

基于卷积神经网络的现场勘查照片分类方法

李卓容¹, 唐云祁^{1*}, 蔡能斌²

¹中国人民公安大学侦查学院, 北京 100038;

²上海市刑事科学技术研究院重点实验室, 上海 200083

摘要 随着人工智能技术的迅速发展与广泛应用, 智能化勘查方法正成为刑事科学技术领域新的研究热点, 而实现现场勘查照片自动识别与分类是智能化勘查的重要研究内容。面向公安机关实战应用需求, 提出了一种基于卷积神经网络的现场勘查照片自动分类算法。基于真实案件照片, 建立了现场勘查照片数据集, 包含现场勘查照片 13164 张, 负类照片 4008 张。根据现场勘查照片数据特性, 设计了现场勘查照片分类网络 (CriSNet), 通过对卷积层增加归一化处理以及改进 bottleneck 模块, 实现对现场勘查照片的精确分类。实验结果表明: CriSNet 模型的分类精度优于基准网络 1 个百分点, 具有较好的鲁棒性, 同时在分辨率低、品质较差的情况下, 仍能保持较好的分类性能。

关键词 图像处理; 卷积神经网络; 现场勘查照片; 图像分类

中图分类号 TP317.4

文献标志码 A

DOI: 10.3788/LOP212827

Method for Classifying Crime Scene Photographs Based on Convolution Neural Network

Li Zhuorong¹, Tang Yunqi^{1*}, Cai Nengbin²

¹School of Criminal Investigation, People's Public Security University of China, Beijing 100038, China;

²Shanghai Key Laboratory of Crime Scene Evidence, Shanghai 200083, China

Abstract With the rapid development and wide application of artificial intelligence, intelligent investigation is becoming a new research hotspot in forensic science, and the realization of automatic recognition and classification of crime scene photographs is an essential aspect of intelligent investigations. We present an algorithm that automatically classifies crime scene photographs based on a convolution neural network. First, based on the data from criminal cases, a crime scene photograph dataset was constructed comprising 13164 scene photographs and 4008 negative photographs. Second, crime scene photograph net (CriSNet) was designed based on the data characteristics to accurately classify crime scene photographs by adding normalization processing to the convolution layer and improving the bottleneck module. The experimental results show that the accuracy of CriSNet is 1 percentage point better than that of the benchmark with good robustness, and CriSNet can still maintain excellent performance under low resolution and poor-quality conditions.

Key words image processing; convolution neural network; crime scene photo; image classification

1 引言

近年来, 基于人工智能的智能化勘查方法成为刑事科学技术领域新的研究热点。其以自动化或智能化的现场勘查方法为研究目标, 涉及现场摄影、痕迹物证发现及提取、检验鉴定等各方面。现场勘查照片是犯罪现场摄影的结果, 用于记录原始犯罪现场的基本情

况, 是犯罪现场最直观反映, 能够为未参与现场勘查的侦查人员以及法官还原案情。

现场勘查照片自动分类具有重要的研究价值。一方面, 现场勘查照片自动分类是实现犯罪现场智能摄影的先决条件。现场摄影工作是现场勘查的起始工作, 需最大限度地降低摄影人员对犯罪现场的人为破坏。因此, 智能化的无人摄影是犯罪现场勘查最理想

收稿日期: 2021-10-28; 修回日期: 2021-11-22; 录用日期: 2021-12-21; 网络首发日期: 2022-01-11

基金项目: 公安部技术研究计划项目(2020JSYJC21)、中央高校基本科研业务费项目(2021JKF203)、上海市现场物证重点实验室开放课题基金(2021XCWZK04)

通信作者: *tangyunqi@ppsuc.edu.cn

的选择。实现现场勘查照片自动分类是实现智能化犯罪现场无人摄影的前提。另一方面,现场勘查照片自动识别与分类算法不受社会其他领域关注,属内发的交叉融合问题,有利于促进刑事科学技术与人工智能交叉融合深度发展^[1-8]。

目前,现有智能化勘查方面的成果主要包含痕迹物证的搜索与分类^[1-3,9]和现场拍照与重建^[9-15]两方面。Li等^[1-3]使用卷积神经网络实现室内指纹定位及自动搜索。Bucheli^[10]提出一种基于可见光和图像的方法,对现场证据进行收集。Lega^[11]提出一个通过航拍图像获取现场照片及其他证据的环境犯罪案件数据采集系统。Tijtgat^[12]提出一种基于 GPU 处理器的 NVIDIA Jetson TX2 模块无人机捕获图像的目标检测系统。Araújo^[13]提出基于无人机的自动化现场勘查系统,配置预训练的 YOLO-v3 网络从多个角度拍摄现场照片。文献^[14-15]也提出基于无人机以及航拍图像的自动化现场勘查照片拍摄方法,但过于追求缩短耗时而导致精度一般,所拍摄现场勘查照片需要进一步人工处理和分类。Wang^[16]提出一种基于卷积神经网络的现场平面图分类方法。

现场勘查照片自动分类与识别问题本质上是图像分类识别问题。卷积神经网络以其独特的卷积结构和自适应学习方法突破了传统方法在图像识别方面的许多瓶颈^[29,32]。卷积神经网络发展至今,在图像处理领域取得了大量成果。AlexNet^[17]是第一个现代深度卷积神经网络模型。VGG^[18]网络的创新在于首次在卷积层组合使用更小的过滤器进行特征提取。GoogLeNet^[19]的核心思想是通过不垂直堆砌层的方法得到更深的网络。ResNet^[20]通过 bottleneck 结构在输入、输出之间建立了一条直接的关联通道,提升了信息

在网络中的传播效率,智能化的深入应用使得轻量级网络成为近年来的研究热点。MobileNet^[21-22]通过深度可分离卷积以及加入线性 bottleneck 和反转残差模块提高了计算效率。EfficientNet^[23]从平衡分辨率、加深网络和加宽网络 3 个维度,实现网络在效率和准确率上的优化。

本文通过人工筛选分类,构建包含 3676 张现场方位照片、4656 张现场概貌照片、4832 张现场重点部位照片及 4008 张非现场勘查照片的数据集。针对现场勘查照片数据特性,构建基于卷积神经网络的现场勘查照片分类网络 (Crime scene photos network, CriSNet),与基线网络相比,CriSNet 具有良好的分类性能和抗干扰能力,在分辨率低、品质损失较大情况下,仍能达到较好精度。

2 问题分析

我国刑事科学技术领域将现场勘查照片划分为 4 个类别,分别是现场方位照片、现场概貌照片、现场重点部位照片与现场细目照片^[24]。国际法庭科学领域认为细目照片并非直接与案情相关。此外,公安部于 2015 年提出的“一长四必”现场勘查新机制^[25]也未对现场细目照片做出具体要求。因此,这里采用国际标准将现场勘查照片划分为 3 个类别,并应用深度学习理论研究现场勘查照片自动识别与分类算法。现场方位照片用于记录犯罪现场所处位置及其与周围环境的关系;现场概貌照片用于记录现场全貌及内部各部分之间关系^[26];现场重点部位照片用于记录犯罪现场重要部位的状况、特点以及与犯罪行为有关痕迹、物品及所在位置^[26-28]。3 种类型的典型照片如图 1 所示。



图 1 数据集展示。(a)现场方位照片;(b)现场概貌照片;(c)现场重点部位照片

Fig. 1 Display of dataset. (a) Orientation; (b) outline; (c) key part

现场勘查照片分类问题具备较强的行业特性,不能简单归类为图像场景分类问题。同一类别照片包含的场景及元素极其复杂,各类别照片的类内差

异较大,类间差异较小。现实生活中犯罪现场复杂多变,如图 2 所示。图 2(a)~图 2(c)为某些案件中的现场概貌照片,涉及的场景分别为宾馆大厅、工地



图 2 现场勘查照片。(a)宾馆大厅的概貌照片;(b)工地宿舍的概貌照片;(c)烟酒便利店的概貌照片;(d)重点部位照片

Fig. 2 Crime scene photos. (a) Outline of the hotel lobby; (b) outline of the dormitory; (c) outline of the convenience store; (d) key part

宿舍、烟酒便利店,图 2(d)为某一案件中的现场重点部位照片。由图 2 可知,受房间大小和元素布局的影响,同一类别的现场勘查照片间差异较大,但抛开案件背景信息,就照片本身而言,现场重点部位照片和概貌照片难以区分,图 2(c)和图 2(d)可以是现场重点部位照片,也可以是概貌照片。

由于现场勘查照片类别边界模糊,其分类问题区别于普通场景分类问题。不能简单根据大场景、中场景、小场景加以区分,须联系图片背景上下文内容及场景纵深信息,进一步学习图像重点与类别之间的关系。同时,通过对卷积层增加归一化处理以及改进 bottleneck 模块,学习鲁棒的图像纹理特征的同时联系上下文背景,实现对现场勘查照片高精度的分类,并且使其具有良好的抗干扰能力。

3 CriSNet 结构

3.1 网络模型结构

针对现场勘查照片分类问题特性,设计 CriSNet,该网络结构如图 3 所示。在第 1 个卷积层后引入改进模块,后接 9 个 inception 模块,分别在第 2 个、第 7 个 inception 模块后连接一个最大池化层,并在最后一个 inception 模块后连接一个平均池化层,最后连接包含 4 个神经元的全连接层,使用 Softmax 分类器输出概率结果。现场勘查照片自动分类问题虽然数据特征较为复杂,但是类别较少,因此仅采用 inceptionV1 结构。网络具体结构及参数如表 1 所示。表 1 中仅展示卷积层、池化层和全连接层。卷积层均采用 ReLU 激活函数,初始化方式为“Xavier”。

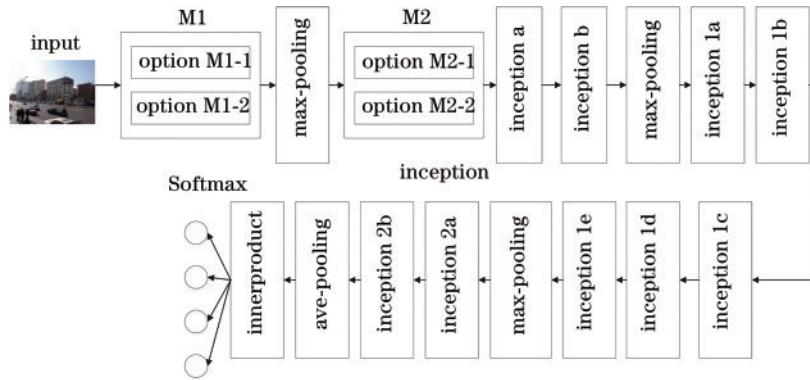


图 3 CriSNet 结构

Fig. 3 Structure of CriSNet

表 1 CriSNet 结构参数

Table 1 Parameters of CriSNet

Type	Number	Kernel size / stride	Output size
Convolution/BN/Scale	1	7×7/S2	[64, 112, 112]
Max-pooling	1	3×3/S2	[64, 56, 56]
Bottleneck	1	1×1/S1 3×3/S1 1×1/S1 1×1/S1	[256, 56, 56]
Inception a	2	3×3/S1 5×5/S1 1×1/S1	[480, 56, 56]
Max-pooling	1	3×3/S2	[480, 28, 28]
Inception b	5	1×1/S1 3×3/S1 5×5/S1 1×1/S1	[832, 32, 32]
Max-pooling	1	3×3/S2	[832, 14, 14]
Inception c	2	1×1/S1 3×3/S1 5×5/S1 1×1/S1	[1024, 14, 14]
Ave-pooling	1	7×7/S1	[1024, 8, 8]
Innerproduct	1	-	[1, 4]

图 3 中 M1、M2 为 2 个选择性优化模块, 经过该模块时, 仅能选择其中的一个优化选项, 若没有优化选项被选择, 则取消该模块。M1 通过对第 1 个卷积层进行归一化处理达到优化效果, 其中包括 option M1-1 和 option M1-2 两个优化选项; M2 对 bottleneck 模块进行优化调整, 其中包括 option M2-1 和 option M2-2 两个优化选项, 具体结构如图 4 所示。

1) Option M1-1 与 Option M1-2 优化选项为局部响应归一化(Local response normalization, LRN)层与

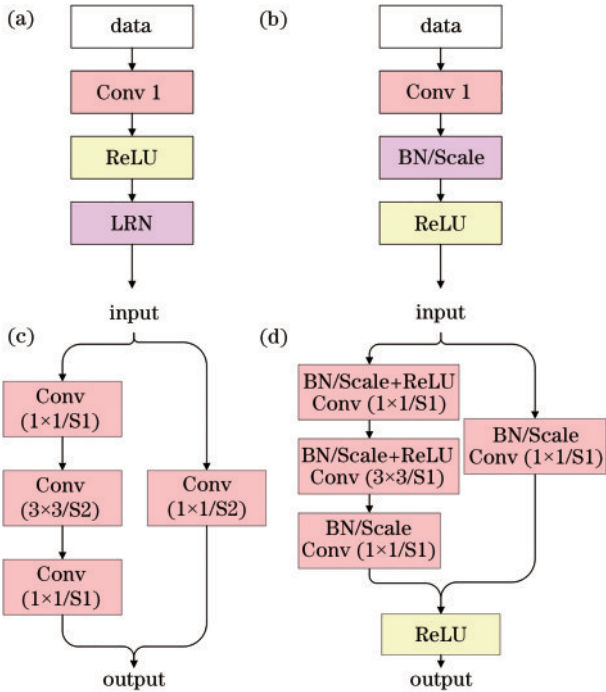


图 4 优化选项具体结构。(a) Option M1-1; (b) Option M1-2; (c) Option M2-1; (d) Option M2-2

Fig. 4 Specific structure of optimization options. (a) Option M1-1; (b) Option M1-2; (c) Option M2-1; (d) Option M2-2

批归一化处理(Batch normalization, BN)层如图 4(a)、(b)所示。对网络结构的第 1 个卷积层分别采取两种归一化方式, 局部响应归一化操作应用在激活函数 ReLU 之后, 批归一化操作应用在激活函数 ReLU 之前。LRN 层对当前单个训练样本所有神经元的输入进行计算, 利用统一的均值和方差进行归一化; BN 层针对单个神经元进行归一化操作, 近似给神经网络引入噪声, 提高了泛化性。

2) Option M2-1 与 Option M2-2 优化选项为 bottleneck 模块的两种优化如图 4(c)、(d)所示。

① 将下采样操作从第 1 个卷积层放到第 3 个卷积层。下采样操作放在步长为 2 的卷积层时, 会丢失 3/4 特征信息, 当下采样操作放在第 3 个卷积层时可以减少损失, 3×3 的卷积核可以保留更多有效信息, 如图 4(c)所示。

② 将 BN 层与 ReLU 操作提前至权重层之前, 作为权重层的预激活处理, 构建对称的激活函数, 使残差块通路与输出同时被激活。当残差块数量较多时, 这种结构会导致 ReLU 函数截断残差块之间的连接, 由此导致精度下降, 但当残差块数量较少时, 通过网络权重更新即可达到健康状态, 如图 4(d)所示。

3.2 Bottleneck 预激活层

在原始残差模块中, 激活函数 $x_{i+1} = f(x_i)$ 对恒等映射和下一个残差单元均会产生影响。文献[30]通过构建不对称激活函数, 如图 5(b)所示, 将 BN 层与 ReLU 重塑为预激活层, 使模型具有更强的泛化能力。

由于 $h(x_i) = x_i$, 残差网络的第 L 层可表示为

$$x_L = x_i + \sum_{i=i}^{L-1} F(x_i, W) \quad (1)$$

显然, 第 L 层是输入与残差之和, \mathcal{E} 为损失函数, 通过反向传播可以得到

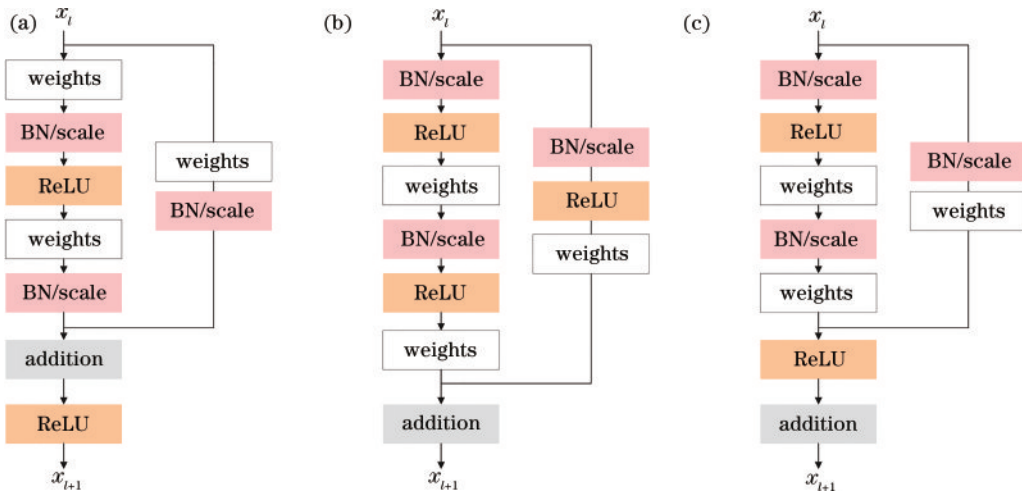


图 5 3 种不同残差块结构。(a)原始残差块; (b)文献[30]优化残差块; (c)本文提出的预激活残差块

Fig. 5 Three different kinds of residual block structures. (a) Original residual blocks; (b) pre-activated residual blocks optimized by Ref. [30]; (c) proposed pre-activated residual blocks in this paper

$$\frac{\partial \mathcal{E}}{\partial x_i} = \frac{\partial \mathcal{E}}{\partial x_L} \frac{\partial x_L}{\partial x_i} = \frac{\partial \mathcal{E}}{\partial x_L} \left(1 + \frac{\partial}{\partial x_i} F(x_i, W_i) \right). \quad (2)$$

$\frac{\partial}{\partial x_i} F(x_i, W_i)$ 不会恒等于 -1 , 此时梯度从 L 层传

递至 1 层不会消失, 深层网络训练效果从而得到改善。残差模块的核心在于恒等映射 $h(x) = x$, 即构建通路让两个层可以直接传递信息。而原始的残差模块由于需要经过 ReLU, 并没有直接从第 1 层到第 $l+1$ 层的通路, 因此文献[30]构建不对称的 ReLU 函数, 并将后激活处理改为预激活处理, 在深层残差网络中获得较好效果。

CriSNet 结构中, 同样将 bottleneck 模块的 BN 层与 ReLU 提前作为预激活层, 如图 5(c) 所示。不同的是, 对于 bottleneck 结构中的最后一个卷积层, 仍采用对称的 ReLU 处理, 即对通路和输出值都进行激活。由于 CriSNet 结构仅连接一个 Bottleneck 模块, 在网络训练过程中通过权重调整, 可使激活层输出值频繁大于 0, 因此从第 1 层到第 $l+1$ 层的通路没有被激活函数截断, 同时预激活层处理使网络表现出更好的性能。

4 实验结果及分析

4.1 实验数据

由于现场勘查照片属于涉密信息, 缺少开源数据集, 因此需首先建立现场勘查照片数据集。利用从全国公安机关现场勘验信息系统(简称“现勘系统”)中导出的来自 12 个省份的真实案件现场勘查照片, 构建现场勘查照片数据集。该数据集以盗窃、抢劫、诈骗类案件为主, 共计 17172 幅现场勘查照片, 其中现场勘查照片 13164 张, 分为现场方位照片、现场概貌照片和现场重点部位照片 3 类。另外, 因真实案件中技术人员有时将现场平面图错误上传至现场勘查照片分类, 选取 4008 张现场图作为负样本。按照 8:1:1 随机分为训练集、验证集、测试集, 验证集共有照片 1750 张, 测试集共有照片 1700 张, 具体数据集分布情况如表 2 所示。

表 2 数据集构成
Table 2 Composition of the dataset

Classification	Orientation	Outline	Key part	Negative sample
Training set	2926	3756	3832	3208
Verification set	400	450	500	400
Test set	350	450	500	400
Summary	3676	4656	4832	4008

4.2 实验环境及设置

实验使用操作系统为 Ubuntu 18.04 的台式计算机, 处理器为 Inter(R) Core(TM) i7-7800X@3.50 GHz, 电脑内存为 16 GB, 显卡型号为 GeForce RTX 2080Ti。实验使用 Caffe^[31] 框架, Epoch 设置为 30, Gamma 值为 0.1, 权值衰减为 0.00005, 基础学习率设

置为 0.001, 学习率下降策略为“step down”, 优化器使用随机梯度下降(Stochastic gradient descent, SGD)。输入数据为 224×224 的 RGB 三通道图像数据, 对网络进行训练。

4.3 评价标准

分类模型常见评价指标有混淆矩阵、受试者工作特征曲线(Receiver operating characteristic curve, ROC)和曲线下面积值(Area under the curve of ROC, AUC)3 种, 用以评价模型分类性能及拟合度。由混淆矩阵衍生二级指标准确率(Accuracy)、精确率(Precision)、召回率(Recall)和三级指标 F1 得分。由于本文数据集应用场景的特殊性, 提出“错误分类置信度分布”指标。

1) 准确率、精确率、召回率、F1 得分。准确率针对整个模型, 评估分类模型所有判断正确结果占总观测值的比重, 计算公式为 $A = \frac{1}{n} \sum_{i=1}^n I(f(x_i) = y_i)$; 精确率针对模型所有预测正例结果中正确预测占比, 计算公式为 $P = \frac{f_{TP}}{f_{TP} + f_{FP}}$; 召回率针对所有真实值为正例的结果中模型正确预测占比, 计算公式为 $R = \frac{f_{TP}}{f_{TP} + f_{FN}}$; 一般来说, 精确率与召回率是一对矛盾变量, F1 得分综合精确率与召回率指标, 计算公式为 $F_1 = \frac{2R \cdot P}{R + P}$ (f_{TP} 为将正类预测为正类数; f_{FP} 为将负类预测为正类数; f_{FN} 为将负类预测为负类数)。

2) ROC 及 AUC 值。ROC 经常被用作分类问题最重要的评价指标之一, 能够反映模型的判断能力及阈值对分类器泛化性能的影响, 同时可以通过曲线形状直观比较分类器性能优劣。ROC 形状基本不受正负样本分布影响, 能够降低测试集带来的干扰, 更客观衡量模型本身性能。AUC 值是 ROC 下方的面积, 同时考虑分类器对正负样本的分类能力, 能够合理有效对分类模型做出评价, 其计算公式为 $f_{AUC} = \frac{1}{2} \sum_{m=1}^{i-1} (x_{i+1} - x_i)(y_i + y_{i+1})$ 。

3) 错误分类置信度分布。为进一步衡量模型性能, 提出了一种新颖的评价指标“错误分类置信度分布”, 即网络对于分类错误图片的打分分布情况。受限于数据集大小、数据特异性等因素, 网络模型对于某些图片无法正确分类, 这里认为如果网络模型对错误分类图片的打分大部分在 0.9 以上, 那么其性能差于对错误分类图片打分较低的网络模型。

由于刑事案件的严谨性与特异性, 在应用自动识别与分类算法的基础上, 技术人员会再次对类别模糊的图片数据进行人工确认, 如果网络以高置信度错误分类图片, 有可能对技术人员的判断造成影响。因此, 衡量网络对错误分类图片是否给出过高预测分数十分

必要。通过对模型“错误分类置信度分布”的可视化,进一步直观评估模型性能。

4.4 实验结果及分析

4.4.1 性能分析

CriSNet 的两个选择性优化模块精度及 F1 得分如表 3 所示,同时按照文献[30]方法对 bottleneck 模块进行优化。

对比方案(a)与方案(b)可知,对第 1 个卷积层增加批归一化处理有助于网络有效学习,精度以及 F1 得分均有所上升,局部响应归一化效果不佳。由于 CriSNet 网络结构仅采用 1 个 bottleneck 模块,所以对

于下采样卷积层顺序的调整减少了参数量与训练耗时,但精度一般。对比方案(d)与文献[30]方法可知,由于 bottleneck 模块较少,构建非对称激活函数并未达到较好优化效果,而较少的残差模块同时使对称激活函数没有频繁截断网络层级间通路,通过权重的调整,短暂震荡之后即可达到健康状态。对比方案(b)与方案(c)可知,虽然方案(c)降低了参数量,大幅缩短了训练耗时,但是精度和 F1 得分均不如方案(b)。对比方案(b)与方案(d)可知,将 BN 层与 ReLU 激活函数作为权重层的预激活处理有助于网络性能提升,精度提升 0.3 个百分点,网络性能更优。

表 3 不同优化方案结果

Table 3 Results of different optimization schemes

Scheme	Method	Time /min	Accuracy /%	F1 Score /%
Scheme (a): Conv/LRN	Option M1-1	61	93.00	92.87
Scheme (b): Conv/BN	Option M1-2	57	93.53	93.11
Scheme (c): 3×3 Conv/S2	Option M1-2+Option M2-1	28	93.32	92.96
Scheme (d): proposed	Option M1-2+Option M2-2	52	93.83	94.16
Ref. [30]	-	61	92.86	92.72

为更好评价本文网络性能,针对系列经典网络做了对比实验,包括 AlexNet^[19]、VGGNet^[20]、GoogLeNet^[21]、ResNet^[22]、MobileNet^[24]、EfficientNet^[25]。由于现场勘查照片数据集类别较少,且规模难以满足网络层数过多的深度网络的训练要求,经过预实验,选择 AlexNet、VGG16、GoogLeNet-V1、ResNet50、MobileNet-V3、

EfficientNet-B0 作为基线网络。特别地,为使网络更好收敛,将 ResNet50 网络所有卷积层初始化方式替换为“Xavier”初始化,使用带有 SE (Squeeze-and-excitation) 结构,且激活函数为 ReLU 的 EfficientNet-B0 网络模型。精确率、准确率、召回率、F1 得分及参数量指标如表 4 所示。

表 4 基线网络模型对比

Table 4 Comparision of the baseline network model

Parameter	AlexNet /%	VGGNet /%	GoogLeNet /%	ResNet /%	MobileNet /%	EfficientNet /%	CriSNet /%
Accuracy	91.53	90.82	92.06	92.34	91.34	92.75	93.83
Precision	91.88	91.88	92.51	92.46	92.26	93.14	94.01
Recall	92.46	91.70	92.97	93.36	92.45	92.72	94.38
F1 Score	92.12	91.75	92.70	92.77	92.34	92.92	94.16

由表 4 可知,CriSNet 的准确率高于基线网络最佳精度 (EfficientNet) 1.08 个百分点,同时准确率及召回率均能保持较好数值,综合 F1 得分,CriSNet 模型性能更佳。不同网络结构测试精度如图 6 所示。

根据 CriSNet 测试结果生成混淆矩阵,如图 7 所示。由图 7 可知,网络对于概貌照片和重点部位照片难以区分,而对于方位照片分类效果则较好。采用绘制 ROC 方法对基线网络及 CriSNet 模型性能进行评估,如图 8 所示。通过直观对比及计算 AUC 值,可以看出 CriSNet 性能优于基线网络。

采用“错误分类置信度分布”评价指标对 CriSNet 及基线网络模型进行分析,具体分布数据如表 5 所示。CriSNet 对于现场勘查照片自动分类任务的学习更佳,对于大部分错误分类照片并不够自信,在一定程度上可以减小后期技术人员进一步人工确认的工作量。

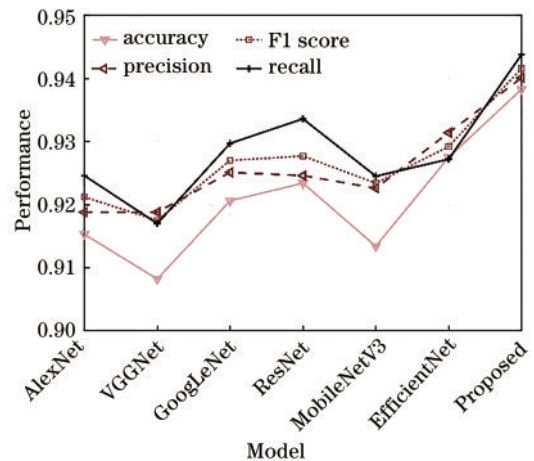


图 6 基线网络模型对比

Fig. 6 Comparison of baseline network model

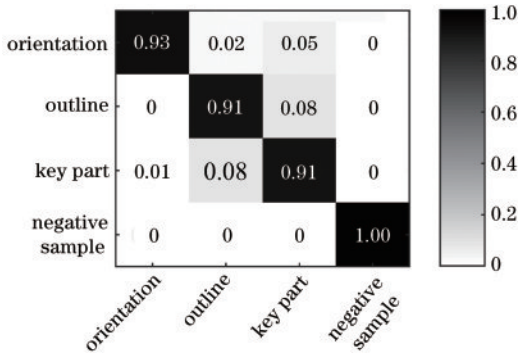


图 7 CriSNet混淆矩阵

Fig. 7 Confusion matrix of CriSNet

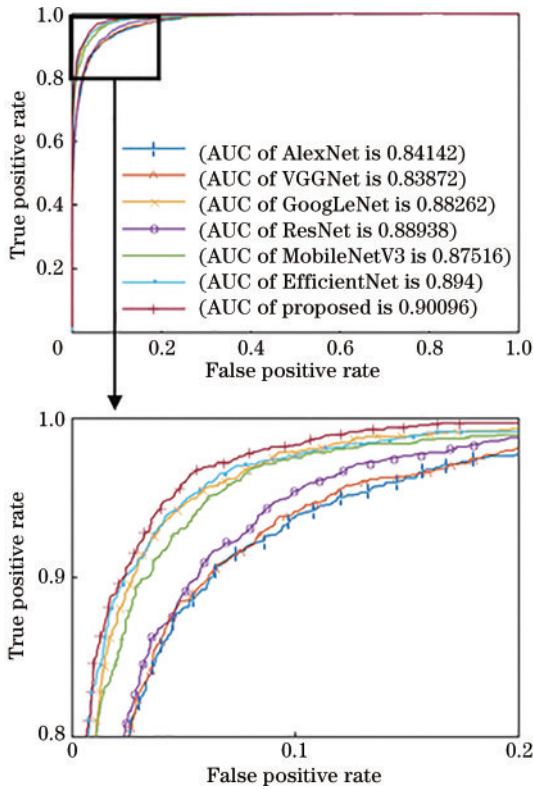


图 8 网络模型 ROC

Fig. 8 ROC of network model

分布可视化如图 9 所示。

4.4.2 鲁棒性分析

现场勘查照片多为尼康 D7000 相机拍摄,原始分辨率多为 1000×1000 以上,真实案件中由于拍摄设备不统一极有可能产生分辨率过大或过小的现场照片,为节省计算成本在上述实验中将训练集图片裁剪至

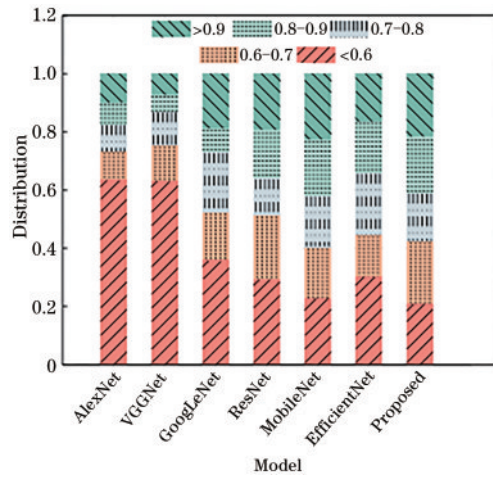


图 9 错误分类置信度分布柱状图

Fig. 9 Distribution histogram of misclassification confidence

224×224 大小。因此,增加训练集分辨率对分类精度影响测试,模拟真实案件可能情况。在一般认知中,分类精度会随着训练集分辨率的上升而上升,本实验仅对低分辨率情况进行测试,如表 6 所示。

实验结果表明,当分辨率下降至 96×96 时,CriSNet 精度虽有所下降,但仍比精度最好的基线网络(VGGNet)高 0.51 个百分点;当分辨率进一步下降为 48×48 时,所有基线网络准确率均下降至 90% 以下,CriSNet 准确率仍保持在 91.03%。CriSNet 模型对于不同分辨率训练集均能达到较好分类精度。

由于拍摄手法不正确等因素可能使拍摄现场照片质量不高,本实验以加入随机椒盐噪声模拟真实情况,分别使用随机加入的 3 种不同密度的椒盐噪声现场照片对网络模型进行测试,结果如表 7 所示。当密度为 0.01 和 0.1 时,基线网络精度均有明显下降,但对 CriSNet 精度影响不大;当密度为 0.5 时,各个网络模型分类精度均大幅度下降,EfficientNet 此时的测试精度最高,CriSNet 精度仍高于 ResNet 近 2.34 个百分点。

可以看出,轻微噪声以及品质损失对于 CriSNet 影响并不大,网络特征学习鲁棒性较好,当噪声明显、品质损失严重时,尽管精度下降明显,但仍优于基准网络。

通过分析训练集分辨率以及噪声对网络分类精度的影响,可以看出 CriSNet 模型抗干扰能力较强,网络性能较好,对于分辨率较小以及品质不佳的照片表现良好。

表 5 错误分类置信度分布

Table 5 Distribution of misclassification confidence

Scores	AlexNet / %	VGGNet / %	GoogLeNet / %	ResNet / %	MobileNet / %	EfficientNet / %	CriSNet / %
>0.9	63.89	63.46	36.30	29.55	23.13	30.50	21.29
0.8-0.9	09.72	12.18	16.30	21.97	17.16	14.40	21.29
0.7-0.8	09.03	11.54	20.74	12.88	17.91	21.18	16.67
0.6-0.7	07.64	05.77	08.15	16.67	19.40	17.79	19.44
<0.6	09.72	07.05	18.52	18.94	22.38	16.10	21.29

表 6 训练集分辨率对分类精度影响

Table 6 Influence of the resolution of training set on the accuracy of network

Resolution	Accuracy / %						
	AlexNet	VGGNet	GoogLeNet	ResNet	MobileNet	EfficientNet	CriSNet
48×48	84.40	89.83	88.51	88.86	78.53	88.75	91.03
96×96	89.20	92.06	90.26	91.60	85.04	90.98	92.57
224×224	91.53	90.82	92.06	92.34	91.43	92.75	93.83

表 7 噪声对网络分类精度影响

Table 7 Influence of noise on the accuracy of network

Noise density	Accuracy / %						
	AlexNet	VGGNet	GoogLeNet	ResNet	MobileNet	EfficientNet	CriSNet
0.01	91.09	91.14	92.11	91.43	91.50	92.50	93.60
0.10	88.69	90.17	91.03	91.20	90.25	92.00	93.60
0.50	76.00	75.83	76.34	82.29	76.25	86.50	84.63

4.4.3 错误分析

结合测试集分类错误情况研究分析了数据集图片

特点后,发现房间的装修风格、布局和光线可能是影响分类精度的因素,其中,典型分类错误图片如图 10 所示。



图 10 典型分类错误图片。(a) (b) (c) 现场概貌;(d) 现场重点部位

Fig. 10 Representative misclassification pictures. (a) (b) (c) Outline; (d) key part

当室内现场概貌照片的家具与房间背景色调一致时,如图 10(a)、图 10(b)所示,可能会误导网络将其判断为现场重点部位照片。同样,当房间光线较暗,使照片整体呈现相似的色调,如图 10(c)所示,或房间整体布局太过简单,都会导致网络将其误判为现场重点部位照片。特别地,当现场重点部位照片呈现出大片排列整齐的格子状花纹时,如图 10(d)所示,网络会非常自信地将其判断为现场方位照片。这些错误可能是由于卷积神经网络在对图片进行分类时,学习了图片的纹理和色彩信息导致的。进一步推测,对于现场方位照片,网络模型可能选择学习了建筑物照片的格子状纹理;对于现场概貌照片,网络模型可能选择学习了室内照片的色彩和条纹;对于现场重点部位照片,网络模型可能选择学习了物品照片的边缘信息。

图 10(a)~图 10(c)的正确类别应为现场概貌照片,图 10(d)的正确类别应为现场重点部位照片。6 个网络模型给出的预测类别及置信度如表 8 所示,分类错误图片及其置信度已加粗。随着网络加深和性能提升,色彩和光线的影响逐渐减弱,但是 6 个网络模型均以高置信度误判了图 10(d)所示图片,因此在目前网络结构中,纹理的影响可能大于色彩。

表 8 分类错误图片置信度

Table 8 Confidence of misclassification pictures

Network	Picture	Ground truth	Predict label	Score
AlexNet	Fig. 10(a)	Outline	Outline	0.9967
	Fig. 10(b)	Outline	Outline	0.9948
	Fig. 10(c)	Outline	Key part	0.8179
	Fig. 10(d)	Overview	Orientation	0.9999
VGGNet	Fig. 10(a)	Outline	Outline	0.9954
	Fig. 10(b)	Outline	Outline	0.5684
	Fig. 10(c)	Outline	Key part	0.7912
GoogLeNet	Fig. 10(d)	Overview	Outline	0.9999
	Fig. 10(a)	Outline	Outline	0.9954
	Fig. 10(b)	Outline	Outline	0.5684
	Fig. 10(c)	Outline	Outline	0.9916
ResNet	Fig. 10(d)	Overview	Orientation	0.9995
	Fig. 10(a)	Outline	Outline	0.9037
	Fig. 10(b)	Outline	Orientation	0.6594
	Fig. 10(c)	Outline	Key part	0.7885
	Fig. 10(d)	Overview	Outline	0.8860

表 8 (续)

Network	Picture	Ground truth	Predict label	Score
MobileNet	Fig. 10(a)	Outline	Outline	0.9934
	Fig. 10(b)	Outline	Key part	0.7860
	Fig. 10(c)	Outline	Outline	0.9975
	Fig. 10(d)	Overview	Orientation	0.9907
EfficientNet	Fig. 10(a)	Outline	Key part	0.8836
	Fig. 10(b)	Outline	Key part	0.8839
	Fig. 10(c)	Outline	Key part	0.8799
	Fig. 10(d)	Overview	Key part	0.8851
CriSNet	Fig. 10(a)	Outline	Outline	0.9728
	Fig. 10(b)	Outline	Outline	0.9941
	Fig. 10(c)	Outline	Outline	0.9287
	Fig. 10(d)	Overview	Orientation	0.9881

5 结 论

针对刑事技术领域现场勘查照片自动分类问题,在综合分析现场勘查照片数据特性基础上,提出了一种新的网络结构 CriSNet,实现现场勘查照片高精度自动分类。该方法具有良好的分类性能和抗干扰能力,精度优于基线网络 1 个百分点以上,准确率与召回率均高于基线网络近 1 个百分点。

该方法可对输入的所有现场勘查照片进行分类,对低分辨率、质量较差的照片鲁棒性强,更适用于实际应用场景。当分辨率下降至 48×48 时,基线网络精度均低于 90%,CriSNet 精度仍保持 91.03%,轻微噪声以及品质损失对于 CriSNet 影响不大;当噪声明显、品质损失严重时,尽管精度下降明显,但仍优于基线网络。CriSNet 的不足之处在于条纹特征对于网络分类性能影响较大。后续研究可以讨论如何削弱条纹特征对于现场勘查照片分类的影响并进一步扩大数据集以获得更好的性能。

参 考 文 献

- [1] 李新春,纪小璐,魏武,等.基于OCAE-SOM的室内指纹定位算法研究[J].激光与光电子学进展,2021,58(8):0812001.
Li X C, Ji X L, Wei W, et al. Research on fingerprint location algorithm based on OCAE-SOM[J]. Laser & Optoelectronics Progress, 2021, 58(8): 0812001.
- [2] 许浩,王旭东,吴楠.基于卷积神经网络的室内可见光指纹定位方法[J].激光与光电子学进展,2021,58(17):1706008.
Xu H, Wang X D, Wu N. Indoor visible light fingerprint positioning scheme based on convolution neural network [J]. Laser & Optoelectronics Progress, 2021, 58(17): 1706008.
- [3] 曹燕平,李晓记,胡云云.基于可见光指纹的室内高精度定位方法[J].激光与光电子学进展,2019,56(16):160601.
Cao Y P, Li X J, Hu Y Y. Visible light fingerprint-based high-accuracy indoor positioning method[J]. Laser &

Optoelectronics Progress, 2019, 56(16): 160601.

- [4] 张红颖,杨维民,王汇三.基于优化三维变形模型参数的人脸重建方法[J].激光与光电子学进展,2021,58(20):2015008.
Zhang H Y, Yang W M, Wang H S. Face reconstruction method based on optimized three-dimensional morphable model parameters[J]. Laser & Optoelectronics Progress, 2021, 58(20): 2015008.
- [5] 高刘雅,孙冬,卢一相.基于轻量级注意机制的人脸检测算法[J].激光与光电子学进展,2021,58(2):0210010.
Gao L Y, Sun D, Lu Y X. Face detection algorithm based on a lightweight attention mechanism network[J]. Laser & Optoelectronics Progress, 2021, 58(2): 0210010.
- [6] 姚丽莎,徐国明,赵凤.基于卷积神经网络局部特征融合的人脸表情识别[J].激光与光电子学进展,2020,57(4):041513.
Yao L S, Xu G M, Zhao F. Facial expression recognition based on local feature fusion of convolutional neural network[J]. Laser & Optoelectronics Progress, 2020, 57(4): 041513.
- [7] 刘美菊,曹永战,朱树云,等.基于卷积神经网络的特征融合视频目标跟踪方法[J].激光与光电子学进展,2020,57(4):041502.
Liu M J, Cao Y Z, Zhu S Y, et al. Feature fusion video target tracking method based on convolutional neural network[J]. Laser & Optoelectronics Progress, 2020, 57(4): 041502.
- [8] 高德勇,康自兵,王松,等.结合稠密轨迹与视频显著性特征的人体动作识别[J].激光与光电子学进展,2020,57(24):241003.
Gao D Y, Kang Z B, Wang S, et al. Human-body action recognition based on dense trajectories and video saliency [J]. Laser & Optoelectronics Progress, 2020, 57(24): 241003.
- [9] 于福升,余江,鲁远甫,等.基于残差网络的虹膜图像性别分类[J].激光与光电子学进展,2021,58(16):1610022.
Yu F S, Yu J, Lu Y F, et al. Gender classification of iris image based on residual network[J]. Laser & Optoelectronics Progress, 2021, 58(16): 1610022.
- [10] Bucheli S R, Pan Z G, Glennie C L, et al. Terrestrial laser scanning to model sunlight irradiance on cadavers under conditions of natural decomposition[J]. International Journal of Legal Medicine, 2014, 128(4): 725-732.
- [11] Lega M, Ferrara C, Persechino G, et al. Remote sensing in environmental police investigations: aerial platforms and an innovative application of thermography to detect several illegal activities[J]. Environmental Monitoring and Assessment, 2014, 186(12): 8291-8301.
- [12] Tijtgat N, van Ranst W, Volckaert B, et al. Embedded real-time object detection for a UAV warning system [C]//2017 IEEE International Conference on Computer Vision Workshops, October 22-29, 2017, Venice, Italy. New York: IEEE Press, 2017: 2110-2118.
- [13] Araújo P, Mendonça M, Fontinele J, et al. Towards autonomous investigation of crime scene by using drones [J]. Sensors & Transducers, 2019, 234(6): 30-36.
- [14] Cunha R R, Arrabal C T, Dantas M M, et al. Laser

- scanner and drone photogrammetry: a statistical comparison between 3-dimensional models and its impacts on outdoor crime scene registration[J]. *Forensic Science International*, 2022, 330: 111100.
- [15] Araújo P, Miranda R, Carmo D, et al. Air-SSLAM: a visual stereo indoor SLAM for aerial quadrotors[J]. *IEEE Geoscience and Remote Sensing Letters*, 2017, 14(9): 1643-1647.
- [16] 王凯旋, 李卓容, 王晓宾, 等. 刑事案件现场图自动分类算法[J]. *激光与光电子学进展*, 2020, 57(4): 041009. Wang K X, Li Z R, Wang X B, et al. Automated classification method for crime scene sketches[J]. *Laser & Optoelectronics Progress*, 2020, 57(4): 041009.
- [17] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks[J]. *Communications of the ACM*, 2017, 60(6): 84-90.
- [18] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[EB/OL]. (2019-08-23)[2021-05-06]. <https://arxiv.org/abs/1908.08681v1>.
- [19] Szegedy C, Liu W, Jia Y Q, et al. Going deeper with convolutions[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition, June 7-12, 2015, Boston, MA. New York: IEEE Press, 2015.
- [20] He K M, Zhang X Y, Ren S Q, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, 37(9): 1904-1916.
- [21] Howard A G, Zhu M L, Chen B, et al. MobileNets: efficient convolutional neural networks for mobile vision applications[EB/OL]. (2017-04-17)[2021-08-08]. <https://arxiv.org/abs/1704.04861>.
- [22] Sandler M, Howard A, Zhu M L, et al. MobileNetV2: inverted residuals and linear bottlenecks[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 4510-4520.
- [23] Tan M X, Le Q V. EfficientNet: rethinking model scaling for convolutional neural networks[EB/OL]. (2019-04-28)[2021-08-09]. <https://arxiv.org/abs/1905.11946>.
- [24] 廖翔. 刑事案件现场照相新特点及应对[J]. *警察技术*, 2007(4): 4-6.
- Liao X. New features and countermeasures of crime scene photography[J]. *Police Technology*, 2007(4): 4-6.
- [25] 王丽杰. “一长四必”现场勘查新机制相关问题探究[J]. *法制与经济*, 2016(5): 172-174.
- Wang L J. Research on the new mechanism of “one long and four necessities” site survey[J]. *Legal and Economy*, 2016(5): 172-174.
- [26] 牛学军. 道路交通事故现场勘查实务指南[M]. 北京: 中国人民公安大学出版社, 2013.
- Niu X J. Practical Guide to road traffic accident scene investigation[M]. Beijing: Chinese People's Public Security University Press, 2013.
- [27] 谢军, 王志群. 刑事影像技术[M]. 北京: 中国人民公安大学出版社, 2014.
- Xie J, Wang Z Q. Criminal imaging technology[M]. Beijing: Chinese people's Public Security University Press, 2014.
- [28] 徐远清, 黄晨. 公安科技概论[M]. 南昌: 江西高校出版社, 2009.
- Xu Y Q, Huang C. Introduction to public security science and technology[M]. Nanchang: Jiangxi Higher Education Press, 2009.
- [29] Li Z W, Yang W J, Peng S H, et al. A survey of convolutional neural networks: analysis, applications, and prospects[EB/OL]. (2020-04-01)[2021-08-06]. <https://arxiv.org/abs/2004.02806>.
- [30] He K M, Zhang X Y, Ren S Q, et al. Identity mappings in deep residual networks[M]//Leibe B, Matas J, Sebe N, et al. Computer vision-ECCV 2016. Lecture notes in computer science. Cham: Springer, 2016, 9908: 630-645.
- [31] Jia Y Q, Shelhamer E, Donahue J, et al. Caffe: convolutional architecture for fast feature embedding[J]. *MM 2014-Proceedings of the 2014 ACM Conference on Multimedia*, 2014: 675-678.
- [32] 鄢然, 廖记登, 吴小勇, 等. 基于卷积神经网络的砂石骨料分类方法研究[J]. *激光与光电子学进展*, 2021, 58(20): 2010015.
- Yan R, Liao J D, Wu X Y, et al. Research on classification method of sand and gravel aggregate based on convolutional neural network[J]. *Laser & Optoelectronics Progress*, 2021, 58(20): 2010015.