}$ 和光照强度的特征 $F_{OS} \in \mathbb{R}^{C \times H_{OS} \times W_{OS}}$ 。其中, $H_{IM}, H_{OS} = H/8, W_{IM}, W_{OS} = W/8, C = 1024$ 。

2)特征融合。在完成特征提取后,本文采用元素求和的方法将相同特征维度的特征向量进行融合(Song等,2018),按照下式进行池化操作,具体为

$$z = \text{pooling}(F_{IMB} + F_{OS}) \quad (3)$$

式中, z 为融合特征, pooling 为全局特征函数,将输出的预测结果记为 R_c 。

3)合成区域估计。借助于特征融合网络的结果

输出,本文进一步开展合成区域的估计。当结果预测为假时,将边界光照图中标记为异常的像素点及其轮廓标记为合成区域,以获取到合成图像的掩码图。

2.3.2 同源分析

本文建立的一致性检测模块是利用物理方法对图像中的物体及阴影关系进行一致性分析,用以辅助深度学习开展合成检测,重点考虑单个光源情景下的光源一致性检测。本文首先建立共面点成像关系图,用于检测图像是否满足点光源要求,如图6所示。

共面点成像是一个平面射影变换 H ,包括了由不动点组成的直线,称为轴 l ,以及一个不在轴上的光源,称为同调顶点。根据 Riess 和 Angelopoulou (2010)提出的方法, H 与顶点光源、轴 l 的关系为

$$H = I + (\mu - 1) \frac{v l^T}{v^T l} \quad (4)$$

式中, μ 是交叉比, v 是顶点光源,轴 l 是两平面的交线。按照 Riess 和 Angelopoulou (2010)提出的方法,

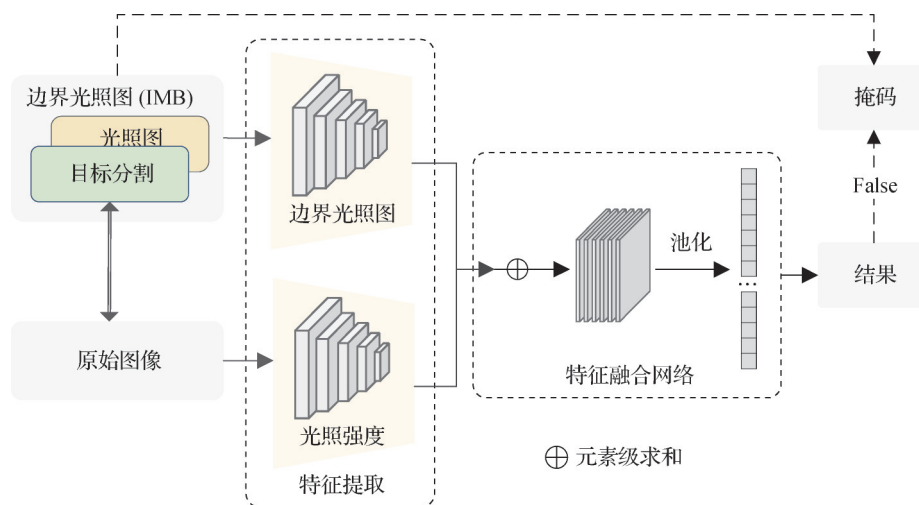


Fig. 5 Framework of feature fusion method

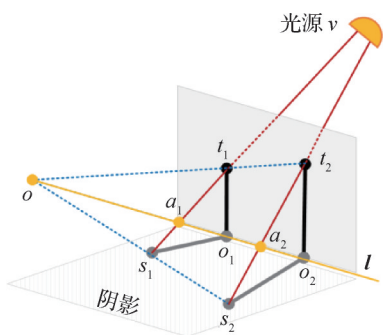


图6 共面点成像关系图

Fig. 6 Imaging relationship diagram of coplanar points

假设在二维平面上, t_1s_1 和 t_2s_2 与轴 l 的交点分别为 a_1 和 a_2 。根据共面点成像的性质, t_1s_1 与 t_2s_2 在轴 l 上相交于点 o 。需要注意的是, 在3D真实世界中, t_2s_2 不能与轴 l 相交, 本文按照 Riess 和 Angelopoulou (2010) 提出的思路, 只在理想情况下进行讨论。

根据交比的几何属性, 在具有光源顶点的情况下, t_i, s_i 以及交点 a_i 具有不变性, 即

$$\{v, t_1; s_1, a_1\} = \{v, t_2; s_2, a_2\} \quad (5)$$

式中, $\{v, t_1; s_1, a_1\}$ 计算方法为

$$\{v, t_1; s_1, a_1\} = \frac{vs_1 \cdot t_1 a_1}{va_1 \cdot t_1 s_1} \quad (6)$$

同时, 式(5)可以连续拓展到整个黎曼球面, 即复平面与无穷远点。 t_2t_1, s_2s_1 与 a_2a_1 之间具备以下关系, 具体为

$$((t_2 \times t_1) \times (s_2 \times s_1)) \cdot (a_2 \times a_1) = 0 \quad (7)$$

光源顶点具有以下属性

$$v = (t_2 \times s_2) \times (t_1 \times s_1) \quad (8)$$

当图像中的物体和阴影不能满足式(5)的约束时, 则判断为合成, 检测结果记为 R_{DE} 。

2.4 结果判断

结果融合模块将综合考虑分类结果 R_C 及一致性检测结果 R_{DE} , 形成 R , 具体为

$$R = \alpha R_C + \beta R_{DE} \quad (9)$$

式中, α 和 β 是结果融合参数, 参数值大小由消融实验确定, 具体见 4.2 节。当 R_{DE} 输出异常时, R 的最终值由 αR_C 决定。

3 合成图像物理属性检测数据集

3.1 基本信息与图像采集

受限于合成图像检测数据集对采集环境有一定

要求, 并且物理属性设置较为复杂, 现阶段的数据集还不能支持对物理属性的研究。为解决此类问题, 本文设计、采集并制作了合成图像物理属性检测 (SIPD) 数据集。SIPD 数据集的主体为微景观图像, 数据总量为 22 203 幅, 其中, 采集图像 15 012 幅, 合成图像 7 191 幅, 图像分辨率为 $8\,256 \times 5\,504$ 像素、 $3\,656 \times 2\,664$ 像素和 $2\,560 \times 1\,440$ 像素。原始图像由相机拍摄, 合成图像由操作人员利用 Photoshop 软件进行制作。拍摄相机的基本信息为: Nikon z7, Expeed6 影像处理器, 493 个对焦点, 可支持 4 575 万像素。镜头采用腾龙 90 mm f/2.8 的微距镜头。

SIPD 数据集设置了街景和人物两个拍摄组, 街景组采用自动对焦、人物组采用定焦拍摄。同时, SIPD 数据集中的采集数据与合成数据都包含了详细的物理属性, 例如光源位置、光源色温、机位、物距、景深以及复杂背景等。SIPD 数据集将物理属性均匀分配到了街景组和人物组当中, 研究人员可以根据具体需要选择对应的组进行模型训练和测试。

3.1.1 街景拍摄组

街景拍摄组考虑了采集背景、车型、颜色等物理属性的多样性。车型分为微型车、小型车、越野车、中型车和大型车, 如图 7(a) 所示。街景组中的车辆均设置了替换组, 每个替换组均包含两辆相同车型、不同颜色的车辆, 主要用于在拼接操作中实现相同或相似车型的替换, 保证图像整体风格的一致性和真实性。为了模拟真实车辆的颜色, SIPD 数据集依据艾仕得 (Axalta) 发布的第 69 期《全球量产汽车色彩流行统计报告》(Axalta Coating Systems Ltd, 2021) 中的统计数据, 选取了最常见的车辆颜色, 包括白色、黑色、灰色、红色、蓝色、棕色、绿色等。具体的颜色分配情况如图 7(b) 所示。

街景拍摄组的物理变量包括光源位置、光源色温、机位 3 类。街景拍摄组采集数据的物理属性、分配的图像数量, 以及对应的示意图, 如图 8 所示。

3.1.2 人物拍摄组

人物拍摄组主要设置了黑色、实物、天空和真实 4 类背景, 实物组和真实组选择了复杂且颜色丰富的建筑作为背景, 并且与前景人物相互交织, 主要用于考察检测模型在多样、复杂和重叠的颜色背景下, 对合成图像和真实图像的识别能力。人物组采集数据选取了 12 个颜色多样的模拟人物, 人物高度为 33~46 mm, 人物宽度为 25 mm。

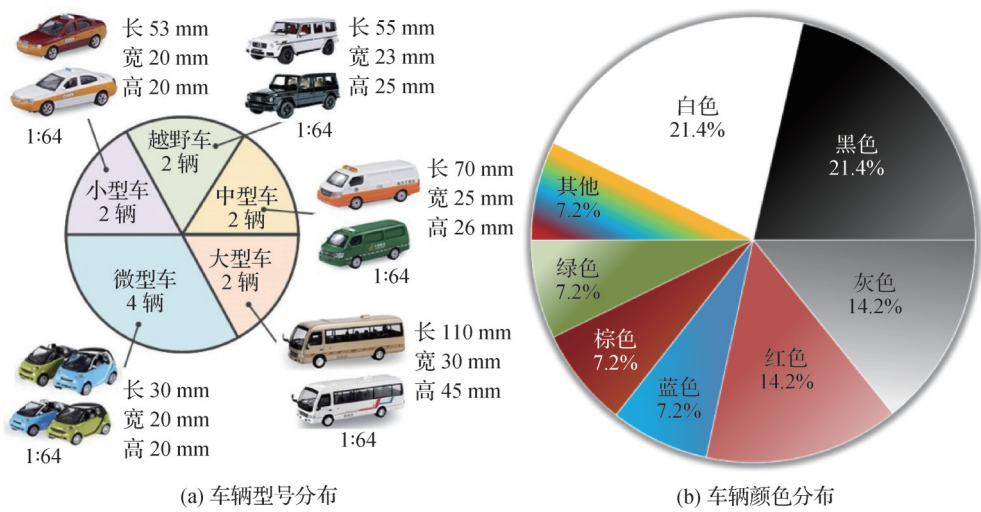


图7 街景拍摄组车辆型号及颜色分布

Fig. 7 Vehicle models and colors distribution of the street-view shooting group
((a) distribution of vehicle models; (b) distribution of vehicle colors)

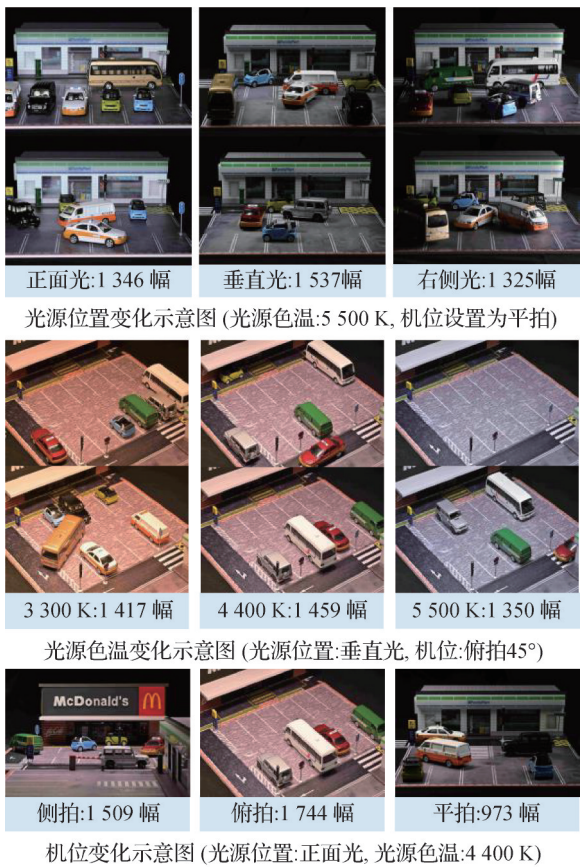


图8 街景拍摄组采集数据

Fig. 8 The street-view shooting group data collection

人物拍摄组的物理变量包括物距、景深和复杂背景3类,具体的物理变量取值、分配的图像数量以及其对应的示意图分别如图9所示。需要说明的

是,图9中的物距类数据采集分为单个物体和多个物体,多个物体的采集依据物体到镜头的距离标记为200 mm、300 mm、400 mm和500 mm。在采集过程中,可能在多个距离位置进行物体摆放。

3.2 图像合成

SIPD数据集的合成图像主要采用人工制作的方法,包括拼接、复制粘贴和删除3类。拼接类合成图像2 591幅,复制粘贴类合成图像2 020幅,删除类合成图像2 580幅,各类图像比例大致均衡,如图10(a)所示。合成图像考虑的物理属性包括光源位置、光源色温、机位、物距、景深和复杂背景6类,每类按照单一变量原则开展图像制作,具体的物理属性占比如图10(b)所示。

图11是合成图像的示意图及其对应的掩码图,每个展示组都包含拼接、复制粘贴和删除3类。街景类合成图像总量为1 671幅,包括光源位置合成图像558幅、光源色温合成图像558幅、机位合成图像555幅,各类图像比例分配均匀,如表1所示。人物类合成图像总量为5 520幅,包含物距合成图像3 680幅、景深合成图像1 840幅,每类均包含单个物体和多个物体的合成图像,如表2所示。SIPD数据集对每类物理属性中的图像按照7:2:1的方式设置了训练、验证和测试组,每个组内分配的合成方法数量比例相近。数据集下载链接为<https://doi.org/10.57760/sciencedb.j00240.00069>。

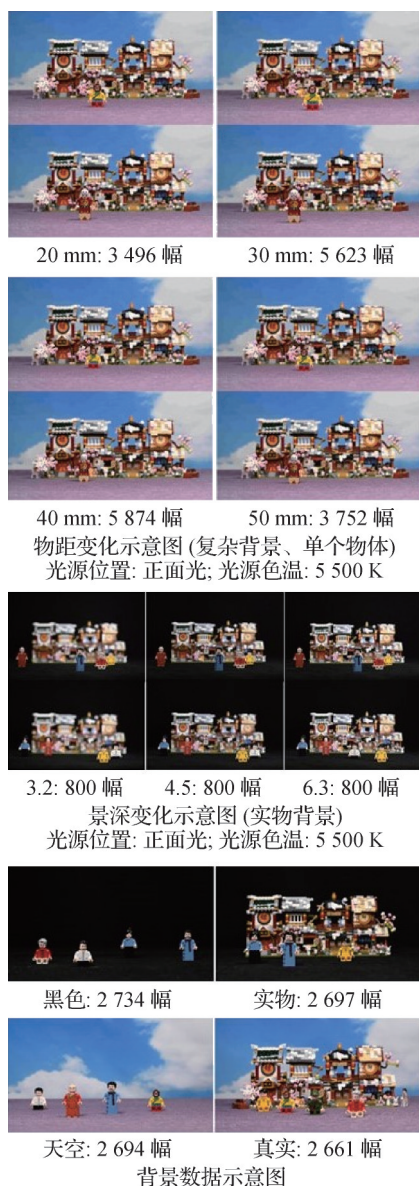


图9 人物拍摄组采集数据

Fig. 9 The character shooting group data collection

4 实验内容与结果分析

4.1 实验设置与基准模型

本文设置了消融实验、对比实验、增强实验和 SIPD 数据集实验,采用的评价指标为 F1 分数(F1 score)、曲线下面积(area under curve, AUC)和交并比(intersection over union, IoU)。本文对比的数据集包括 NIST 16(NIST, 2017)、Coverage(Wen 等, 2016)、CASIA(Dong 等, 2013)和 IMD 2020(Novozámský 等, 2021)。特别地,为配合本文方法的阴影检测,在上

述 4 个数据集的测试集中选取了 2 872 幅具有阴影的测试数据,以验证本文方法的有效性,选取的图像示例如图 12 所示。测试数据中的图像大多为真实场景,图像中的物体均在两个以上,并且有较为清晰的阴影区域。本文提出的 LSFNet 采用 GeForce GTX 1080Ti 进行训练, ResNet 模型的批处理大小为 32, 初始学习率为 0.000 1。本文对比模型与方法包括 LSTM-Encoder (long short-term memory encoder) (Bappy 等, 2017)、ManTra-Net (manipulation tracing network) (Wu 等, 2019)、SPAN (spatial pyramid attention network) (Hu 等, 2020)、ObjectFormer (Wang 等, 2022)、PSCC-Net (progressive spatio-channel correlation network) (Liu 等, 2022)、CFL-Net (contrastive forgery localization network) (Niloy 等, 2023)、Zhou 等人 (2023)、HiFi-Net (hierarchical fine-grained network) (Guo 等, 2023) 以及 APSC-Net (Qu 等, 2024)。本文使用了上述方法的程序,并参考 Liu 等人 (2022)、Wang 等人 (2022) 和 Qu 等人 (2024) 的部分实验结果。

4.2 消融实验

4.2.1 融合模块参数 α 与 β 的比例关系

本文对式(9)中融合模块参数 α 与 β 的比例关系展开了实验。实验的基准数据集为 NIST 16(NIST, 2017)数据集,测试数据为在 NIST 16 数据集中选取的阴影图像,实验结果如表 3 所示。

4.2.2 物理方法有效性验证实验

为验证深度神经网络和物理检测方法的有效性,本文对一致性检测方法的效果开展实验,测试数据与 4.1 节相同,结果如表 4 所示。“不包含”表示一致性检测分支的结果不计入总结果。“包含(7:3)”是以“7:3”的比例叠加计算分支预测结果。实验结果证明,本文使用的物理方法可以辅助深度神经网络方法开展合成检测。同时,根据表 3 和表 4 的实验结果,“7:3”的实验组表现出更优的 AUC 和 F1 分数,本文将按照“7:3”的实验组开展后续工作。

4.3 对比实验

本文使用 F1 分数和 AUC 作为测试基准,与 4.1 节中的测试数据保持一致,结果如表 5 所示。实验结果表明,LSFNet 在多个数据集上的性能显著优于主流方法。

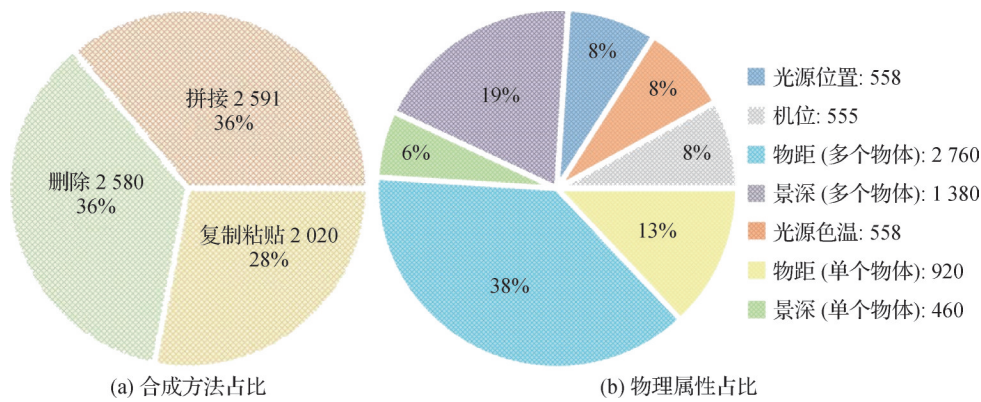


图 10 合成图像物理属性取值范围和分配的图像占比
Fig. 10 Value range of physical attributes of synthetic images and the proportion of allocated images
(a) proportion of synthetic methods; (b) proportion of physical attributes)



图 11 合成图像及掩码图像示例

Fig. 11 Examples of synthetic images and mask images

在 NIST 16 数据集上, LSFNet 的 AUC 相较于 LSTM-Encoder、SPAN、ObjectFormer、PSCC-Net、Zhou 等人 (2023) 提出的方法以及 HiFi-Net 分别高出 17.8%、3.8%、0.3%、2.1%、7.8% 和 1.9%; F1 分数相较于上述方法分别高出 9.0%、28.0%、1.8%、6.1%、0.6% 和 2.4%。

表 1 SIPD 数据集街景类合成数据制作数量
Table 1 Synthetic data of street-view category generation statistics for SIPD dataset

				/幅
合成方法	物理属性			合计
	光源位置	光源色温	机位	
拼接	438	438	435	1 311
复制粘贴	60	60	60	180
删除	60	60	60	180
合计	558	558	555	1 671

表 2 SIPD 数据集人物类合成数据制作数量
Table 2 Synthetic data of character category generation statistics for SIPD dataset

/幅

合成方法	物理属性				合计
	物距		景深		
	单个	多个	单个	多个	
拼接	120	360	200	600	1 280
复制粘贴	400	1 200	60	180	1 840
删除	400	1 200	200	600	2 400
合计	920	2 760	460	1 380	5 520



图 12 经过筛选的主流数据集示例图

Fig. 12 Example images from the mainstream datasets after filtering

在 Coverage 数据集上, LSFNet 的 AUC 相较于 LSTM-Encoder、SPAN、PSCC-Net、Zhou 等人(2023)方法以及 HiFi-Net 分别高出 21.2%、1.6%、2.5%、4.0% 和 0.2%; F1 分数分别高出 8.0%、24.0%、7.9%、2.1% 和 2.3%。

在 CASIA 数据集上, LSFNet 的 AUC 相较于 APSC-Net 和 HiFi-Net 分别高出 1.2% 和 4.9%, F1 分数分别高出 0.5% 和 3.8%。

结果表明, 本文框架通过深度融合神经网络与

表 3 结果融合模块参数确定
Table 3 Parameter determination for result fusion module

$\alpha:\beta$	AUC	F1
3:7	72.2	71.6
4:6	81.1	74.3
5:5	90.3	76.8
6:4	<u>94.1</u>	79.5
7:3	94.2	80.2
8:2	92.6	81.5
9:1	91.4	<u>81.2</u>
1:0	83.2	80.3

注:加粗、下划线字体分别表示各列最优、次优结果。

表 4 物理方法有效性实验
Table 4 Effectiveness evaluation of physics-based methods

检测方法	NIST 16		Coverage		CASIA		IMD 2020	
	AUC	F1	AUC	F1	AUC	F1	AUC	F1
不包含	83.2	80.0	89.5	73.7	85.5	53.4	89.3	61.2
包含(7:3)	94.2	80.2	93.6	79.3	90.3	58.1	91.2	66.9

物理检测方法,在合成图像检测任务中实现了更高的准确率。值得注意的是,APSC-Net在NIST 16数据集上的F1分数略高1.7%,分析认为,APSC-Net设计的共享网络增强了局部特征提取的鲁棒性,对于设计较为简单的数据集,具有更强优势。整体上,LSFNet通过结合边界光照图检测与物理属性一致性分析,在AUC及其他关键指标上具有显著优势。例如,在Coverage数据集上,LSFNet的AUC比APSC-Net高出2.2%,验证了本文方法对复杂光照场景的适应性。

为进一步验证LSFNet的IoU值,本文参考Qu等人(2024)方法设计了对比实验,并使用全量的数据集开展测试,结果如表6所示。根据实验结果,LSFNet在NIST 16、Coverage和CASIA数据集上的IoU值分别为41.2%、51.3%和78.4%,均显著高于PSCC-Net。特别是在Coverage数据集上, IoU值达到51.3%,远超PSCC-Net的19.7%,进一步验证了方法的优越性。APSC-Net在NIST 16和Coverage数据集上的IoU值分别为39.9%和49.0%,均低于LSFNet,但在CASIA数据集上, APSC-Net的IoU值高达

表5 优化后的模型AUC和F1分数
Table 5 AUC and F1 scores of optimized models

对比方法	/%					
	NIST 16		Coverage		CASIA	
	AUC	F1	AUC	F1	AUC	F1
LSTM-Encoder	76.4	71.2	72.4	71.3	79.5	51.3
SPAN	90.4	52.2	92.0	55.3	80.3	47.4
ObjectFormer	93.9	78.4	92.6	74.1	87.9	55.9
PSCC-Net	92.1	74.1	91.1	71.4	85.0	52.4
CFL-Net	93.5	79.2	93.4	73.1	86.3	53.3
Zhou 等人 (2023)	86.4	79.6	89.6	77.2	86.9	56.6
HiFi-Net	92.3	77.8	93.4	77.0	85.4	54.3
APSC-Net	93.2	81.9	91.4	78.5	89.1	57.6
LSFNet(本文)	94.2	80.2	93.6	79.3	90.3	58.1

注:加粗字体表示各列最优结果。

79.9%,略高于LSFNet的78.4%。这种差异可能源于APSC-Net采用了更复杂的自适应感知模块和自校准机制,能够更好地处理高分辨率图像和复杂合成场景。

表6 模型IoU对比结果
Table 6 Comparison of model IoU results

方法	/%		
	NIST 16	Coverage	CASIA
PSCC-Net	24.7	19.7	40.1
APSC-Net	39.9	49.0	79.9
LSFNet(本文)	41.2	51.3	78.4

注:加粗字体表示各列最优结果。

4.4 增强实验

本文进一步开展了数据增强和攻击实验,以验证LSFNet的鲁棒性。

4.4.1 数据增强

表7是数据增强的实验结果,包括图像反转、JPEG压缩以及添加噪声。旋转角度设置为 90° , JPEG压缩质量设置为50,噪声采用方差为3的高斯噪声。实验结果表明,经过旋转、JPEG压缩和高斯噪声等操作对合成图像检测带来了一定的影响。特别是在添加高斯噪声后再次叠加JPEG压缩,会给合成图像检测带来更大的影响。例如在NIST 16数据集中,无操作的检测组AUC为94.2%,F1分数为

80.2%。叠加了旋转和高斯噪声的检测组,AUC降低了14.2%,F1分数降低了4.3%。叠加了JPEG压缩和高斯噪声的检测组,也对AUC和F1分数具有很大的影响,较无操作检测组分别降低了18.1%和5.5%。同时,本文发现旋转操作对提出方法的影响不大,主要是因为旋转后的图像并未对光照图的分类产生影响,只影响了物理检测方法的效果,造成检测结果的小幅降低。

表7 数据增强方法比较实验
Table 7 Comparative experiments on data augmentation

操作	/%					
	NIST 16		Coverage		CASIA	
	AUC	F1	AUC	F1	AUC	F1
无操作	94.2	80.2	93.6	79.3	90.3	58.1
旋转	91.5	80.0	93.2	72.5	86.4	53.7
旋转 + JPEG	87.9	79.3	91.8	70.9	84.4	52.9
旋转 + 噪声	80.0	75.9	90.3	70.5	83.7	51.5
JPEG + 噪声	76.1	74.7	82.4	69.8	82.6	50.4

4.4.2 噪声攻击

为进一步验证噪声攻击对其他检测方法的影响,本文参考Liu等人(2022)的实验结果,以AUC为评价指标,对NIST 16的测试图像开展尺寸改变、高斯模糊、高斯噪声和JPEG压缩等噪声攻击实验,结果如表8所示。尺寸变化包括0.78和0.25,高斯模糊的内核大小为3和15,高斯噪声标准差为3和15, JPEG压缩的质量因子为100和50,对比方法包括ManTra-Net、SPAN、PSCC-Net、HiFi-Net、APSC-Net等。

实验结果表明,LSFNet在噪声攻击场景下展现出较强的综合抗干扰能力:在高斯模糊 $k=3$ 与高斯噪声 $\sigma=15$ 的实验中,其AUC分别达89.1%和77.6%,显著优于APSC-Net等对比方法,体现了对高频噪声与模糊的鲁棒性;在JPEG压缩 $q=100$ 和 $q=50$ 的实验中,LSFNet的AUC达到87.9%和85.6%,进一步验证了其抗压缩伪影的优势。然而,在尺寸变化 $0.78\times$ 实验中,LSFNet的AUC为85.9%,与APSC-Net的AUC值85.6%相接近,但受尺寸缩放导致的阴影检测失效问题限制,其定位精度未能显著提升。这一现象表明,物理方法对几何变换的敏感性与局部特征关联性仍有一定的空间,下一步可尝试引入

表 8 NIST 16数据集噪声攻击比较实验(AUC)
Table 8 Comparative experiments on noise attacks on NIST 16 dataset (AUC)

方法	尺寸变化		高斯模糊		高斯噪声		JPEG 压缩	
	0.78×	0.25×	$k = 3$	$k = 15$	$\sigma = 3$	$\sigma = 15$	$q = 100$	$q = 50$
	/%							
ManTra-Net	76.5	74.2	77.3	72.1	65.4	56.7	75.9	72.4
SPAN	84.3	81.4	82.1	79.5	71.7	65.4	80.9	77.6
PSCC-Net	84.8	83.5	85.9	80.3	78.2	75.4	85.4	84.8
HiFi-Net	83.3	82.0	84.3	81.9	80.1	75.3	83.9	79.2
APSC-Net	85.6	84.9	88.6	85.3	81.3	77.8	85.3	83.8
LSFNet(本文)	85.9	84.2	89.1	81.7	80.2	77.6	87.9	85.6

注:加粗字体表示各列最优结果。

多尺度动态感知模块或自适应阴影补偿机制,使模型更好地适应复杂形变。

4.5 SIPD数据集实验

4.5.1 数据集复杂度分析

本部分实验旨在验证 SIPD 数据集的高难度与 LSFNet 方法的优越性。为此,本文在 SIPD 数据集上进行了对比实验,实验基于未经过训练的模型展开,对比方法涵盖 LSTM-Encoder、SPAN、ObjectFormer、PSCC-Net、HiFi-Net 以及 APSC-Net 等主流方法,实验结果如表 9 所示。实验结果表明,LSFNet 在 SIPD 数据集上展现出较好的检测性能,其 AUC 达 70.6%,略优于 APSC-Net(70.3%)和 PSCC-Net(68.6%),同时 F1 分数为 67.8%,显著高于 SPAN(41.5%)和 LSTM-Encoder(45.7%),表明 LSFNet 具有较好的泛化能力。本文以未经过训练的模型为基础,按照表 8 的方法制作了 SIPD 数据集的噪声攻击类数据,方法包括尺寸变化、高斯模糊、高斯噪声和 JPEG 压缩 4 类,对比方法包括 SPAN 和 PSCC-Net,结果如表 10 所示。

根据表 9 和表 10,LSFNet 均表现出较好的检测结果,特别是在 SIPD 数据集的抗干扰实验中,LSFNet 表现出很强的稳定性。具体而言,LSFNet 在 SIPD 数据集抗干扰实验中的平均 AUC 达到 62.55%,分别较 SPAN 和 PSCC-Net 高出 5.9% 和 1.45%。由于 SIPD 数据集包含显著的光照与阴影差异等物理属性干扰,LSFNet 在尺寸变化和 JPEG 压缩等场景中表现尤为突出,充分体现了其对复杂

表 9 SIPD数据集对比实验
Table 9 Comparative experiments on SIPD dataset

对比方法	/%	
	AUC	F1
LSTM-Encoder	51.4	45.7
SPAN	65.0	41.5
ObjectFormer	68.3	64.9
PSCC-Net	68.6	67.2
HiFi-Net	67.3	65.4
APSC-Net	70.3	66.4
LSFNet(本文)	70.6	67.8

注:加粗字体表示各列最优结果。

物理特征的高鲁棒性。相比之下,SPAN 等方法的平均 AUC 在 56.65% 至 61.84% 之间,明显低于其在 NIST 16 等数据集上的表现,这表明对 SIPD 数据集的检测难度更大,现有方法需要进一步优化以提升其泛化性。

4.5.2 数据集分项测试

本部分实验主要验证数据在不同属性中的分布合理性以及 LSFNet 的稳定性。表 11 展示了 LSFNet 在 SIPD 数据集上的分项测试结果,测试了 LSFNet 对拼接、复制粘贴和删除操作的检测性能。实验结果表明,LSFNet 在拼接、复制粘贴和删除的检测效果较为均衡,这不仅表明该方法具有很强的稳定性,也证明了 SIPD 数据集的合成图像质量较高且数据分布合理。

表 10 SIPD 数据集噪声攻击对比实验(AUC)

Table 10 Comparative experiments on noise attacks on SIPD dataset (AUC)

/%

方法	尺寸变化		高斯模糊		高斯噪声		JPEG 压缩		平均
	0.78×	0.25×	$k = 3$	$k = 15$	$\sigma = 3$	$\sigma = 15$	$q = 100$	$q = 50$	
SPAN	60.1	54.3	61.5	58.4	59.4	46.3	60.7	52.5	56.65
PSCC-Net	66.3	59.2	65.7	60.8	61.2	53.4	62.4	59.8	61.10
HiFi-Net	67.9	60.7	64.3	62.1	60.7	54.3	65.3	59.4	61.84
APSC-Net	70.6	60.0	63.2	59.4	61.2	53.2	63.1	60.4	61.39
LSFNet(本文)	68.2	61.0	65.1	61.9	61.2	55.0	66.1	61.9	62.55

注:加粗字体表示各列最优结果。

表 11 LSFNet 在 SIPD 数据集的分项测试结果

Table 11 Subcategory test results of LSFNet on SIPD dataset

/%

合成方法	AUC	F1
拼接	70.3	68.2
复制粘贴	71.7	69.4
删除	69.9	65.8
总体	70.5	67.7

4.5.3 模型增益分析

本部分实验旨在验证 SIPD 数据集的有效性以及模型增益效果。具体而言,一是验证数据集是否可用于训练,二是评估模型增益效果。本文利用 SIPD 数据集的训练集对 LSFNet 进行训练,结果如表 12 所示。实验结果表明,LSFNet 在经过训练后,AUC 达到 88.4%,F1 分数达到 79.3%,较未经训练的结果分别提升了 17.8%和 11.5%。这一结果证明了 SIPD 数据集不仅可用于检测模型的训练、测试和

表 12 利用 SIPD 数据集训练和测试的结果

Table 12 Results of training and testing using SIPD dataset

/%

类别	AUC	F1
未经训练的结果	70.6	67.8
经过训练的结果	88.4	79.3
差值	+17.8	+11.5

注:+表示数值提升。

优化升级,还能显著提升模型性能。

本文进一步验证了训练后模型的抗干扰能力,以表 10 的结果为基础,本文将经过 SIPD 数据集训练后的 LSFNet 进行测试,如表 13 所示。根据实验结果,经过训练的 LSFNet 比未经训练的 AUC 的平均值提升了 18.1%。特别是在高斯噪声实验组中 AUC 提升最为明显,达到了 20% 以上。同时,训练后的 LSFNet 可以有效应对尺寸变化、高斯模糊等干扰条件下的合成图像检测,以此证明了 SIPD 数据集可以有效支撑 LSFNet 的训练和测试工作。

表 13 LSFNet 在 SIPD 数据集上的增益分析(AUC)

Table 13 LSFNet gain analysis based on SIPD dataset (AUC)

/%

方法	尺寸变化		高斯模糊		高斯噪声		JPEG 压缩		平均
	0.78×	0.25×	$k = 3$	$k = 15$	$\sigma = 3$	$\sigma = 15$	$q = 100$	$q = 50$	
LSFNet(未经训练)	68.2	61.0	65.1	61.9	61.2	55	66.1	61.9	62.55
LSFNet(经过训练)	82.3	78.9	85.9	76.7	82.7	75.2	83.9	79.6	80.65
差值	+14.1	+17.9	+20.8	+14.8	+21.5	+20.2	+17.8	+17.7	+18.1

注:+表示数值提升。

4.5.4 跨数据集泛化能力评估

本部分实验旨在验证 SIPD 数据集在模型训练中的有效性,以及评估 LSFNet 在跨数据集泛化能力方面的表现。本文参考 Qu 等人(2024)提出的方法,基于 SIPD 数据集实施消融实验。具体而言,使用 SIPD 数据集重新训练了开源方法 PSCC-Net 和 LSFNet,并分别在使用和不使用 SIPD 数据集的情况下进行对比实验。实验结果如表 14 所示。表 14 中,“未训”表示模型在未使用 SIPD 数据集训练时的性能,“训练”表示使用 SIPD 数据集训练后的性能。测试数据集选用 NIST 16、Coverage 和 CASIA,以更全面地评估模型在不同场景下的泛化能力。

通过 SIPD 数据集训练后,PSCC-Net 和 LSFNet 在多个测试数据集上的性能均取得了显著提升。PSCC-Net 在 NIST 16、Coverage 和 CASIA 数据集上的

IoU 分别从 24.7%、19.7% 和 40.1% 提升至 40%、31.8% 和 58.6%,相对提升幅度分别为 61.9%、61.4% 和 46.1%。这表明 SIPD 数据集显著增强了 PSCC-Net 在高分辨率、复杂合成场景下的定位能力。LSFNet 的提升更为显著,其在 NIST 16、Coverage 和 CASIA 数据集上的 IoU 分别从 41.2%、51.3% 和 78.4% 提升至 62.3%、70.9% 和 89.3%,相对提升幅度分别为 51.2%、38.2% 和 13.9%。这说明 LSFNet 对 SIPD 数据集的适应性更强,尤其是在处理复杂光照和多种合成类型时表现出更高的鲁棒性。

整体上看,SIPD 数据集对 PSCC-Net 和 LSFNet 的性能提升具有显著效果,尤其是在复杂合成场景下的定位精度和泛化能力方面,LSFNet 的表现尤为突出,表明其在处理多种合成类型和高噪声干扰时具有更强的适应性。

表 14 跨数据集泛化能力评估(IoU)

Table 14 Cross-dataset generalization evaluation (IoU)

方法	NIST 16数据集			Coverage数据集			CASIA数据集		
	未训	训练	数值提升	未训	训练	数值提升	未训	训练	数值提升
	/%								
PSCC-Net	24.7	40.0	61.9	19.7	31.8	61.4	40.1	58.6	46.1
LSFNet(本文)	41.2	62.3	51.2	51.3	70.9	38.2	78.4	89.3	13.9

4.5.5 数据集可视化分析

本文选取了 LSFNet 方法在复制粘贴、拼接、删除等合成图像中的可视化分析结果,如图 13 所示。光照分析和同源分析分别展示了光照图和同源一致性分析等过程。光照分析是将光照图预测到合成区域制作成为热力图,热力图中深色区域表示合成,浅色区域表示真实。同源一致性分析对各个物体及其阴影进行了可视化,并对物体和阴影间的角度进行了标注,异常角度可以用于检测合成。

在光照分析中,光照图可以对合成区域进行标注,例如图 13 复制粘贴类的合成图像中,将绿色车辆从右侧复制粘贴到左侧,光照分析的可视化结果对此进行了明显的标注。同理,第 2 列和第 3 列也表现出相同的趋势。但是整体而言,合成位置的标注不够明确,对于图像中未修改的区域,也可能被标记

为合成,这是本文下一步的研究重点。

本文关注到同源分析可以有效地对图像中的物体及其阴影之间的角度进行区分,特别是在复制粘贴组中,合成后的物体大多没有阴影。例如第 2 列和第 3 列中的合成图像,在阴影检测过程中未正常检测到阴影区域,同源分析结果存在异常。同时,第 1 列的合成物体阴影区域范围过小,同源分析认为物体和阴影之间的角度为 180 度,也与图像中的其他角度存在明显的差异。图 13 中的拼接类和删除类也反映出类似的检测效果,光源分析结果可以对合成区域进行粗略标注。在同源分析过程中,拼接组中的部分拼接物体,存在阴影与其他物体不一致的情况,可以根据此类不一致性进行合成判断。在删除组中,同源分析过程检测到了没有对应物体的阴影区域,也可以作为合成图像的判断依据。

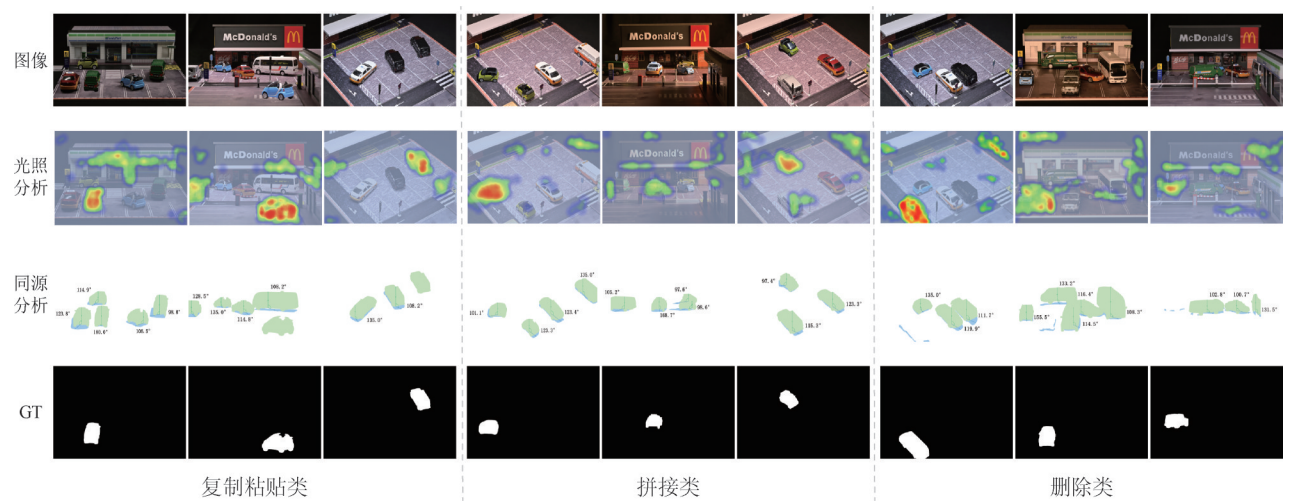


图 13 合成图像可视化分析结果

Fig. 13 Visual analysis results of synthetic images

5 结 论

本文提出并构建了LSFNet,这是一种结合深度神经网络与物理一致性的合成图像检测网络,通过深入研究同源顶点成像规律,融合光照与阴影信息,显著提升了检测的准确性与鲁棒性。为全面评估LSFNet的性能,在多个主流数据集中选取真实场景图像进行实验。实验结果表明,LSFNet相较于其他主流算法具有更高的准确率,能够精确标记合成区域,有效应对高质量合成图像的检测场景。然而,本文方法在处理无阴影图像的合成检测时仍存在局限性,未来研究将着重解决这一问题。

本文创新性地构建了SIPD数据集,旨在解决现有合成图像检测数据集缺乏物理属性的问题。该数据集涵盖多种物理属性,图像分辨率高,由专业人员进行后期制作,确保了数据质量。实验证明,SIPD数据集能够有效支持现有合成图像检测方法的训练、验证和测试,为该领域提供了新思路和数据支撑。此外,SIPD数据集在物体识别、光照检测、景深检测等多种研究场景中也具有广泛的应用潜力。未来,将继续对SIPD数据集进行完善和扩展,以支持更多领域的研究工作。

参考文献(References)

Atak I G and Yasar A. 2024. Image forgery detection by combining visual transformer with variational autoencoder network. Applied

Soft Computing, 165: #112068 [DOI: 10.1016/j. asoc. 2024. 112068]

Axalta Coating Systems Ltd. 2021. Axalta extends automotive color leadership with the 69th Global Automotive Color Popularity Report [EB/OL]. [2021-12-21].
<https://ir.axalta.com/news/press-releases/detail/52/axalta-extends-automotive-color-leadership-with-the-69th-global-automotive-color-popularity-report>

Bai R Y. 2024. Image manipulation detection and localization using multi-scale contrastive learning. Applied Soft Computing, 163: #111914 [DOI: 10.1016/j.asoc.2024.111914]

Bappy J H, Simons C, Nataraj L, Manjunath B S and Roy-Chowdhury A K. 2019. Hybrid LSTM and encoder-decoder architecture for detection of image forgeries. IEEE Transactions on Image Processing, 28(7): 3286-3300 [DOI: 10.1109/TIP.2019.2895466]

Bappy M J H, Roy-Chowdhury A K, Bunk J, Nataraj L and Manjunath B S. 2017. Exploiting spatial structure for localizing manipulated image regions//Proceedings of 2017 IEEE International Conference on Computer Vision (ICCV). Venice, Italy: IEEE: 4980-4989 [DOI: 10.1109/ICCV.2017.532]

Bunk J, Bappy J H, Mohammed T M, Nataraj L, Flenner A, Manjunath B S, et al. 2017. Detection and localization of image forgeries using resampling features and deep learning//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Honolulu, USA: IEEE: 1881-1889 [DOI: 10.1109/CVPRW.2017.235]

Chen J Y, Yang G B, Ding X L, Guo Z Q and Wang S. 2022. Robust detection of dehazed images via dual-stream CNNs with adaptive feature fusion. Computer Vision and Image Understanding, 217: #103357 [DOI: 10.1016/j.cviu.2022.103357]

Chen X, Yan B, Zhu J W, Wang D, Yang X Y and Lu H C. 2021a. Transformer tracking//Proceedings of 2021 IEEE/CVF Conference

- on Computer Vision and Pattern Recognition (CVPR). Nashville, USA: IEEE: 8122-8131 [DOI: 10.1109/CVPR46437.2021.00803]
- Crowson K, Biderman S, Kornis D, Stander D, Hallahan E, Castricato L, et al. 2022. VQGAN-CLIP: open domain image generation and editing with natural language guidance//Proceedings of the 17th European Conference on Computer Vision — ECCV 2022. Tel Aviv, Israel: Springer: 88-105 [DOI: 10.1007/978-3-031-19836-6_6]
- Ding B, Long C J, Zhang L and Xiao C X. 2019. ARGAN: attentive recurrent generative adversarial network for shadow detection and removal//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision (ICCV). Seoul, Korea (South): IEEE: 10212-10221 [DOI: 10.1109/ICCV.2019.01031]
- Dong J, Wang W and Tan T N. 2013. CASIA image tampering detection evaluation database//Proceedings of 2013 IEEE China Summit and International Conference on Signal and Information Processing. Beijing, China: IEEE: 422-426 [DOI: 10.1109/ChinaSIP. 2013. 6625374]
- Gholap S and Bora P K. 2008. Illuminant colour based image forensics//TENCON 2008—2008 IEEE Region 10 Conference. Hyderabad, India: IEEE: 1-5 [DOI: 10.1109/TENCON.2008.4766772]
- Guo R Q, Dai Q Y and Hoiem D. 2011. Single-image shadow detection and removal using paired regions//Proceedings of 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Colorado, USA: IEEE: 2033-2040 [DOI: 10.1109/CVPR.2011. 5995725]
- Guo X, Liu X H, Masi I and Liu X M. 2025. Language-guided hierarchical fine-grained image forgery detection and localization. *International Journal of Computer Vision*, 133(5): 2670-2691 [DOI: 10.1007/s11263-024-02255-9]
- Hadwiger B C and Riess C. 2022. Deep metric color embeddings for splicing localization in severely degraded images. *IEEE Transactions on Information Forensics and Security*, 17: 2614-2627 [DOI: 10.1109/TIFS.2022.3189527]
- Han R D, Wang X F, Bai N N, Wang Y H, Hou J P and Xue J R. 2024. HDF-Net: capturing homogeny difference features to localize the tampered image. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(12): 10005-10020 [DOI: 10.1109/TPAMI.2024.3432551]
- Hao J, Zhang Z X, Yang S C, Xie D and Pu S L. 2021. TransForensics: image forgery localization with dense self-attention//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal, Canada: IEEE: 15035-15044 [DOI: 10.1109/ICCV48922.2021.01478]
- Hao Q X, Ren R Y, Niu S Z, Wang K, Wang M S and Zhang J W. 2024. UGEE-Net: uncertainty-guided and edge-enhanced network for image splicing localization. *Neural Networks*, 178: #106430 [DOI: 10.1016/j.neunet.2024.106430]
- He K M, Gkioxari G, Dollár P, and Girshick R. 2017. Mask R-CNN//Proceedings of 2017 IEEE International Conference on Computer Vision. Venice, Italy: IEEE: 2961-2969 [DOI: 10.1109/ICCV. 2017.322]
- He K M, Zhang X Y, Ren S Q and Sun J. 2016. Deep residual learning for image recognition//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, USA: IEEE: 770-778 [DOI: 10.1109/CVPR.2016.90]
- Hu X F, Zhang Z H, Jiang Z Y, Chaudhuri S, Yang Z H and Nevatia R. 2020. SPAN: spatial pyramid attention network for image manipulation localization//Proceedings of the 16th European Conference on Computer Vision — ECCV 2020. Glasgow, UK: Springer: 312-328 [DOI: 10.1007/978-3-030-58589-1_19]
- Hu X W, Wang T Y, Fu C W, Jiang Y T, Wang Q and Heng P A. 2021. Revisiting shadow detection: a new benchmark dataset for complex world. *IEEE Transactions on Image Processing*, 30: 1925-1934 [DOI: 10.1109/TIP.2021.3049331]
- Huang X, Hua G, Tumblin J and Williams L. 2011. What characterizes a shadow boundary under the sun and sky?//Proceedings of 2011 International Conference on Computer Vision. Barcelona, Spain: IEEE: 898-905 [DOI: 10.1109/ICCV.2011.6126331]
- Johnson M K and Farid H. 2006. Exposing digital forgeries through chromatic aberration//Proceedings of the 8th Workshop on Multimedia and Security. Geneva, Switzerland: ACM: 48-55 [DOI: 10.1145/1161366.1161376]
- Johnson M K and Farid H. 2007. Exposing digital forgeries in complex lighting environments. *IEEE Transactions on Information Forensics and Security*, 2(3): 450-461 [DOI: 10.1109/TIFS.2007.903848]
- Kee E, O'Brien J F and Farid H. 2013. Exposing photo manipulation with inconsistent shadows. *ACM Transactions on Graphics*, 32(3): 1-12 [DOI: 10.1145/2487228.2487236]
- Lahiri A, Jain A K, Agrawal S, Mitra P and Biswas P K. 2020. Prior guided GAN based semantic inpainting//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, USA: IEEE: 13693-13702 [DOI: 10.1109/CVPR42600.2020.01371]
- Lalonde J F, Efros A A and Narasimhan S G. 2010. Detecting ground shadows in outdoor consumer photographs//Proceedings of the 11th European Conference on Computer Vision 2010. Heraklion, Greece: Springer: 322-335 [DOI: 10.1007/978-3-642-15552-9_24]
- Le H, Vicente T F Y, Nguyen V, Hoai M and Samaras D. 2018. A+D Net: training a shadow detector with adversarial shadow attenuation//Proceedings of the 15th European Conference on Computer Vision—ECCV 2018. Munich, Germany: Springer: 680-696 [DOI: 10.1007/978-3-030-01216-8_41]
- Lee C H, Liu Z W, Wu L Y and Luo P. 2020. MaskGAN: towards

- diverse and interactive facial image manipulation//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, USA: IEEE: 5548-5557 [DOI: 10.1109/CVPR42600.2020.00559]
- Lin T Y, Dollár P, Girshick R, He K M, Hariharan B and Belongie S. 2017. Feature pyramid networks for object detection//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, USA: IEEE: 936-944 [DOI: 10.1109/CVPR.2017.106]
- Liu W X, Zhang H, Lin X Y, Zhang Q, Li Q, Liu X X, et al. 2024. Attentive and contrastive image manipulation localization with boundary guidance. *IEEE Transactions on Information Forensics and Security*, 19: 6764-6778 [DOI: 10.1109/TIFS.2024.3424987]
- Liu X H, Lin Z, Zhang J M, Zhao H D, Tran Q, Wang X G, et al. 2020. Open-edit: open-domain image manipulation with open-vocabulary instructions//Proceedings of the 16th European Conference on Computer Vision — ECCV 2020. Glasgow, UK: Springer: 89-106 [DOI: 10.1007/978-3-030-58621-8_6]
- Liu X H, Liu Y J, Chen J and Liu X M. 2022. PSCC-Net: progressive spatio-channel correlation network for image manipulation detection and localization. *IEEE Transactions on Circuits and Systems for Video Technology*, 32 (11): 7505-7517 [DOI: 10.1109/TCSVT.2022.3189545]
- Merlin N R G, Santhosh K B, Cristin R, and Sekar R. 2019. Improved segmentation techniques for image manipulation with shadow based digital forgery using vanishing point computation *International Journal of Scientific and Research*. 8(12): 503-507
- Nasiri M and Behrad A. 2019. Using expectation-maximization for exposing image forgeries by revealing inconsistencies in shadow geometry. *Journal of Visual Communication and Image Representation*, 58: 323-333 [DOI: 10.1016/j.jvcir.2018.12.007]
- Ng T T and Chang S F. 2004. A Data Set of Authentic and Spliced Image Blocks. 203-2004-3. Columbia University.
- Nillius P and Eklundh J O. 2001. Automatic estimation of the projected light source direction//Proceedings of 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR). Kauai, USA: IEEE: 1076-1083 [DOI: 10.1109/CVPR. 2001.990650]
- Niloy F F, Bhaumik K K and Woo S S. 2023. CFL-Net: image forgery localization using contrastive learning//Proceedings of 2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). Waikoloa, USA: IEEE: 4631-4640 [DOI: 10.1109/WACV56688.2023.00462]
- NIST. 2017. Nimble challenge 2017 evaluation [EB/OL]. [2024-11-01]. <https://www.nist.gov/itl/iad/mig/nimble-challenge-2017-evaluation/>
- Novozámský A, Mahdian B and Saic S. 2021. Extended IMD2020: a large-scale annotated dataset tailored for detecting manipulated images. *IET Biometrics*, 10 (4): 392-407 [DOI: 10.1049/bme2.12025]
- Palmer C J, Goddard E and Clifford C W G. 2022. Face detection from patterns of shading and shadows: the role of overhead illumination in generating the familiar appearance of the human face. *Cognition*, 225: #105172 [DOI: 10.1016/j.cognition.2022.105172]
- Pan W Y, Ma W T, Zhao S, Gu L C, Shi G L, Xia Z H, et al. 2024. Image manipulation detection with cascade hierarchical graph representation. *IEEE Transactions on Circuits and Systems for Video Technology*, 34 (9): 8672-8683 [DOI: 10.1109/TCSVT. 2024.3390127]
- Park J, Cho D, Ahn W and Lee H K. 2018. Double JPEG detection in mixed JPEG quality factors using deep convolutional neural network//Proceedings of the 15th European Conference on Computer Vision. Munich, Germany: Springer: 656-672 [DOI: 10.1007/978-3-030-01228-1_39]
- Pomari T, Ruppert G, Rezende E, Rocha A and Carvalho T. 2018. Image splicing detection through illumination inconsistencies and deep learning//Proceedings of 2018 IEEE International Conference on Image Processing (ICIP). Athens, Greece: IEEE: 3788-3792 [DOI: 10.1109/ICIP.2018.8451227]
- Qu C F, Zhong Y W, Liu C Y, Xu G T, Peng D I, Guo F J, et al. 2024. Towards modern image manipulation localization: a large-scale dataset and novel methods//Proceedings of 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 10781-10790
- Ren S Q, He K M, Girshick R and Sun J. 2017. Faster R-CNN: towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39 (6): 1137-1149 [DOI: 10.1109/TPAMI.2016.2577031]
- Riess C and Angelopoulos E. 2010. Scene illumination as an indicator of image manipulation//Proceedings of the 12th International Conference on Information Hiding. Calgary, Canada: Springer: 66-80 [DOI: 10.1007/978-3-642-16435-4_6]
- Salloum R, Ren Y Z and Jay Kuo C C. 2018. Image splicing localization using a multi-task fully convolutional network (MFCN). *Journal of Visual Communication and Image Representation*, 51: 201-209 [DOI: 10.1016/j.jvcir.2018.01.010]
- Sekhar P N R L C and Shankar T N. 2021. Splicing forgery localisation using colour illumination inconsistencies. *International Journal of Electronic Security and Digital Forensics*, 13 (3): 346-358 [DOI: 10.1504/IJESDF.2021.114956]
- Shi J, Xu N, Zheng H T, Smith A, Luo J B and Xu C L. 2022. Space-Edit: learning a unified editing space for open-domain image color editing//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans, USA: IEEE: 19698-19707 [DOI: 10.1109/CVPR52688.2022.01911]
- Song W W, Li S T, Fang L Y and Lu T. 2018. Hyperspectral image classification with deep feature fusion network. *IEEE Transactions on*

- Geoscience and Remote Sensing, 56(6): 3173-3184 [DOI: 10.1109/TGRS.2018.2794326]
- Tan R T, Nishino K and Ikeuchi K. 2004. Color constancy through inverse-intensity chromaticity space. *Journal of the Optical Society of America A*, 21(3): 321-334 [DOI: 10.1364/JOSAA.21.000321]
- van de Weijer J, Gevers T and Gijsenij A. 2007. Edge-based color constancy. *IEEE Transactions on Image Processing*, 16(9): 2207-2214 [DOI: 10.1109/TIP.2007.901808]
- Vicente T F Y, Hoai M and Samaras D. 2018. Leave-one-out kernel optimization for shadow detection and removal. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(3): 682-695 [DOI: 10.1109/TPAMI.2017.2691703]
- Wang J K, Wu Z X, Chen J J, Han X T, Shrivastava A, Lim S N, et al. 2022. ObjectFormer for image manipulation detection and localization//*Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. New Orleans, USA: IEEE: 2354-2363 [DOI: 10.1109/CVPR52688.2022.00240]
- Wang T Y, Hu X W, Wang Q, Heng P A and Fu C W. 2020. Instance shadow detection//*Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Seattle, USA: IEEE: 1877-1886 [DOI: 10.1109/CVPR42600.2020.00195]
- Wen B H, Zhu Y, Subramanian R, Ng T T, Shen X J and Winkler S. 2016. COVERAGE — A novel database for copy-move forgery detection//*Proceedings of 2016 IEEE International Conference on Image Processing (ICIP)*. Phoenix, USA: IEEE: 161-165 [DOI: 10.1109/ICIP.2016.7532339]
- Wu Y, Abd-Almageed W and Natarajan P. 2018. Image copy-move forgery detection via an end-to-end deep neural network//*Proceedings of 2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*. Lake Tahoe, USA: IEEE: 1907-1915 [DOI: 10.1109/WACV.2018.00211]
- Wu Y, AbdAlmageed W and Natarajan P. 2019. ManTra-Net: manipulation tracing network for detection and localization of image forgeries with anomalous features//*Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Long Beach, USA: IEEE: 9535-9544 [DOI: 10.1109/CVPR.2019.00977]
- Xie S N, Girshick R, Dollár P, Tu Z W and He K M. 2017. Aggregated residual transformations for deep neural networks//*Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Honolulu, USA: IEEE: 5987-5995 [DOI: 10.1109/CVPR.2017.634]
- Xing Z H, Wang T Y, Hu X W, Wu H R, Fu C W and Heng P A. 2024. Video instance shadow detection under the sun and sky. *IEEE Transactions on Image Processing*, 33: 5715-5726 [DOI: 10.1109/TIP.2024.3468877]
- Ye G S, Wang J M, Yang Z Z, Zhang Y H, Cui R K and Xuan S. 2023. Survey of image composition based on deep learning. *Journal of Image and Graphics*, 28(12): 3670-3698 (叶国升, 王建明, 杨自忠, 张宇航, 崔荣凯, 宣帅. 2023. 深度学习图像合成研究综述. *中国图象图形学报*, 28(12): 3670-3698) [DOI: 10.11834/jig.220713]
- Zhang W, Cao X C, Zhang J W, Zhu J G and Wang P. 2009. Detecting photographic composites using shadows//*Proceedings of 2009 IEEE International Conference on Multimedia and Expo*. New York, USA: IEEE: 1042-1045 [DOI: 10.1109/ICME.2009.5202676]
- Zhang Y, Goh J, Win L L and Thing V. 2016. Image region forgery detection: a deep learning approach//*Proceedings of 2016 Singapore Cyber-Security Conference (SG-CRC)* 2016. Singapore, Singapore: IOS Press: 1-11 [DOI: 10.3233/978-1-61499-617-0-1]
- Zheng Q L, Qiao X T, Cao Y and Lau R W H. 2019. Distraction-aware shadow detection//*Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Long Beach, USA: IEEE: 5162-5171 [DOI: 10.1109/CVPR.2019.00531]
- Zhou J Z, Ma X C, Du X, Alhammadi A Y and Feng W T. 2023. Pre-training-free image manipulation localization through non-mutually exclusive contrastive learning//*Proceedings of 2023 IEEE/CVF International Conference on Computer Vision (ICCV)*. Paris, France: IEEE: 22289-22299 [DOI: 10.1109/ICCV51070.2023.02042]
- Zhou P, Han X T, Morariu V I and Davis L S. 2018. Learning rich features for image manipulation detection//*Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Salt Lake City, USA: IEEE: 1053-1061 [DOI: 10.1109/CVPR.2018.00116]
- Zhu Y, Shen X Q, Liu S K, Zhang X L and Yan G. 2021. Image splicing location based on illumination maps and cluster region proposal network. *Applied Sciences*, 11(18): #8437 [DOI: 10.3390/app11188437]

作者简介

薛子育,男,博士,工程师,主要研究方向为人工智能安全和人工智能生成内容鉴别。E-mail: xueziyu@abs.ac.cn

刘庆同,男,工程师,主要研究方向为人工智能安全和人工智能生成内容鉴别。E-mail: liuqingtong@abs.ac.cn

周康能,男,博士研究生,主要研究方向为三维内容生成制作。E-mail: elliszk@163.com

王伟,男,副研究员,主要研究方向为人工智能安全和多媒体内容安全。E-mail: wei.wong@ia.ac.cn

姜秀华,女,教授,主要研究方向为智能视音频技术。

E-mail: jiangxiuhua@cuc.edu.cn