

文章编号: 1672-8785(2023)12-0041-08

基于特征增强的模糊红外刑侦目标提取算法研究

于 晓* 姜晨慧

(天津理工大学电气工程与自动化学院, 天津 300384)

摘 要: 本研究旨在解决红外图像刑侦场景中目标边缘模糊和轮廓不完整等问题。提出了一种基于特征增强的模糊刑侦目标提取方法。首先, 设计模糊刑侦目标边缘提取网络模型(Blurry Criminal Investigation Target Edge Extraction Network Combined with Spatial Channel Attention, BCES-Net), 并利用空间通道关注模块(Spatial and Channel Attention Module, STCAM)获取具有强语义信息的特征图像。接着通过建模提取来获取包含语义类别信息的边缘特征和模糊刑侦目标特征。在训练过程中, 基于特定损失函数和多种特征融合技术, 通过反复监督学习和训练校正, 提高了边缘和模糊刑侦目标分割性能。在手部热痕迹数据集上, 与 DeeplabV3+, U-Net, HRNet, PSPNet 等模型相比, BCES-Net 模型在均交并比(mean Intersection over Union, mIoU)、平均精度均值(mean Average Precision, mAP)、准确率等评价指标上显著优越, mIoU 达到 88.3%, mAP 达到 94.35%, 准确率达到 95.5%。本研究创新性地提高了模糊红外刑侦目标提取的准确度, 为实际应用提供了技术支持。

关键词: 红外图像; 特征增强; 注意力模块; 刑侦目标提取

中图分类号: TP391.4 文献标志码: A DOI: 10.3969/j.issn.1672-8785.2023.12.006

Research on Fuzzy Infrared Criminal Investigation Target Extraction Algorithm Based on Feature Enhancement

YU Xiao*, JIANG Chen-hui

(School of Electrical Engineering and Automation, Tianjin University of Technology, Tianjin 300384, China)

Abstract: The aim of this study is to solve the problems of blurred edges and incomplete contours in infrared image criminal investigation scenes. A method of extracting fuzzy criminal investigation targets based on feature enhancement is presented in this paper. Firstly, the BCES-Net network model is designed and the feature images with strong semantic information are obtained by using STCAM. Then the edge features and fuzzy criminal investigation target features containing semantic category information are obtained by modeling extraction. In the training process, based on specific loss functions and multiple feature fusion techniques, the seg-

收稿日期: 2023-08-02

基金项目: 国家自然科学基金项目(61502340); 天津市自然科学基金项目(18JCQNJC01000)

作者简介: 于晓(1985-), 男, 山东临沂人, 副教授, 博士, 主要从事人工智能和图像处理方面的研究。

*通讯作者: E-mail: yx_tjut@163.com

mentation performance of edge and fuzzy criminal investigation targets is improved through repeated supervised learning and training correction. In the hand heat trace data set, compared with DeeplabV3+, U-Net, HR-Net, PSPNet and other models, BCES-Net is significantly superior in mIoU, mAP, accuracy and other evaluation indexes. mIoU reaches 88.3%, mAP reaches 94.35%, and the accuracy reaches 95.5%. This research innovatively improves the extraction accuracy of fuzzy infrared criminal detection targets and provides technical support for practical application.

Key words: infrared image; feature enhancement; attention module; criminal investigation target extraction

0 引言

红外热成像技术在公安刑侦等领域中的应用已经得到了广泛关注^[1]。手部热痕迹识别技术被广泛应用于案发现场的痕迹检测。通过对热痕迹进行检测与分析,可以还原罪犯在现场犯罪的过程,缩小嫌疑人排查范围,从而加速查案进程。在手部热痕迹的识别过程中,热痕迹特征的提取是至关重要的一步,其准确性直接决定了后续热痕迹信息的分析结果和追踪效果。由于手部红外热痕迹图像通常由体表红外辐射的强度分布生成,随着时间的推移,手部局部的热改变会对红外辐射图像产生影响,导致红外热痕迹图像出现模糊和信息丢失的问题(见图1)。因此,在手部热痕迹识别过程中,提取模糊的红外热痕迹图像中的目标与背景是当前红外图像在公安刑侦等领域应用时需要解决的迫切问题之一。本研究旨在探讨提取手部红外热痕迹图像中刑侦目标的方法,以增强犯

罪嫌疑人追踪的准确性和可靠性。

近年来, VGGNet^[2]、ResNet^[3]、MobileNet^[4]和EfficientNet^[5]等深度卷积神经网络已在目标识别方面取得了显著的成果。然而,在这些网络的图像分割算法中,大多都是简单地从图像分类算法中转移而来。由于网络特征表达能力不足,这些算法往往忽略了语义类别的细节(如边缘和模糊刑侦目标)。为了满足语义分割的要求并提高其准确性和效率,需要对图像分割算法进行优化。在分类网络中,频繁采用池化操作和调整卷积步幅时丢失了许多细节信息,可能会对空间分辨率产生不利影响,导致分割性能的降低。为了解决这个问题,可以采用一些技术手段(如反卷积、上采样、跳跃连接和膨胀卷积等),以提高分辨率并保留更多的细节信息。此外,还可以使用更深的网络架构和更多的训练数据来提高分割性能。

研究模糊刑侦目标提取时,我们发现它的

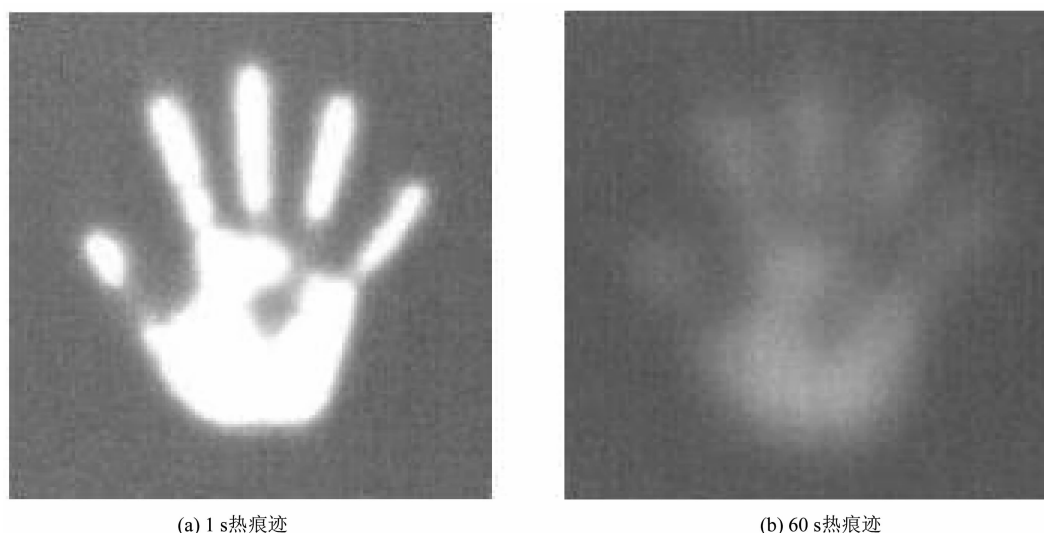


图1 红外手部热痕迹图像

分割难度在于目标特征较浅、语义信息较少以及背景之间的尺寸不同。较小的感受野难以获取全局语义信息,但较大的感受野则可能导致模糊的刑侦目标特征遗失。在图像分割领域已经有很多研究工作,虽然其中的一些已经取得了良好的成果,但仍然需要进一步优化分割性能以满足特定需求。

相对于传统的目标提取方法,深度神经网络利用数据驱动提取特征,可以学习到深层、数据集特定的特征表示,在大规模图像分割任务中表现出优异的性能。Chen L C 等人^[6]提出了一种新型深度卷积神经网络模型。他们采用 Atrous 卷积、CRFs 和多尺度上下文信息等技术,同时获取高分辨率的细节信息和宽范围的上下文信息,并通过 ASPP 模块和 CRFs 后处理进一步提高分割精度。Hao Y 等人^[7]提出了一种基于边缘引导流的交互式分割方法。该方法可以在不需要像素级注释的情况下提供高质量的训练数据,从而提高图像分割性能。Yu Y 等人^[8]提出了一种基于无监督卷积神经网络(Convolutional Neural Network, CNN)的边缘感知超像素算法。超像素提供了图像的有效表示,并适用于后续的视觉任务。该算法基于一个无监督的 CNN,使用边缘感知的方法来生成超像素。Zhang D 等人^[9]提出了一种基于交叉任务一致性理论的方法。他们将边缘先验信息纳入语义分割任务中,以获得更好的结果。Liu H 等人^[10]提出了一种名为 MEA-Net 的新型编码器-解码器结构。该结构使用 E1 和 E2 中的边缘模块来处理分割任务中的边缘信息。Chen L C 等人^[11]提出了一种新型的编码器-解码器架构。他们采用学习率自适应的 ASPP 模块进行语义编码,然后将其输出传递到可分离卷积层以提高效率,并提出了基于全图像随机扰动的数据增强方法和多尺度训练方法来进一步提高模型的鲁棒性和泛化能力。

以上提到的深度学习方法和技術已经在场景分割任务中取得了巨大进步。而本文旨在提高刑侦场景分割任务中模糊刑侦目标和边缘等

的分割精度,特别是改善手部热红外刑侦目标的提取以及刑侦目标和边缘的分割性能。因此,本文设计了一种结合 STCAM 的 BCES-Net 模型,将网络高层特征经过空洞卷积池化金字塔处理,用全局语义信息指导浅层高分辨图像特征训练,以保持浅层特征的分辨率和语义信息。同时,将 STCAM 应用于多种特征融合方法,设计新的损失函数以修正边缘和刑侦目标类别,成功提高了分割性能并降低了内轮廓噪声。本文研究结果对手部热红外刑侦目标的提取与识别等实际应用具有一定意义。

1 相关工作

1.1 空洞空间卷积池化金字塔

空洞空间卷积池化金字塔(Atrous Spatial Pyramid Pooling, ASPP)是一种用于图像语义分割的卷积神经网络模块^[11]。它被设计用于解决卷积神经网络在提取不同尺度物体时受限的问题。该模块通过在网络最后一层添加一个多尺度金字塔池化模块,提高了网络的感受野、分辨率和上下文信息。ASPP 模块运用多种采样率的空洞卷积来提取各种尺度的特征,有助于丰富特征表达。在每个尺度上,ASPP 模块对特征图进行金字塔池化,汇总尺度不同的空间上下文信息。最后将这些特征图拼接,并通过卷积层得到最终的分割结果。ASPP 模块的具体结构如图 2 所示。

相比于其他分割模块,ASPP 模块拥有更广的有效感受野、更多的上下文信息以及更强大的特征提取能力,在各种图像分割任务中取得了很好的效果。近年来,该模块已经被广泛应用于自然图像语义分割、医学图像分割等诸多领域,成为图像分割中不可或缺的重要组成部分。

1.2 空间注意力机制

目前的图像语义分割模型通过深度卷积神经网络来提取高层图像特征,以获取更丰富的语义信息。然而,这些高层特征在细节信息方面表现不佳。与此相反,浅层图像特征富含细节,但缺乏语义。简单地融合高层和低层特征

并不能显著提升分割精度。为了扩展空间接收域,一些方法(如 Deeplabv3^[11]和 PSPNet^[12])采用了多尺度特征提取策略。然而,这些方法主要关注局部特征关系,因此所获得的上下文语义信息相对受限^[13]。在一些复杂的场景中,可能无法处理图像中存在的杂乱噪声问题。CCNet^[14]的交叉注意力机制将特征图划分成若干小块,并计算每个小块与其他小块的交叉注意力权重。这一过程可以增强特征的表示能力,捕获不同区域之间的相关性。EMANet^[15]则引入了一种基于期望最大化算法的注意力机制,将互补的信息编码到不同的分支中,通过期望最大化将分支信息融合到一起,从而达到最优的分割结果。Zhong Z 等人^[16]提出了一种高效的压缩注意力网络结构——SANet。他们采用残差注意力模块对特征图进行加权,使网络更加注重包含目标信息的区域。然而,这些方法并未考虑像素和类别在空间上的直

接联系。这一点对于减少噪声信息和增加空间上下文的解释性非常重要。因此,我们提出了一种 STCAM 模块,即利用空间和通道注意力在像素空间中获得更强的语义信息。

2 BCES-Net 网络模型

本文提出了一种由 5 个阶段构成的 BCES-Net 网络模型。第一阶段对输入图像进行浅层特征提取,以获取浅层特征。接下来的三个阶段逐渐增加网络深度,提取更深层次的特征。第五阶段使用 ASPP 模块对深层特征进行处理,从而进一步提升了模型的感知能力。ASPP 模块对主干网络输出的特征进行了不同膨胀率的处理,从而增强了模型的感知能力和上下文语义信息。浅层特征与 ASPP 处理后的深层特征的融合有助于获取满足不同尺度目标的语义信息和增加部分细节信息。然后,这些特征被送入 STCAM 模块,以提高对边缘和模糊图像的特征提取能力,从而获取边缘特征和模

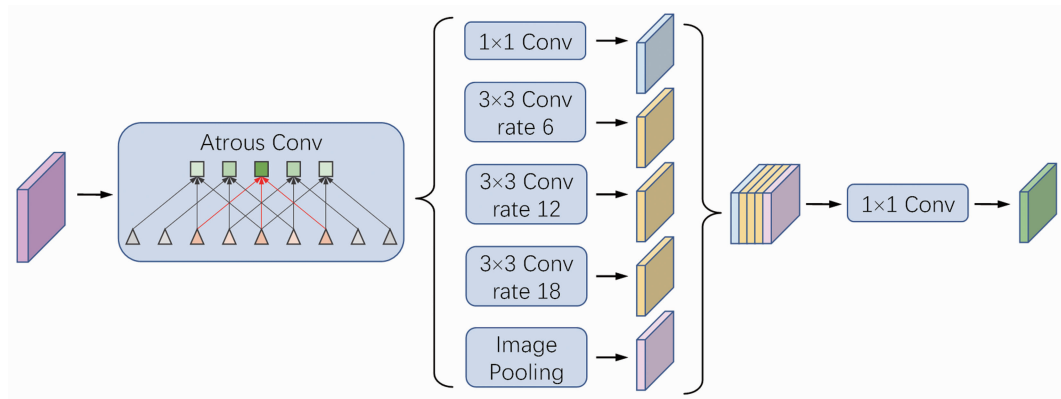


图 2 ASPP 模块结构示意图

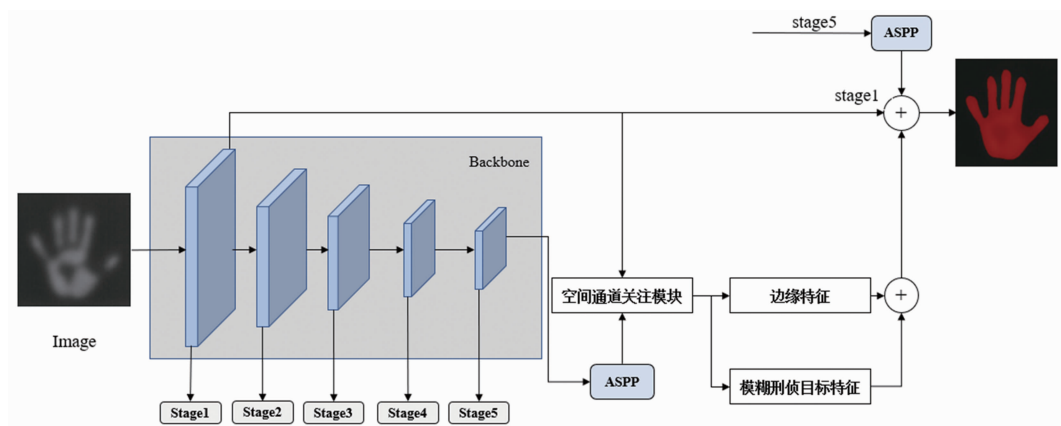


图 3 本文算法模型结构图

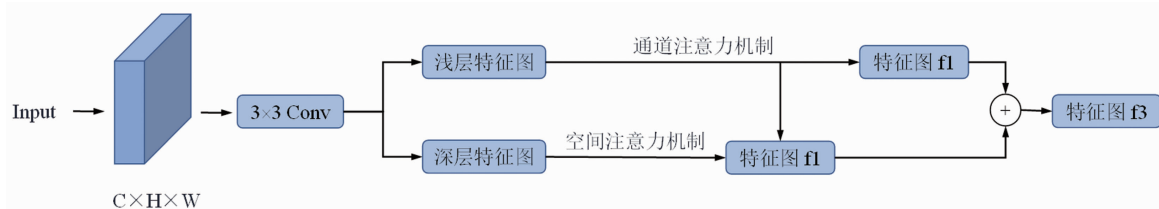


图 4 STCAM 模块结构图

模糊刑侦目标特征。最后, 通过特征融合(合并), 将浅层特征、边缘特征、模糊刑侦目标特征以及经过 ASPP 处理的深层特征生成最终的图像分割结果, 并综合考虑浅层和深层特征的信息, 提高了模型的表现力。本文算法模型结构如图 3 所示。

2.1 STCAM 模块

针对当前语义分割算法在处理模糊的刑侦目标时可能造成特征丧失, 从而影响准确分割的问题, 我们设计了一种 STCAM 模块。该模块将空间关注和通道关注相结合, 以获得富含强大语义信息的图像特征。具体而言, 我们将高层特征中蕴含的丰富语义信息反馈至浅层, 通过 STCAM 模块对浅层特征进行训练, 以赋予其更丰富的细节和语义信息。这一策略使得我们在像素空间中获取了更为丰富的语义信息, 从而有效地解决了模型中浅层特征在像素空间中缺乏语义信息的问题。STCAM 模块的结构如图 4 所示。

STCAM 模块结合了空间和通道注意力机制, 以增强特征表示能力并优化边缘和模糊特征的刑侦目标细节分割。该模块以大小为 $C \times H \times W$ 的特征图作为输入。其中, C 表示通道数, H 和 W 是输入图的高度和宽度维度。首先, 我们对输入特征图应用的可分离卷积核大小为 3×3 、膨胀因子为 1。然后, 将输出分成两组: 浅层和深层特征图。我们对浅层特征图应用通道注意力, 获得一个沿通道方向的权重向量。将通道向量乘以可学习权重矩阵, 然后应用 *Softmax* 操作计算每个通道的权重, 并使用这些权重计算浅层特征值的加权和 $f1$ 。接下来, 分别对浅层和深层特征图进行全局平

均池化, 以获取它们的平均值。然后使用两个不同的卷积核获得两组权重映射 $map1$ 和 $map2$, 捕捉浅层和深层特征图之间像素级的空间依赖关系。我们通过对权重映射与可学习参数矩阵的乘积应用 *Softmax* 操作来计算出通道权重。最后, 使用加法或级联的方式结合浅层特征值 $f1$ 和深层特征值 $f2$, 以获取特征图 $f3$ 。特征图 $f3$ 可作为下一模块的输入, 或作为语义分割模型的最终输出。

2.2 损失函数设计

针对边缘和模糊刑侦目标特征进行监督学习, 不仅监督主干网络的最终分割特征图, 还增加了边缘损失函数和模糊刑侦目标损失函数来监督语义边缘和语义小目标的学习过程。为了更加有针对性地指导和监督模型的学习过程, 本文采用交叉熵损失函数作为监督信号。在二分类的情况下, 模型需要预测的结果仅包括两种情形, 即对于每个类别, 分别对应着概率 p 和 $1-p$ 的预测。计算公式如下:

$$L = \frac{1}{N} \sum_i L_i = \frac{1}{N} \sum_i [-y_i \log(p_i) + (1 - y_i) \log(1 - p_i)] \quad (1)$$

式中, y_i 表示样本 i 的 label。对于边缘损失函数, 真实值为 1 表示该像素点是边缘, 真实值为 0 表示不是边缘; 对于刑侦目标损失函数, 真实值为 1 表示该像素点属于刑侦目标, 真实值为 0 表示不属于刑侦目标。 p_i 表示样本 i 被预测为正类的概率。

根据多任务学习思想^[17], 结合 3 个模块的预测损失, 定义最终损失函数:

$$L = L_s + L_e + L_c \quad (2)$$

式中, L_s 为 STCAM 模块的损失; L_e 为边缘提取模块的损失; L_c 为刑侦目标提取模块的

损失。

3 实验结果与分析

3.1 数据集介绍

本文中所构建的数据集是使用 Fluke Tix640 红外热像仪拍摄的。为了验证本文提出的改进模型在提取热红外刑侦目标方面的性能,我们对采集的手部热痕迹红外数据图像采用了随机翻转、旋转和裁剪等数据扩充方法,以增加数据集的多样性。最终,我们的手部热痕迹红外图像数据集包括 1320 张图像。其中 90% 用于训练数据,10% 用于验证数据。在实验中,图像的尺寸为 512×512 像素。

3.2 实验环境与评估标准

实验采用 PyTorch 深度学习框架搭建深度神经网络模型,计算机的处理器型号为 Intel (R) Xeon (R) Platinum 8255C,操作系统为 Ubuntu20.04,内存达到 43GB, GPU 型号为 NVIDIA GeForce RTX 3080。

在对模型进行评估时,我们选择了 mIoU、mAP 和准确率这几个指标来评估模型的性能。分类问题的结果通常只有两类:正例(Positive)和负例(Negative)。模型的预测结果与真实值分类为四类:真正例(TP)、假正例(FP)、真负例(TN)和假负例(FN)。其中,TP 表示实际值和预测结果都是正例的样本数,FP 表示实际值是负例但预测结果是正例的样本数,TN 表示实际值和预测结果都是负例的样本数,FN 表示实际值是正例但预测结果是负例的样本数。准确率表示正确分类的样本个数占总样本个数的比例,即

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)$$

mIoU 的公式为

$$mIoU = \frac{1}{K+1} \sum_{j=0}^K \frac{p_{jj}}{\sum_{i=0}^K p_{ji} + \sum_{i=0}^K p_{ij} - p_{jj}} \quad (4)$$

式中,类别总数(包含背景类)为 $K+1$,真实值为 j 、预测结果为 i 的像素数为 p_{ji} 。假设假正表示真实值为 j 但被误分类为 i 的像素数

p_{ji} ;而假负表示真实值为 i 但被误分类为 j 的像素数 p_{ij} 。

mAP 是一个取值在 0 到 1 之间的指标,其值越接近 1 表示检测模型的性能越好,检测能力越强。mAP 的公式如下:

$$mAP = \frac{\sum P_A}{N_C} \quad (5)$$

式中, N_C 为类别数量, P_A 为各类别的平均精度。

3.3 实验结果分析

为验证本文算法的有效性,我们在实验中选择了 DeeplabV3+^[18]、U-Net^[19]、HRNet^[20]、PSPNet^[21]等四种经典的图像提取模型作为对比方法。我们从定性与定量两个方面来评价本文方法的性能。在实验过程中,将采用的 Batch Size 统一设置为 2, Epoch 统一为 100,学习率统一设置为 0.0001。为了保证模型训练的精度和运行时间,我们选择 Adam 作为网络训练的优化器。部分样本的提取结果如图 5 所示。

由图 5 可知,本文方法在红外刑侦目标提取方面的表现优于其他模型。为了更加全面、准确地评估本文所提出的改进模型的性能,我们对不同模型的评价指标进行了比较分析(结果见表 1)。在 HRNet 模型中,18、32 和 48 分别代表不同宽度的模型版本,反映了网络层次的通道数。从表 1 中可以看出,相较于 DeeplabV3+、U-Net、HRNet、PSPnet 等模型,本文所提出的模型在 mIoU、mAP、准确率等评价指标上均表现出优异的性能,说明 BCES-Net 模型在热红外刑侦目标提取方面能够实现较为准确的结果。本文模型的 mIoU 值得到了有效提升(达到 88.3%),mAP 值达到 94.35%,准确率达到 95.5%。因此,我们可以得出以下结论:本文所提出的 BCES-Net 模型在准确提取红外刑侦目标方面具有一定的优势。这也为相关的实际应用提供了新的技术支持。

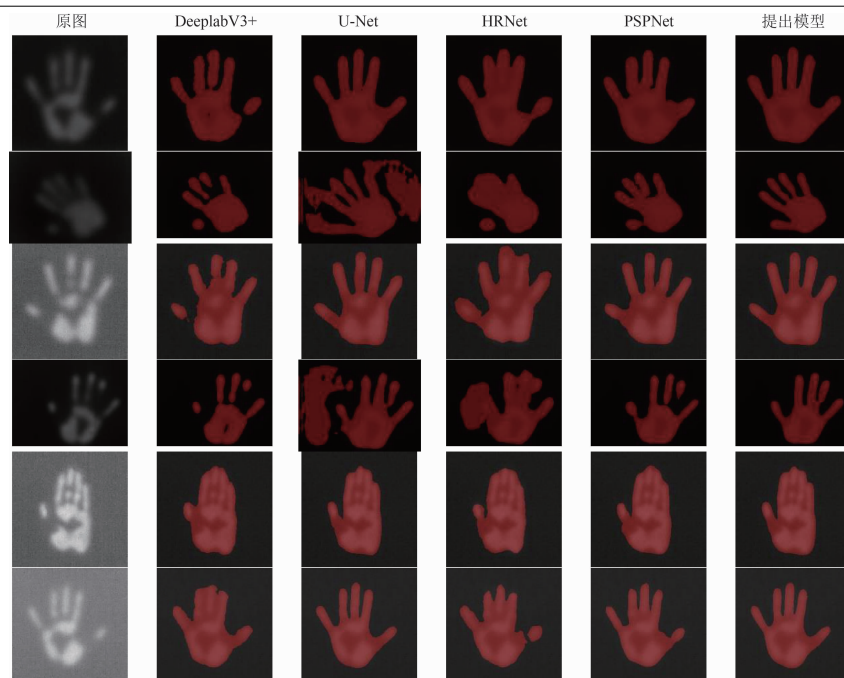


图 5 对比实验结果

表 1 不同深度神经网络模型在手部热痕迹数据集上的对比实验结果

模型	骨干网络	mIoU/%	mAP/%	准确率/%
DeeplabV3+	MobileNet	83.96	90.33	93.84
DeeplabV3+	Xception	82.05	88.5	83.12
U-Net	ResNet50	75.53	90.52	88.73
HRNet	HRNetv2_w18	85.76	92.45	94.47
HRNet	HRNetv2_w32	86.65	93	94.79
HRNet	HRNetv2_w48	77.58	89.67	90.26
PSPNet	MobileNet	82.02	88.82	93.05
PSPNet	ResNet50	83.65	89.93	93.37
提出模型	否	88.3	94.35	95.5

4 结束语

本文提出了一种基于特征增强的模糊刑侦目标提取算法,展示了深度学习技术在刑侦领域的潜在价值。该算法成功地克服了红外图像中目标边缘模糊和轮廓不完整等挑战,实现了对模糊图像中刑侦目标的高精度提取。从理论角度看,这项研究为图像处理领域提供了重要的指导,并为进一步研究奠定了基础。在实际应用中,该算法具有广泛的潜在价值,可用于犯罪调查、安全监控等领域,提高了刑侦工作的效率和准确性。与现有方法相比,BCES-Net 算法表现出了明显的改进,尤其是在手部热痕迹数据集上。然而仍需应对未来的挑战,

包括提高算法的稳健性和泛化能力,以适应更复杂的刑侦图像场景。此外,将该算法整合到实际刑侦系统中可能是一个有潜力的发展方向,从而更好地满足刑侦工作的实际需求。因此,未来的研究方向包括进一步改进算法性能、探索新的应用场景,并积极适应刑侦领域的实际需求,不断提高图像分析和处理的水平,为社会安全和犯罪调查提供更强大的工具和方法。

参考文献

- [1] Akula A, Ghosh R, Kumar S, et al. Moving target detection in thermal infrared imagery using spatiotemporal information [J]. *Journal of the*

- Optical Society of America A: Optics and Image Science and Vision*, 2013, **30**(8): 1492–1501.
- [2] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition [J]. *arXiv*: 1409.1556, 2014.
- [3] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition [C]. Las Vegas: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.
- [4] Howard A G, Zhu M, Chen B, et al. Mobilenets: efficient convolutional neural networks for mobile vision applications [J]. *arXiv*: 1704.04861, 2017.
- [5] Tan M, Le Q. Efficientnet: rethinking model scaling for convolutional neural networks [J]. *arXiv*: 1905.11946, 2019.
- [6] Chen L C, Papandreou G, Kokkinos I, et al. Deeplab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, **40**(4): 834–848.
- [7] Hao Y, Liu Y, Wu Z, et al. Edgeflow: achieving practical interactive segmentation with edge-guided flow [C]. Montreal: 2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW), 2021.
- [8] Yu Y, Yang Y, Liu K. Edge-aware superpixel segmentation with unsupervised convolutional neural networks [C]. Anchorage: 2021 IEEE International Conference on Image Processing (ICIP), 2021.
- [9] Zhang D, Zheng R. Trianglenet: edge prior augmented network for semantic segmentation through cross-task consistency [J]. *arXiv*: 2210.05152, 2022.
- [10] Liu H, Feng Y, Xu H, et al. MEA-Net: multilayer edge attention network for medical image segmentation [J]. *Scientific Reports*, 2022, **12**(1): 7868.
- [11] Chen L C, Zhu Y, Papandreou G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation [C]. Munich: 15th European Conference on Computer Vision (ECCV), 2018.
- [12] Zhao H, Shi J, Qi X, et al. Pyramid scene parsing network [C]. Honolulu: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.
- [13] 任莎莎, 刘琼. 小目标特征增强图像分割算法 [J]. *电子学报*, 2022, **50**(8): 1894–1904.
- [14] Huang Z, Wang X, Huang L, et al. CCNet: criss-cross attention for semantic segmentation [C]. Seoul: 2019 IEEE/CVF International Conference on Computer Vision (ICCV), 2019.
- [15] Li X, Zhong Z, Wu J, et al. Expectation-maximization attention networks for semantic segmentation [C]. Seoul: 2019 IEEE/CVF International Conference on Computer Vision (ICCV), 2019.
- [16] Zhong Z, Lin Z Q, Bidart R, et al. Squeeze-and-attention networks for semantic segmentation [C]. Seattle: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020.
- [17] Takikawa T, Acuna D, Jampani V, et al. Gated-SCNN: gated shape CNNs for semantic segmentation [C]. Seoul: 2019 IEEE/CVF International Conference on Computer Vision (ICCV), 2019.
- [18] Chen L C, Papandreou G, Schroff F, et al. Rethinking atrous convolution for semantic image segmentation [J]. *arXiv*: 1706.05587, 2017.
- [19] Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation [C]. Munich: Medical Image Computing and Computer-Assisted Intervention (MICCAI), 2015.
- [20] Wang J, Sun K, Cheng T, et al. Deep high-resolution representation learning for visual recognition [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020, **43**(10): 3349–3364.
- [21] Kingma D P, Ba J. Adam: a method for stochastic optimization [J]. *arXiv*: 1412.6980, 2014.