

doi: 10.16104/j.issn.1673-1891.2026.01.011

基于 ConvNeXt 网络的监控图像斑点特征识别方法

尹 金

(安徽工商职业学院信息工程学院, 安徽 合肥 230031)

摘要:为解决在复杂监控场景下对斑点特征识别准确率低、鲁棒性差的问题,提出一种基于 ConvNeXt 网络的监控图像斑点特征识别方法。采用视觉显著性检测模型精准定位监控图像中的斑点显著区域,并将其高效分割提取出来;将分割得到的斑点显著区域输入至基于深度可分离卷积机制改进的 ConvNeXt 网络中,利用该机制特有的高效计算特性,对斑点显著区域进行多尺度特征深度提取,充分挖掘不同尺度下的斑点特征信息;依据提取的多尺度斑点特征,借助 Softmax 激活函数完成对监控图像中斑点类型的准确识别,为监控图像斑点分析提供了精准的解决方案。结果表明:斑点显著区域划分+ConvNeXt 网络的平均质量缺陷指数(quality defect index, QDI)为 0.9~1.0,高于基础的 ConvNeXt 网络,证明斑点显著区域划分对应提升识别方法性能的作用。基于 ConvNeXt 网络的识别方法的平均 QDI 在 0.8~1.0 波动,高于基于支持向量机的识别方法、基于 HSV 空间融合 Retinex 算法的识别方法和基于多特征聚合的识别方法,证明了其在监控图像斑点特征识别任务中的准确率和鲁棒性。

关键词: ConvNeXt 网络; 监控图像; 斑点特征; 显著区域划分; 识别方法

中图分类号: TP391.41 **文献标志码:** A **文章编号:** 1673-1891(2026)01-0093-10

Spot Feature Recognition Method for Surveillance Images Based on ConvNeXt Network

YIN Jin

(School of Information Engineering, Anhui Vocational College of Industry and Commerce,
Hefei 230031, Anhui, China)

Abstract: Aiming at the problems of low accuracy and poor robustness in spot feature recognition under complex surveillance scenarios, a spot feature recognition method for surveillance images based on the ConvNeXt network is proposed. First, a visual saliency detection model is adopted to accurately locate the salient regions of spots in surveillance images, and segment and extract them efficiently. Then, the segmented salient regions of spots are input into the improved ConvNeXt network based on the depthwise separable convolution mechanism. Taking advantage of the unique efficient computing characteristics of this mechanism, deep extraction of multi-scale features is carried out on the salient regions of spots to fully mine the spot feature information at different scales. Finally, based on the extracted multi-scale spot features, the Softmax activation function is used to accurately identify the spot types in surveillance images, which provides

收稿日期: 2025-09-27

基金项目: 安徽省高校自然科学重点研究项目(2023AH052663、2024AH050137);安徽省优秀青年教师培育项目(YQYB2025194);安徽省职业与成人教育学会教育教学研究重点项目(AZCJ2024024);安徽工商职业学院校级质量工程项目(2024xjks07)。

作者简介: 尹金(1986—),女,安徽蚌埠人,讲师,硕士,主要研究方向为机器学习、图像处理。E-mail: qwa12305@sina.com。

a precise solution for the spot analysis of surveillance images. The experimental results show that the average quality defect index (QDI) of the method combining spot salient region segmentation with the ConvNeXt network fluctuates between 0.9 and 1.0, which is higher than that of the basic ConvNeXt network, proving that spot salient region segmentation plays a positive role in improving the performance of the recognition method. The average QDI of the recognition method based on the ConvNeXt network fluctuates between 0.8 and 1.0, which is higher than that of the support vector machine-based recognition method, the HSV space fusion Retinex algorithm-based recognition method, and the multi-feature aggregation-based recognition method, demonstrating its high accuracy and strong robustness in the task of spot feature recognition for surveillance images.

Keywords: ConvNeXt network; surveillance images; spot features; salient region segmentation; recognition method

0 引言

监控图像中往往包含了大量有价值的信息,为科学决策提供参考和依据^[1]。斑点特征是监控图像中常见的表现形态和重要视觉线索,在图像识别与分析中发挥了重要作用^[2]。因此,对斑点特征进行准确识别是监控图像分析领域的重点课题。

针对上述课题,很多专家和学者进行了广泛研究。节连彬等^[3]针对变电站消防监控场景,根据提取的色调-饱和度-强度(hue-saturation-intensity, HSI)特征,利用 K-means 聚类算法进行火灾隐患区域识别;但 HSI 特征无法全面涵盖监控场景中斑点特征的所有关键信息,对于复杂监控场景下斑点特征的描述能力有限,导致对斑点特征识别准确率不高。王晶晶等^[4]将支持向量机(support vector machine, SVM)模型应用于实时采集的井下烟雾图像,先提取图像特征,再把特征输入模型,经过处理后输出分类结果,判断图像中是否存在烟雾;然而,直接对整幅图像进行处理,会受到图像中非斑点区域的干扰,从而影响烟雾识别的效果。陈卡等^[5]针对运动目标图像,提取轮廓并查找最小包围圆,通过统计各圆心之间距离及圆心连线的斜率实现目标检测;然而,运动目标的形状和姿态会随着时间的变化而变化,导致提取的轮廓出现毛刺或断裂,进一步降低了目标检测的准确性。张荣国等^[6]提出多特征聚合边界引导网络,提取视频帧显著目标的空间与运动特征,并耦合不同分辨率下的边界及时空特征以定位目标,用多层特征注意聚合模块

融合相异特征,训练时借助混合损失准确分割显著边界、获取目标;但这种融合方式对于斑点特征这种特定类型特征的提取和识别是否足够有效并不明确,无法充分挖掘斑点特征的关键信息,影响识别准确率。

上述方法难以处理监控图像中的复杂背景的干扰,导致无法有效捕捉斑点区域的特征信息,影响了识别效果。而 ConvNeXt 网络具备分层架构,这一特性使其在特征提取过程中能够同步捕获局部区域的精细细节以及整体场景的全局上下文信息。局部区域的精细细节有助于精准定位斑点特征的具体形态和位置,而整体场景的全局上下文信息则有助于理解斑点特征与周围环境的关系,从而增强对复杂场景的适配能力。此外,ConvNeXt 网络还引入了深度可分离卷积机制,这种机制在减少模型参数和计算量的同时,能够有效提取多尺度特征,进一步提升了特征提取的效率和准确性。为此,本文提出一种基于 ConvNeXt 网络的监控图像斑点特征识别方法。首先,以视觉显著性检测模型精准定位并高效分割监控图像中的斑点显著区域,规避非斑点区域干扰;其次,将分割区域输入基于深度可分离卷积机制改进的 ConvNeXt 网络,利用其高效计算特性深度提取多尺度斑点特征;最后,借助 Softmax 激活函数准确识别斑点类型,以此提升识别准确性与鲁棒性,助力监控系统智能化应用。

1 监控图像斑点特征识别研究

基于 ConvNeXt 网络的监控图像斑点特征识别方法主要分为 2 部分:前一部分借助视觉显著性检

测划分监控图像斑点区域,降低复杂背景的干扰;后一部分借助ConvNeXt提取斑点显著区域特征,识别斑点类别。方法的简要流程如图1所示。

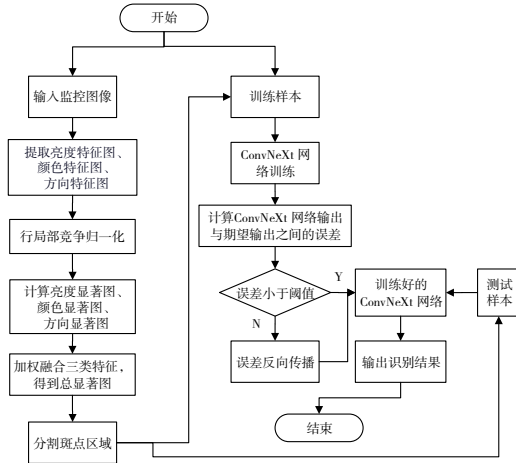


图1 监控图像斑点特征识别方法研究流程

1.1 监控图像斑点显著区域划分

在光照变化、遮挡、动态背景等复杂条件下,监控图像中包含的信息往往呈现高度的不确定性和复杂性,导致后续识别方法难以区分目标与背景,出现漏检或误检^[7]。因此,为提高ConvNeXt网络的鲁棒性和准确性,需要从监控图像中划分出斑点区域,所采用的方法为视觉显著性检测模型,通过多尺度特征融合模拟人眼快速定位显著区域的过程,从图像中提取显著区域^[8]。具体过程如下。

步骤1:输入待识别的监控图像,记为 $A(i,j)$ 。

步骤2:对图像实施高斯低通滤波操作,以实现图像的平滑处理,进而得到处理后的图像 $\tilde{A}(i,j)$ 。

步骤3:多尺度特征提取。

针对待识别的监控图像,首先开展灰度化处理操作,将其转换为灰度图像。随后,构建高斯金字塔模型,利用该模型对灰度化处理后的图像 $\tilde{A}(i,j)$ 实施降采样处理,进而生成具有不同尺度的多尺度图像序列,如式(1)所示。

$$\hat{A}(i,j) = \frac{[R_A(i,j) + G_A(i,j) + B_A(i,j)]}{3} \quad (1)$$

式中: $R_A(i,j)$ 、 $G_A(i,j)$ 、 $B_A(i,j)$ 分别为视频监控图像

在点 (i,j) 处的红色(R)、绿色(G)、蓝色(B)分量。

从待处理图像中提取红色-绿色(R-G)及蓝色-黄色(B-Y)这2个具有对立特性的颜色通道信息,随后针对这2个颜色通道分别构建高斯金字塔结构^[9]。

使用Gabor滤波器在4个方向(0° 、 45° 、 90° 和 135°)上提取边缘特征,构建方向金字塔。

步骤4:中心-周围差分,提取特征图。

1) 亮度特征图:计算不同尺度间的亮度差异,生成6张亮度特征图 $I(l,h)$ 。

2) 颜色特征图:计算红-绿和蓝-黄通道的中心-周围差,生成6张 $R-G(l,h)$ 特征图、6张 $R-Y(l,h)$ 特征图。

3) 方向特征图:计算不同方向的中心-周围差,生成24张方向特征图,这些特征图由6个不同尺度对与4个方向组合而成,并将其记为 $C(l,h,\theta)$,如式(2)所示。

$$\begin{cases} I(l,h) = |I(l) \ominus I(h)| \\ \begin{cases} R-G(l,h) = |[R(l)-G(l)] \ominus [R(h)-G(h)]| \\ B-Y(l,h) = |[B(l)-Y(l)] \ominus [B(h)-Y(h)]| \end{cases} \\ C(l,h,\theta) = |C(l,\theta) \ominus C(h,\theta)| \end{cases} \quad (2)$$

式中: l 为中心尺度; h 为邻域尺度; $I(l)$ 和 $I(h)$ 分别为中心尺度 l 和邻域尺度 h 的灰度值; \ominus 为高斯差分操作,即对中心尺度图像进行上采样至周围尺度分辨率后相减; $R(l)$ 和 $G(l)$ 分别为中心尺度 l 的红色通道和绿色通道值; $R(h)$ 和 $G(h)$ 分别为邻域尺度 h 的红色通道和绿色通道值; $B(l)$ 和 $Y(l)$ 分别为中心尺度 l 的蓝色通道和黄色通道值; $B(h)$ 和 $Y(h)$ 分别表示邻域尺度 h 的蓝色通道和黄色通道值; $C(l,\theta)$ 和 $C(h,\theta)$ 分别为中心尺度 l 和邻域尺度 h 在方向 θ 上的特征值。

步骤5:对特征图 $I(l,h)$ 、 $R-G(l,h)$ 、 $R-Y(l,h)$ 和 $C(l,h,\theta)$ 进行局部竞争归一化,目的是抑制背景噪声干扰,同时增强显著区域的特征表达^[10]。具体操作方法为:计算特征图的全局最大值,将特征图除以(最大值+常数)以压缩动态范围。

上述特征图局部竞争归一化的特征图分别记为 $\bar{I}(l, h)$ 、 $\bar{R} - \bar{G}(l, h)$ 、 $\bar{B} - \bar{Y}(l, h)$ 和 $\bar{C}(l, h, \theta)$ 。

步骤 6: 使用跨尺度合并策略(将不同尺度的特征图相加)计算显著图,如式(3)所示。

$$\begin{cases} \ddot{I} = \bigoplus_l \bigoplus_h \bar{I}(l, h) \\ \ddot{S} = \bigoplus_l \bigoplus_h v_{R-G}(l, h) \times \bar{R} - \bar{G}(l, h) + \\ \quad \bigoplus_l \bigoplus_h v_{B-Y}(l, h) \times \bar{B} - \bar{Y}(l, h) \\ \ddot{C} = \sum_{\theta \in (0^\circ, 45^\circ, 90^\circ, 135^\circ)} \bigoplus_l \bigoplus_h \bar{C}(l, h, \theta) \end{cases} \quad (3)$$

式中: $v_{R-G}(l, h)$ 和 $v_{B-Y}(l, h)$ 分别为 $R - G(l, h)$ 和 $R - Y(l, h)$ 的权重, \ddot{I} 、 \ddot{S} 和 \ddot{C} 分别为亮度显著图、颜色显著图和方向显著图, \bigoplus 为跨尺度相加后上采样至原始尺寸。

步骤 7: 加权融合 3 类特征, 得到总显著图 H , 如式(4)所示。

$$H = \frac{1}{3} \ddot{I} + \frac{1}{3} \ddot{S} + \frac{1}{3} \ddot{C} \quad (4)$$

步骤 8: 通过胜者全取机制, 从 H 中选择最显著的点作为注意力焦点, 并对已关注的区域进行临时抑制, 避免重复检测。

视觉显著性区域通常对应于关键目标或异常事件, 在这一研究中, 视觉显著性区域就是监控图像斑点所在区域, 将这部分区域从监控图像中分割出来, 得到监控图像斑点区域图像。

1.2 ConvNeXt 网络识别图像斑点类型

在这一环节, 以分割出的斑点区域为目标, 通过 ConvNeXt 网络识别图像斑点类型。ConvNeXt 通过借鉴 Transformer 的设计思想, 对传统卷积神经网络进行了升级, 提升了卷积神经网络的特征提取和学习能力, 克服传统方法的局限性, 提高斑点特征识别的准确性和效率^[11]。ConvNeXt 网络结构图如图 2 所示。

由图 2 可知, 首先需要输入训练样本(带有类别标签的监控图像斑点区域显著图像)对 ConvNeXt 网络进行训练。训练样本记为 (H_i, Y_i) , $i = 1, 2, \dots, N$, 其中, H_i 是第 i 个训练用监控图像斑点区域显著图像, Y_i 是第 i 个训练用图像类别概率^[12]。对于输

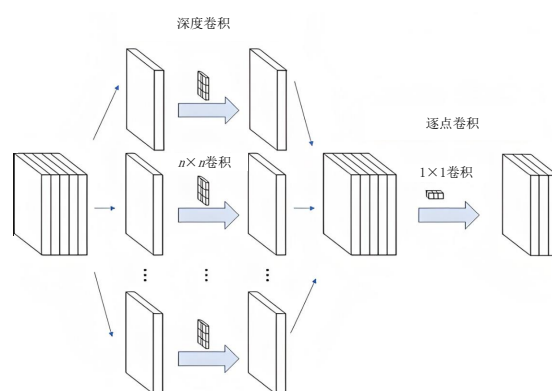


图 2 ConvNeXt 网络结构图

入的监控图像 H_i , 运用尺寸为 4×4 的卷积核, 并将步长设定为 4, 以此对输入图像执行下采样操作。下采样后监控图像的输出尺寸如式(5)所示。

$$\begin{cases} J_{\hat{H}_i} = \frac{J_H - K + 2L}{\mu} + 1 \\ W_{\hat{H}_i} = \frac{W_H - K + 2L}{\mu} + 1 \end{cases} \quad (5)$$

式中: $J_{\hat{H}_i}$ 和 $W_{\hat{H}_i}$ 分别为下采样后监控图像的高度和宽度, J_H 和 W_H 分别为监控图像原始高度和宽度, K 为深度卷积核尺寸, L 为填充, μ 为步长。

针对下采样图像 \hat{H}_i 进行深度可分离卷积, 提取多尺度斑点特征^[13]。深度可分离卷积公式如式(6)所示。

$$\begin{cases} g_k = \text{conv}(\hat{H}_i, k) \\ G_i = \{g_1, g_2, \dots, g_k\} \end{cases} \quad (6)$$

式中: g_k 为第 k 个深度卷积核提取的斑点特征图, 反映了输入图像在该位置与特定卷积核响应后的特征; $\text{conv}(\hat{H}_i, k)$ 为卷积运算; G_i 为下采样图像 \hat{H}_i 的斑点特征图。

输入 G_i 到全局上下文信息嵌入模块^[14], 进行扩大感受野, 如式(7)所示。

$$U_k = \left[\left[\sum_{a=1}^{J_H} \sum_{b=1}^{W_H} \omega g_k(a, b) \right]^2 + \delta \right]^{\frac{1}{2}} \quad (7)$$

式中: ω 为自适应嵌入权重, a 和 b 分别为下采样后监控图像任意的高度和宽度, δ 为极小常数, U_k 为全局上下文信息嵌入模块输出结果。

利用注意力机制, 捕捉监控图像斑点特征图显

著特征,提高卷积神经网络特征提取能力,其计算公式如式(8)所示。

$$Z_k = f(\kappa \times \text{GAP}(U_k) + \nu) \otimes U_k \quad (8)$$

式中: Z_k 为注意力权重特征图, f 为 Sigmoid 激活函数, $\text{GAP}()$ 为全局平均池化, κ 和 ν 分别为权重和偏置。

基于通过 Softmax 激活函数计算斑点分类概率分布 \hat{Y}_i ,将 ConvNeXt 网络的原始输出转换为概率分布。针对 ConvNeXt 网络,首先,计算其输出结果 \hat{Y}_i

与期望输出 Y_i 之间的误差。其次,利用误差反向传播算法,将误差从输出层向输入层逐层反向传递;在反向传播过程中,根据误差梯度对 ConvNeXt 网络的参数进行更新和优化,最终完成 ConvNeXt 网络的训练流程^[15]。最后,将待识别的监控图像输入到训练好的 ConvNeXt 网络中,输出分类识别结果。

基于上述描述的监控图像斑点显著区域划分及 ConvNeXt 网络识别图像斑点类型的伪代码如图 3~5 所示。

```
import numpy as np
import cv2
import torch
import torch.nn as nn
import torchvision.transforms
from torch.utils.data import DataLoader, Dataset
# 假设的 Gabor 滤波器函数
def apply_gabor_filter(image, angle):
    # 应用 Gabor 滤波器返回边缘特征
    pass
# 假设的深度可分离卷积函数
def depthwise_separable_convolution(image, kernel):
    # 执行深度可分离卷积操作
    pass
# 假设的全局上下文信息嵌入模块
def global_context_embedding(feature_map):
    # 扩大感受野并嵌入全局上下文信息
    pass
# 假设的注意力机制函数
def attention_mechanism(feature_map):
    # 计算注意力权重并应用注意力机制
    pass
# 监控图像斑点显著区域划分伪代码
def saliency_region_detection(image):
    # 步骤 1: 输入监控图像
    # image = cv2.imread('input_image.jpg')
    # 步骤 2: 高斯低通滤波
    blurred_image = cv2.GaussianBlur(image, (5, 5), 0)
    # 步骤 3: 多尺度特征提取
    # 灰度化
    gray_image = cv2.cvtColor(blurred_image, cv2.COLOR_BGR2GRAY)
    # 构建高斯金字塔
    pyramid = [gray_image]
    for _ in range(3): # 降采样 3 次
        pyramid.append(cv2.pyrDown(pyramid[-1]))
    # 提取 R-G 和 B-Y 颜色通道并构建高斯金字塔(此处简化处理)
    # 实际应用中需分离 R, G, B 通道并计算 Y 通道
    # 使用 Gabor 滤波器提取边缘特征
    orientations = [0, 45, 90, 135]
    edge_features = []
    for angle in orientations:
        edge_feature = apply_gabor_filter(gray_image, angle)
        edge_features.append(edge_feature)
    # 步骤 4: 中心-周围差分提取特征图
    # 此处为简化描述实际实现需计算亮度、颜色和方向特征图
    # 已生成特征图 luminance_maps, color_maps, orientation_maps
    # 步骤 5: 局部竞争归一化
    # 步骤 6: 跨尺度合并计算显著图
    def normalize(feature_map):
        max_val = np.max(feature_map)
        return feature_map / (max_val + 1e-6) # 添加极小常数避免除零
    normalized_luminance = [normalize(map) for map in luminance_maps]
    normalized_color = [normalize(map) for map in color_maps]
    normalized_orientation = [normalize(map) for map in orientation_maps]
    # 步骤 7: 加权融合 3 类特征
    total_saliency = 0.5 * luminance_saliency + 0.3 * color_saliency + 0.2 * orientation_saliency
    # 步骤 8: 胜者全取机制选择最显著点
    # 此处简化处理实际应用中需实现胜者全取算法
    saliency_map = total_saliency
    # 找到最显著点并分割出斑点区域
    saliency_region = segment_saliency_region(image, saliency_map)
    return saliency_region

def train_convnext(train_loader, model, criterion, optimizer, num_epochs):
    model.train()
    for epoch in range(num_epochs):
        for images, labels in train_loader:
            loss.backward()
            optimizer.step()
            optimizer.zero_grad()
            outputs = model(images)
            loss = criterion(outputs, labels)
        print(f'Epoch {epoch+1}, Loss: {loss.item():.4f}')
    # 主流程
    if __name__ == "__main__":
        # 1. 监控图像斑点显著区域划分
        input_image = cv2.imread('monitor_image.jpg')
        saliency_region = saliency_region_detection(input_image)
        # 2. 准备训练数据
        train_dataset = CustomDataset(saliency_region) # 假设的自定义数据集
        train_loader = DataLoader(train_dataset, batch_size=32, shuffle=True)
        # 3. 初始化 ConvNeXt 模型
        model = ConvNeXt(num_classes=5) # 有 5 类斑点
        criterion = nn.CrossEntropyLoss()
        optimizer = optim.Adam(model.parameters(), lr=0.001)
        # 4. 训练 ConvNeXt 模型
        train_convnext(train_loader, model, criterion, optimizer, num_epochs=10)
        # 5. 使用训练好的模型进行预测
        model.eval()
        with torch.no_grad():
            output = model(torch.randn(1, 3, 224, 224)) # 假设输入尺寸
            predicted_class = torch.argmax(output, dim=1)
            print(f'Predicted class: {predicted_class.item()}')
```

图 3 算法伪代码

```
class ConvNeXt(nn.Module):
    def __init__(self, num_classes):
        super(ConvNeXt, self).__init__()
        # 网络结构定义
        self.conv1 = nn.Conv2d(3, 64, kernel_size=4, stride=4)
        self.depthwise_conv = nn.Conv2d(64, 128, kernel_size=3, groups=64)
        self.global_context = GlobalContextEmbedding() # 全局上下文模块
        self.attention = AttentionMechanism() # 假设的注意力机制
        self.fc = nn.Linear(128, num_classes)
    def forward(self, x):
        # 下采样
        x = self.conv1(x)
        # 深度可分离卷积
        x = self.depthwise_conv(x)
        # 全局上下文嵌入
        x = self.global_context(x)
        # 注意力机制
        x = self.attention(x)
        # 全局平均池化
        x = nn.functional.adaptive_avg_pool2d(x, (1, 1))
        x = torch.flatten(x, 1)
        # 分类
        x = self.fc(x)
    return x
```

图 4 ConvNeXt 网络识别图像斑点类型伪代码

```
def train_convnext(train_loader, model, criterion, optimizer, num_epochs):
    model.train()
    for epoch in range(num_epochs):
        for images, labels in train_loader:
            loss.backward()
            optimizer.step()
            optimizer.zero_grad()
            outputs = model(images)
            loss = criterion(outputs, labels)
        print(f'Epoch {epoch+1}, Loss: {loss.item():.4f}')
    # 主流程
    if __name__ == "__main__":
        # 1. 监控图像斑点显著区域划分
        input_image = cv2.imread('monitor_image.jpg')
        saliency_region = saliency_region_detection(input_image)
        # 2. 准备训练数据
        train_dataset = CustomDataset(saliency_region) # 假设的自定义数据集
        train_loader = DataLoader(train_dataset, batch_size=32, shuffle=True)
        # 3. 初始化 ConvNeXt 模型
        model = ConvNeXt(num_classes=5) # 有 5 类斑点
        criterion = nn.CrossEntropyLoss()
        optimizer = optim.Adam(model.parameters(), lr=0.001)
        # 4. 训练 ConvNeXt 模型
        train_convnext(train_loader, model, criterion, optimizer, num_epochs=10)
        # 5. 使用训练好的模型进行预测
        model.eval()
        with torch.no_grad():
            output = model(torch.randn(1, 3, 224, 224)) # 假设输入尺寸
            predicted_class = torch.argmax(output, dim=1)
            print(f'Predicted class: {predicted_class.item()}')
```

图 5 训练过程伪代码

2 实验测试

实验测试的目的是验证 ConvNeXt 网络在监控图像斑点特征(如污渍、划痕、异物等)识别任务中

的有效性,通过与传统方法对比,分析其性能优势。

2.1 实验环境设置

在实验测试环节,所依托的软硬件环境具体细节如下。

1) 硬件环境。

GPU 服务器:NVIDIA A100 80GB × 2(用于大规模训练和推理);

CPU: Intel Xeon Platinum 8358(32核,64线程,用于数据预处理);

内存:256 GB DDR4 ECC;

存储:NVMe SSD 4 TB(用于数据集存储)。

2)软件环境。

操作系统:Ubuntu 20.04 LTS;

深度学习框架:PyTorch 1.12.0 + CUDA 11.6;

处理工具:OpenCV 4.7, Albumentations 1.3.0;

可视化工具:TensorBoard、Matplotlib(训练过程监控与结果分析);

编程语言:Python 3.8;

性能分析:NVIDIA Nsight, PyTorch Profiler。

2.2 样本选取

监控图像中的斑点特征往往具有很强研究价值,例如在生产车间的监控中,监控图像中存在的斑点可能就是潜在的故障隐患;在城市安防的监控中,监控图像中存在的斑点可能就是可疑人员或物品;在医疗的监控场景里,监控图像中存在的斑点可能就是患者的病灶表现。在不同的场景中,斑点特征表现不同的形式,因此从不同的数据集中抽取监控图像,作为 ConvNeXt 网络训练样本,从各类不同的监控场景中截取监控图像,组成测试样本集。样本的具体分布情况如表 1 所示。

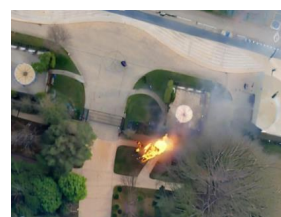
表 1 样本分布情况

| 序号 | 数据集来源 | 监控场景类型 | 斑点类型 | 训练集样本数/个 | 测试集样本数/个 |
|----|--------------|----------|----------------|----------|----------|
| 1 | MVTec AD | 工业生产线上监控 | 划痕、凹痕、锈迹等 | 8 000 | 2 500 |
| 2 | PedX | 城市安防监控 | 遗留包裹、火灾、破损井盖 | 9 000 | 1 800 |
| 3 | MediSpot | 医疗内窥镜影像 | 息肉、出血点、溃疡面 | 6 000 | 1 200 |
| 4 | TrafficGuard | 交通设施监控 | 车牌污损、路面坑洞、标线缺失 | 7 000 | 1 500 |
| 5 | MixSynthetic | 多场景合成数据 | 模拟工业+安防复合缺陷 | 5 000 | 1 000 |
| 合计 | | | | 35 000 | 8 000 |

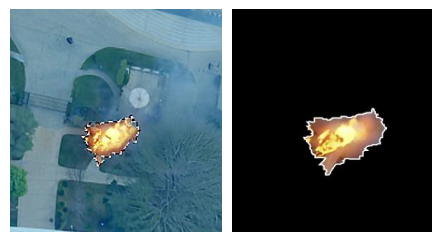
共选取 5 类样本(表 1)进行实验测试,其中,MVTec AD 数据集每张图片分辨率统一为 700×700 像素,提供对应的分割掩码用于精确计算异常区域的位置与范围,标注类型为语义分割图;PedX 数据集包含高清图像,提供行人、车辆、异常物体的边界框或像素级标注,标注精细且丰富;MediSpot 数据集分辨率满足医疗影像需求,包含病灶区域的边界框标注或像素级标注;TrafficGuard 数据集分辨率各异但适用于现代目标检测模型,每个交通异常特征都被精确框选并标注类别,可直接用于目标检测训练;MixSynthetic 数据集作为合成数据,分辨率和标注方式根据实验需求自定义,包含多种场景的复合缺陷标注,如工业划痕与安防遗留包裹的组合标注。这些数据集在分辨率和标注方式上的多样性,为 ConvNeXt 网络提供了丰富且具有挑战性的训练样本。

2.3 斑点显著区域划分结果

基于 1.1 的研究,针对监控图像样本进行监控图像斑点显著区域划分,结果如图 6 所示。



(a)原监控图像



(b)斑点显著区域

图 6 监控图像斑点区域图像划分结果

从图 6 可以看出,通过视觉显著性检测,模型在

监控图像上准确标记出了斑点显著区域,该区域在图像中清晰可见,与周围环境形成鲜明对比。

2.4 斑点识别结果

针对监控图像斑点区域图像,利用训练好的 ConvNeXt 网络进行斑点特征识别,识别结果如表 2 所示。

表 2 斑点特征识别结果示例

| 编号 | 斑点特征描述 | ConvNeXt 识别结果 | 最大类别概率/% | 实际类别标签 | 是否正确 |
|-----|----------------|---------------|----------|----------|------|
| 001 | 路面中央圆形污渍 | 污渍 | 95 | 污渍 | 是 |
| 002 | 井盖边缘裂纹 | 破损 | 88 | 破损(井盖边缘) | 是 |
| 003 | 行人道砖块缺失 | 异物/缺失 | 76 | 砖块缺失 | 是 |
| 004 | 道路积水中的悬浮物 | 异物 | 92 | 异物(漂浮物) | 是 |
| 005 | 路面长条形划痕 | 划痕 | 85 | 划痕 | 是 |
| 006 | 轻微路面凹陷(非明显坑洞) | 轻微破损 | 72 | 轻微凹陷 | 是 |
| 007 | 井盖完全缺失,露出下方空洞 | 严重破损 | 98 | 井盖缺失 | 是 |
| 008 | 路面油渍 | 污渍 | 89 | 污渍(油渍) | 是 |
| 009 | 道路边缘的小石子堆积 | 异物/堆积 | 68 | 小石子堆积 | 是 |
| 010 | 路面无明显斑点(测试负样本) | 无显著斑点 | 99 | 无 | 是 |

从表 2 可以看出,ConvNeXt 网络能够准确识别监控图像中的多种类型的斑点特征,包括污渍、破损、划痕、异物等,并且其识别结果与实际类别标签高度一致,表明 ConvNeXt 网络在斑点特征识别任务上具有较高的准确性。

2.5 性能测试

以平均质量缺陷指数(quality defect index, QDI)为指标,进行方法性能测试,其计算公式如式(9)~(10)所示。

$$Q_i = \alpha s_i + \beta \zeta_i + \gamma \xi_i \quad (9)$$

$$\bar{Q} = \frac{\sum_{i=1}^N Q_i}{N} \quad (10)$$

式(9)~(10)中: \bar{Q} 为平均 QDI, N 为监控图像样本的总数, Q_i 为包含第 i 类斑点监控图像的 QDI 分数, α 、 β 、 γ 为权重系数, s_i 、 ζ_i 和 ξ_i 分别为方法对第 i 类斑点监控图像识别的召回率、精确率和 IoU(交并比)。

针对所研究方法进行消融实验和对比实验。消融部分为是否进行监控图像斑点显著区域划分,基于表 1 所列的实验样本,采用随机有放回抽样策略,从总体样本池中抽取 50 个实验子集,每个子集严格包含 100 个图像样本,以系统探究该消融变量对模型性能的影响。实验设定了基准模型与 3 个消融模

型进行对比实验:基准模型采用完整的 ConvNeXt 网络架构,涵盖注意力机制和深度卷积等所有核心模块;消融模型 1 在基准模型基础上移除注意力机制模块,其余结构保持不变;消融模型 2 将深度卷积模块替换为普通卷积模块,其他部分维持基准模型配置;消融模型 3 则同时移除注意力机制模块并将深度卷积模块替换为普通卷积模块,以此全面评估各模块对模型性能的贡献。消融实验结果如图 7 所示。

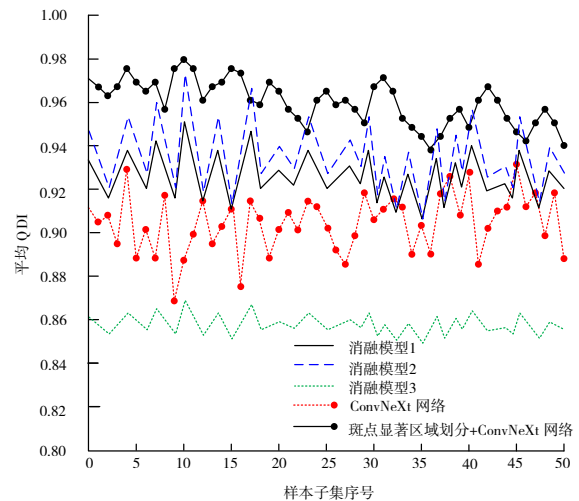


图 7 消融实验结果

从图 7 可以看出,以平均 QDI 为评估指标,基准模型(采用完整 ConvNeXt 网络架构)展现出了最佳性能,这表明注意力机制和深度卷积模块在模型中发挥了重要作用。相较于基准模型、消融模型 1(移

除注意力机制模块)和消融模型2(将深度卷积模块替换为普通卷积模块)的性能均有所下降,说明这2个模块分别对提升模型召回率、精确率及IoU起到了关键作用。而消融模型3(同时移除注意力机制模块并将深度卷积模块替换为普通卷积模块)的性能下降最为显著,进一步验证了注意力机制和深度卷积模块在 ConvNeXt 网络中的协同作用,它们共同促进了模型在斑点显著区域划分及识别任务中的优异表现。

为了进一步验证本文方法的效果,以基于支持向量机的识别方法、基于 HSV 空间融合 Retinex 算法的识别方法和基于多特征聚合的识别方法这3种前人研究方法做对比方法进行识别性能实验,并以平均 QDI 为评估指标,实验结果如图 8 所示。

由图 8 可知,本文所提方法在斑点显著区域划分及识别任务中的性能显著优于3种对比方法。具体而言,相较于3种对比方法,本文方法在 QDI 上均取得了更高的数值。这是由于本文方法通过 ConvNeXt 网络架构,结合对监控图像斑点显著区域的有效划分,能更精准地提取和识别斑点特征,减少背景干

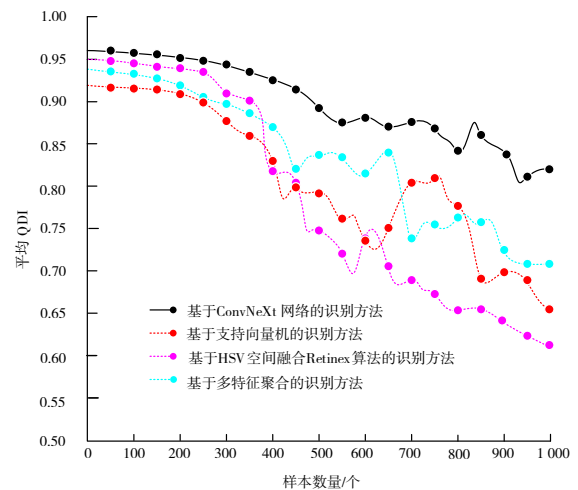


图 8 平均 QDI 对比

扰,从而在平均 QDI 上取得明显优势,有力验证了本文方法的优越性。

为验证 ConvNeXt 网络在斑点显著区域划分及识别任务中的跨场景泛化能力,使用选定的训练集对 ConvNeXt 网络进行训练,调整网络参数以优化在训练集上的性能。在每次实验中,选择一个未参与训练的场景作为测试集。使用训练好的 ConvNeXt 网络对测试集进行斑点显著区域划分及识别,记录并分析模型在测试集上的准确率、召回率、F1 分数等性能指标。4 种方法的实验结果如表 3 所示。

表 3 不同方法的跨场景泛化能力对比

%

| 方法 | 测试集场景类型 | 准确率 | 召回率 | F1 分数 |
|-----------------------------|--------------|------|------|-------|
| 基于 ConvNeXt 网络的识别方法 | MVTec AD | 87.9 | 85.8 | 86.8 |
| | PedX | 85.2 | 82.1 | 83.6 |
| | MediSpot | 88.7 | 86.5 | 87.6 |
| | TrafficGuard | 83.4 | 80.9 | 82.1 |
| | MixSynthetic | 86.1 | 84.3 | 85.2 |
| 基于支持向量机的识别方法 | MVTec AD | 78.5 | 75.2 | 76.8 |
| | PedX | 76.3 | 73.1 | 74.6 |
| | MediSpot | 79.2 | 76.8 | 78.0 |
| | TrafficGuard | 74.1 | 71.0 | 72.5 |
| | MixSynthetic | 77.0 | 74.5 | 75.7 |
| 基于 HSV 空间融合 Retinex 算法的识别方法 | MVTec AD | 80.1 | 77.6 | 78.8 |
| | PedX | 78.9 | 75.9 | 77.4 |
| | MediSpot | 81.5 | 79.2 | 80.3 |
| | TrafficGuard | 76.7 | 73.6 | 75.1 |
| | MixSynthetic | 79.4 | 77.1 | 78.2 |
| 基于多特征聚合的识别方法 | MVTec AD | 82.3 | 80.1 | 81.2 |
| | PedX | 80.7 | 78.3 | 79.5 |
| | MediSpot | 83.6 | 81.4 | 82.5 |
| | TrafficGuard | 79.5 | 76.4 | 77.9 |
| | MixSynthetic | 81.2 | 79.0 | 80.1 |

由表3可知,在各个测试集场景类型中,基于ConvNeXt网络的识别方法在准确率、召回率和F1分数上均显著优于其他3种对比方法。以MediSpot场景为例,ConvNeXt网络准确率达88.7%、召回率86.5%、F1分数87.6%,远高于其他3种对比方法。这充分表明,ConvNeXt网络在斑点显著区域划分及识别任务中具有更强的跨场景泛化能力,能够更好地适应不同场景下的斑点特征,展现出更优的性能。

2.6 多尺度样本实验分析

为进一步验证ConvNeXt网络在处理多尺度样本时的性能表现,从已有数据集中选取不同分辨率的图像样本,划分为小尺度(分辨率为450×450像素)、中尺度(分辨率为800×800像素)和大尺度(分辨率为1400×1400像素)3组,每组包含1000个样本,覆盖工业、安防、医疗等多样场景。通过ConvNeXt网络进行识别,记录其在不同尺度下的准确率、召回率和F1分数,结果如表4所示。

表4 多尺度样本识别性能

| 尺度组别 | 准确率 | 召回率 | F1分数 |
|------|------|------|------|
| 小尺度 | 84.5 | 82.1 | 83.3 |
| 中尺度 | 87.8 | 85.6 | 86.7 |
| 大尺度 | 89.2 | 87.3 | 88.2 |

由表4可知,ConvNeXt网络在不同尺度样本上均表现出较高的识别性能,且随着样本尺度的增大,准确率、召回率和F1分数均呈上升趋势。这表明ConvNeXt网络通过其低层局部特征提取与高层

全局上下文融合多层卷积结构,能够有效适应多尺度输入,充分挖掘斑点特征信息。具体而言,网络中的深度可分离卷积机制在减少参数数量的同时,增强了多尺度特征表达能力,而注意力机制进一步聚焦于显著区域,提升了模型对复杂背景的鲁棒性。

3 结束语

本研究围绕“基于ConvNeXt网络的监控图像斑点特征识别方法”展开并取得了以下成果:通过将斑点显著区域划分与ConvNeXt网络相结合,弥补了传统方法依赖特征提取质量、泛化能力不足的问题,为监控图像分析提供了参考;同时,采用消融实验与对比实验相结合的方法,系统评估了斑点显著区域划分对识别性能的影响,并与基于支持向量机的识别方法、基于HSV空间融合Retinex算法的识别方法、基于多特征聚合的识别方法这3种现有研究方法进行了对比,确保了研究结论的可靠性和准确性。

然而,该方法仍存在一定局限性,在资源受限场景下,模型计算开销较大,难以直接部署于边缘端设备;同时,面对复杂多变的动态场景,对快速移动或遮挡目标的跟踪能力有待提升。其适用范围主要集中在对监控图像中斑点特征识别有较高要求的场景,包括工业质检、安防监控等。未来,将从模型轻量化、多模态融合以及动态场景优化等方面着手,进一步提高识别方法的性能,推动其向更高效、更智能的方向发展。

参考文献:

- [1] WU Z F, HOU F Z. Design of laser image recognition system based on high performance computing of spatiotemporal data[J]. Intelligent Decision Technologies, 2024, 18(4): 2701-2714.
- [2] CHEN S, GUO D T, WANG D G. Image recognition system for mechanical parts based on SSD and YOLOv5[J]. Journal of Electrical Systems, 2024, 20(3): 852-869.
- [3] 节连彬,赵嘉兴,刘毅敏.基于特征识别的变电站消防监控图像处理方法[J].沈阳工业大学学报,2023,45(6):704-709.
- [4] 王晶晶,宋启祥.支持向量机技术在井下烟雾图像检测中的应用[J].煤炭技术,2023,42(11):213-216.
- [5] 陈卡,黄俊杰,包嘉琪,等.基于HSV空间融合Retinex算法的全天候运动目标检测[J].矿冶工程,2023,43(5):17-21+25.
- [6] 张荣国,郑晓鸽,王丽芳,等.多特征聚合的边界引导视频图像显著目标检测[J].中国图象图形学报,2025,30(4):1141-1154.
- [7] ZHANG Z, LEI Z, OMURA M, et al. Dendritic learning-incorporated vision transformer for image recognition[J]. IEEE/CAA

Journal of Automatica Sinica, 2024, 11(2): 539-541.

- [8] 王钰,赵明晶,谢晓明,等.基于曲率滤波和视觉显著性的红外小目标检测[J].北京化工大学学报(自然科学版),2023,50(4):75-86.
- [9] 梁秀雅,冯水春,陈红珍.结合视觉显著性和 EfficientNetV2 的舰船目标检测方法[J].计算机工程与应用,2024,60(5):259-270.
- [10] 梁浩然,方加奇,梁荣华.相位谱提示多特征聚合的视觉显著性检测方法[J].小型微型计算机系统,2023,44(12):2797-2803.
- [11] SU S, ZHANG K, ZHU M J S. Similarity-sequenced multi-view discriminant feature extraction for image recognition[J]. Journal of Modern Optics, 2023, 70(8): 503-516.
- [12] 何德芬,江倩,金鑫,等.基于 ConvNeXt 的伪造人脸检测方法[J].信息安全研究,2025,11(3):231-240.
- [13] 韩源涛,张聪,詹晓芸,等.基于改进 ConvNeXt 模型的农作物害虫细粒度识别[J].农业工程学报,2025,41(4):185-192.
- [14] 张浩,张大海,刘乃毓,等.基于改进 VMD 及 ConvNeXt 的小电流接地系统单相接地故障选线方法[J].高电压技术,2025,51(2):730-741.
- [15] 崔素晓,武哲,崔彦平,等.基于频率切片小波变换和注意力增强 ConvNeXt 模型的行星齿轮箱故障诊断[J].兵工学报,2025,46(2):157-166.

责任编辑:蒋召雪

(上接第 92 页)

- [3] 王奥,吴华瑞,朱华吉.基于特征增强的多方位农业问句语义匹配[J].西南大学学报(自然科学版),2023,45(6):201-210.
- [4] 张文慧,汪美玲,侯志荣.融合语境语义差异特征的短文本匹配模型[J].北京大学学报(自然科学版),2023,59(1):30-38.
- [5] 刘萌,齐孟津,詹圳宇,等.基于深度学习的图像-文本匹配研究综述[J].计算机学报,2023,46(11):2370-2399.
- [6] 刘建平,初新涛,王健,等.面向中文科学数据集的句子级语义匹配模型[J].郑州大学学报(工学版),2024,45(6):56-64.
- [7] WU F, WANG Q, WANG Z, et al. ITCContrast: contrastive learning with hard negative synthesis for image-text matching[J]. The Visual Computer, 2024, 40(12): 8825-8838.
- [8] 陈佳,张鸿.基于特征增强和语义相关性匹配的图像文本检索方法[J].计算机应用,2024,44(1):16-23.
- [9] 王钦晨,段利国,王君山,等.基于 BERT 字句向量与差异注意力的短文本语义匹配策略[J].计算机工程与科学,2024,46(7):1321-1330.
- [10] DUAN J, JIA M, LIAO J, et al. HyperMatch: long-form text matching via hypergraph convolutional networks[J]. Knowledge and Information Systems, 2024, 66(11): 6597-6616.
- [11] 李艺茹,姚涛,张林梁,等.基于多级语义对齐的图像-文本匹配算法[J].北京航空航天大学学报,2024,50(2):551-558.
- [12] 刘东旭,段利国,崔娟娟,等.融合义原相似度矩阵与字词向量双通道的短文本语义匹配策略[J].计算机科学,2024,51(12):250-258.
- [13] 张贞港,余传明.基于知识增强的文本语义匹配模型研究[J].情报学报,2024,43(4):416-429.
- [14] 陈岳林,高铸成,蔡晓东.基于 BERT 与密集复合网络的长文本语义匹配模型[J].吉林大学学报(工学版),2024,54(1):232-239.
- [15] SUN H, QIN X, LIU X. Image-text matching using multi-subspace joint representation[J]. Multimedia Systems, 2023, 29(3): 1057-1071.

责任编辑:李远辉