

DOI:10.19651/j.cnki.emt.2211492

侧窗自适应彩色增强算法在司机行为识别系统中的应用研究

崔育洋 袁小军 田野 姚巍巍 彭联贴 李晨

(株洲中车时代电气股份有限公司数据与智能技术中心 株洲 412001)

摘要: 外界光照的变化容易干扰机车司机室视频的图像质量,出现图像亮度异常现象,导致司机行为识别系统检测精度下降。针对此问题,提出了一种基于侧窗滤波的自适应非线性彩色增强算法,并设计了一种新型司机行为识别系统。首先利用主聚类推定算法,建立图像照度分类模型,将司机室视频图像分类为低光照、正常光照和曝光3种场景。然后采用本文所提算法对低光照图像进行增强,有效提高了图像亮度、对比度和加强了暗区细节信息。最后利用深度学习方法,建立了基于YOLOv3的司机行为检测模型。为证明可行性,选取某铁路局机务段的6A视频在NVIDIA视频分析服务器上进行试验,结果表明本文提出的低光照图像增强算法能够更好地改善图像质量,利用YOLOv3对增强后的低光照场景图像进行目标检测,项点的检测精度达到了97.20%,与优化前相比提高了6.33%,满足机务段视频智能分析的实际需求。

关键词: 司机行为识别;侧窗滤波;自适应;彩色增强;YOLOv3

中图分类号: TP2 **文献标识码:** A **国家标准学科分类代码:** 510.4050

Adaptive color enhancement based on side window filtering and application to driver behavior recognition system

Cui Xiaoyang Yuan Xiaojun Tian Ye Yao Weiwei Peng Liantie Li Chen

(Data and Intelligence R&D Center, Zhuzhou CRRC Times Electric Co., Ltd., Zhuzhou 412001, China)

Abstract: The image quality of locomotive driver's room video is easily disturbed, especially when the image brightness abnormality caused by external lighting changes, which leads to the decrease of system detection accuracy. To address this problem, this paper proposes an adaptive nonlinear color enhancement algorithm based on side-window filtering for pre-processing, and designs a novel driver behavior recognition system scheme. Using the principal clustering presumption algorithm, an image illumination classification model is established to classify 6A driver's room video images into three scenes: low illumination, normal illumination and exposure. Then the algorithm proposed in this paper is used to enhance the low-illumination 6A driver's room video image, which effectively improves the image brightness, contrast and enhances the detail information in dark areas. YOLOv3-based driving behavior detection model is established using a deep learning method. To prove the feasibility of the method, the locomotive 6A video from a railroad bureau's locomotive depot was selected for experiments on an NVIDIA video analysis server. The results show that the low-light image enhancement algorithm proposed in this paper can better improve the image quality, and the object detection accuracy of the item point reached 97.20%, which was improved by 6.33% compared with before optimization, and meet the actual demand of video intelligent analysis in the locomotive depot of railroad bureau.

Keywords: driver behavior recognition; side window filtering; adaptive; color enhancement; YOLOv3

0 引言

随着我国铁路营业里程的增加和机车运营速度的提

高,铁路安全运营显得越来越重要。司机由于工作环境特殊性容易疲劳,导致违章操作而引发行车安全事故^[1]。机务段大数据中心分析员依赖人工方式对机车车载安全防护

系统(6A 系统)视频^[2]和列车运行监控记录系统(LKJ 系统)数据^[3]联动分析,考核司机的违章行为,工作量繁重且耗时较长。司机行为识别系统能够自动化检测乘务员的驾驶行为。然而在机车进出隧道和昼夜行车过程中,外界自然光照的变化容易干扰 6A 司机室视频的成像质量,具体表现为图像过曝、亮度和对比度的整体或局部下降、以及色偏等现象,造成图像信息的损失,进而影响司机行为识别系统的检测精度。

基于深度卷积网络的计算机视觉技术发展迅速^[4-6],在轨道交通的司机行为识别领域中获得了广泛应用^[7-9]。文献[7]提出了一种二维 Inception-v3 的双流卷积网络模型和三维残差卷积神经网络,文献[8]提出了一种改进型卷积深度长短时记忆神经网络和时空增强网络,文献[9]提出了一种将人体关键点识别和目标检测通过目标框识别融合起来的多尺度金字塔神经网络模型,对列车司机行为进行识别。上述科研人员更注重通过改进模型架构或者优化训练过程,忽略了图像数据本身对模型效果的影响。以数据为中心的方法可以本质上改善输入模型的数据,提高神经网络模型的准确性^[10]。对劣化的 6A 视频的图像进行增强处理,能够有效地提高司机行为识别的精度。

图像增强传统方法包括空域法和变换域法。空域法的直方图均衡技术^[11]通过拉伸直方图的分布增强图像。变换域法的同态滤波^[12]根据图像获取过程中照明反射成像原理,在频域中增强对比度和压缩图像亮度范围。基于 Retinex 理论^[13]的图像增强算法将原始图像分解为照射分量和反射分量,并变换到频域中采用不同策略增强。基于深度学习的图像增强算法包括基于参考数据、基于合成数据和无参考数据 3 类方法^[14]。无参考数据方法主流采用基于生成对抗网络(generative adversarial network, GAN)的弱监督学习方法^[15],解决参考数据困难的图像退化问题,例如低光照图像增强和 underwater 图像增强^[16]。目前基于无参考数据的方法研究并不完善,图像退化模型较复杂,缺少完善成对的学习样本,导致增强结果和真实情况不尽符合。

本文在前人的研究基础上,提出了一种基于侧窗滤波的自适应非线性彩色增强算法。并设计了一套新型司机行为识别系统方案,即利用深度学习神经网络 YOLOv3^[17]对增强处理后的图像进行模型推理和违章作业识别。根据试验结果,上述图像增强算法可以改善图像质量,有利于后续神经网络模型的目标检测。同时,本文所设计的司机行为识别系统能够有效地实现项点检测,满足机务段视频智能分析的实际业务需求,这对保障机车安全行驶,减少人员伤亡和财产损失,以及铁路部门的降本增效有着重要意义。

1 司机行为识别系统方案

本文所提出的司机行为识别系统整体设计方案包括数据采集、数据预处理、数据分析处理和数据结果输出 4 个部

分。首先从机车联动分析系统获取 6A 视频并抽帧和 LKJ 文件预处理。然后利用基于主聚类推定^[18]的照度分类器对图像进行分类。如果 6A 司机室图像属于低光照场景的类别,则对其进行图像增强处理,图像整体的对比度和亮度会得到提高,色偏现象可以受到抑制,暗处细节信息能够获得加强,边缘信息得到保留。如果属于正常光照场景,则不需要增强处理。如果属于曝光图像,则直接删除舍弃返回至上一步骤。然后将处理后的图像输入到深度神经网络 YOLOv3 进行模型推理。最后依据列车乘务作业规范,并结合机车 LKJ 数据进行项点检测域的分析,对推理得到的司机行为检测结果进行违章判断逻辑分析处理,输出违章作业识别结果。司机行为识别系统的流程如图 1 所示。

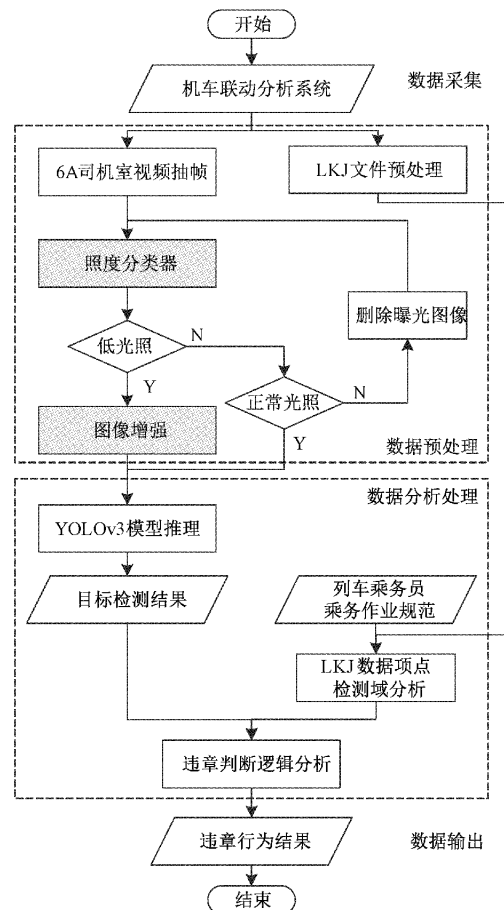


图 1 司机行为识别系统的流程

2 机车 6A 视频图像预处理

2.1 基于主聚类推定的照度分类器

将 RGB 图像转换成 HSV 图像,提取 V 分量得到亮度图像。高斯核函数的尺度类似于视觉的感受野,尺度因子越大,感受野相应越大,图像光照值的全局特性越明显;反之尺度因子越小,感受野也越小,光照值的局部特性越明显,即动态范围压缩效果越好。分别采用大尺度、中尺度和小尺度的高斯滤波器,多尺度对亮度图像的光照值进行滤

波,加权得到光照强度分量的亮度图像。然后利用主聚类推定算法计算亮度图像的亮度平均值,步骤如下:

1)将亮度图像 N 个像素点值 $x_n, n = 1, \dots, N$ 的平均值 μ_x 和标准差 σ_x 设置为主聚类的平均值估计 $\hat{\mu}_N$ 和标准差估计 $\hat{\sigma}_N$ 的初始值,如式(1)所示。

$$\hat{\mu}_N \leftarrow \mu_x, \hat{\sigma}_N \leftarrow \sigma_x \quad (1)$$

2)主聚类方向上采用平均值 μ_w 和标准差 σ_w 的高斯分布 $p(x; \mu_w, \sigma_w)$ 作为核函数, μ_w 和 σ_w 分别由 $\hat{\mu}_N$ 和 $\hat{\sigma}_N$ 确定如式(2)所示。

$$\mu_w \leftarrow \hat{\mu}_N, \sigma_w \leftarrow \hat{\sigma}_N \quad (2)$$

3)对于 N 个像素点的权重 a_n 通过高斯核函数确定,如公式所示。像素点平均值 μ_x 和标准差 σ_x 由高斯权重 a_n 加权所得,如式(3)和(4)所示。

$$a_n = \frac{1}{A} p(x_n; \mu_w, \sigma_w), A = \sum_{k=1}^N p(x_k; \mu_w, \sigma_w) \quad (3)$$

$$\mu_x = \sum_{n=1}^N a_n x_n, \sigma_x = \sqrt{\sum_{n=1}^N a_n (x_n - \mu_x)^2} \quad (4)$$

4)迭代更新 $\hat{\mu}_N$ 和 $\hat{\sigma}_N$, 如式(5)所示。当值更新小于某预定值时停止,否则返回步骤 2)并继续更新。

$$\hat{\mu}_N \leftarrow \mu_x, \hat{\sigma}_N \leftarrow \sqrt{\frac{\sigma_w \sigma_x^2}{\sigma_w^2 - \sigma_x^2}} \quad (5)$$

根据经验阈值和主聚类推定的亮度平均值二者之间的大小关系,将该机车 6A 司机室视频图像分类为低光照、正常光照或者曝光 3 个场景。

2.2 自适应非线性彩色增强算法

面对外界宽动态范围的场景,人眼拥有亮度自适应感知机制,通过收缩瞳孔的大小调节进入眼球的光线量。视网膜对亮度、边缘和颜色等信息进行动态范围压缩处理并完成光电信号的转换,视神经细胞将电信号传递到视觉皮层形成视觉^[19]。自适应非线性彩色增强算法(adaptive non-linear color enhancement, ANCE)^[20]可以很好地模拟人类视觉处理光信号的过程,将亮度增强、对比度增强和颜色恢复分开独立处理,且相互之间达到平衡,同时考虑图像全局和局部的信息,能够更加自适应地增强图像质量。

1) 自适应亮度增强

将 RGB 三通道彩色图像通过灰度变换转换成灰度图像并作归一化处理,如式(6)和(7)所示。

$$I(x, y) = 0.299I_R(x, y) + 0.587I_G(x, y) + 0.114I_B(x, y) \quad (6)$$

$$I_n(x, y) = \frac{I(x, y)}{255} \quad (7)$$

式中: $I_R(x, y)$ 、 $I_G(x, y)$ 和 $I_B(x, y)$ 分别表示图像 RGB 3 个通道在 (x, y) 位置处的像素值, $I(x, y)$ 和 $I_n(x, y)$ 分别表示灰度图像和归一化图像的像素值。然后对归一化灰度图像进行自适应全局非线性亮度增强,其映射关系如式(8)所示。

$$I'_n = \frac{I_n^{(0.75P+0.25)} + 0.4(1-P)(1-I_n) + I_n^{(2-P)}}{2} \quad (8)$$

式中: I'_n 表示亮度增强后的灰度像素值,亮度增强因子 P 的定义如式(9)所示。

$$P = \begin{cases} 0, & L \leq 50 \\ \frac{L-50}{100}, & 50 < L \leq 150 \\ 1, & L > 150 \end{cases} \quad (9)$$

式中: L 表示在图像的灰度累积直方图中频率为 0.1 所对应的灰度值,即用 10% 的像素来评判图像黑暗的程度。动态调整参数 P 可以自适应全局地提高图像的亮度,即图像较暗区域的亮度值大幅提高,图像较亮区域的亮度值小幅提高。这样能够有效防止过大提高图像的亮度,达到压缩图像动态范围的目的。

2) 自适应对比度增强

自适应全局非线性亮度增强使图像所有像素值均得到了一定程度的提高,没有利用领域像素点的信息,容易出现对比度下降的问题。局部对比度增强根据中心点和领域像素值之间的大小关系来改变中心点像素值,即如果邻域像素值更亮,那么中心点像素值会被降低,反之则提高。局部对比度增强不仅能够加强图像的细节信息和提高图像的对比度,还可以抑制动态范围的扩大。人眼认知外界图像的视觉特性和高斯分布比较相近,通过二维高斯离散空间卷积获取邻域像素的亮度信息。二维高斯函数如式(10)所示。

$$G(x, y) = K e^{-\frac{x^2+y^2}{c^2}} \quad (10)$$

$$\iint K e^{-\frac{x^2+y^2}{c^2}} dx dy = 1 \quad (11)$$

式中:参数 c 表示高斯核的尺度。首先对尺寸大小 $M \times N$ 的图像做高斯卷积运算,如式(12)所示。

$$I_{\text{Conv}}(x, y) = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} G(m+x, n+y) I(m, n) \quad (12)$$

然后对图像进行自适应对比度增强处理,如式(13)所示。

$$T(x, y) = 255 I'_n(x, y)^{H(x, y)} \quad (13)$$

$$H(x, y) = \left(\frac{I_{\text{Conv}}(x, y)}{I(x, y)} \right)^Q \quad (14)$$

式中: $H(x, y)$ 表示对比度增强指数, $I_{\text{Conv}}(x, y)$ 表示灰度图像卷积像素值。对比度增强因子 Q 由灰度图像 $I(x, y)$ 的标准差 σ 决定, σ 越小意味着图像的对比度越弱,增强因子 Q 相应越大,自适应地调节对比度增强效果,如式(15)所示。

$$Q = \begin{cases} 4, & \sigma \leq 4 \\ -0.5\sigma + 6, & 4 < \sigma < 10 \\ 1, & \sigma \geq 10 \end{cases} \quad (15)$$

如果中心点像素值高于邻域,卷积运算之后像素值变小,即 $I_{\text{Conv}}(x, y)/I(x, y)$ 值小于 1。参数 Q 的取值范围

大于 1,故对比度增强指数 $H(x,y) < 1$ 。自适应亮度增强灰度图像 I'_n 像素值的取值范围为 $[0,1]$, 所以 $I'_n(x,y)^{H(x,y)} < I'_n(x,y)$, 表明经过自适应对比度增强之后,中心点的像素值会降低。反之如果中心点像素值低于邻域则被提高。自适应对比度增强通过对对比度增强因子 Q 可以有效地提高图像的对比度。

小尺度卷积可以提供领域像素的亮度、对比度等细节信息;大尺度卷积关注全局信息,卷积图像更加平滑和自然;中等尺度卷积可以同时提供图像全局和局部信息。将各个尺度的卷积滤波结果线性相加得到多尺度卷积滤波图像,其信息更加丰富和完整,对比度增强效果更加均衡,如式(16)~(20)所示。

$$G_i(x,y) = Ke^{-\frac{x^2+y^2}{c_i^2}} \quad (16)$$

$$I_{\text{Conv},i}(x,y) = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} G_i(m+x,n+y)I(m,n) \quad (17)$$

$$H_i(x,y) = r_i(x,y)^Q = \left(\frac{I_{\text{Conv},i}(x,y)}{I(x,y)}\right)^Q \quad (18)$$

$$T_i(x,y) = 255I'_n(x,y)^{H_i(x,y)} \quad (19)$$

$$T(x,y) = \sum_i w_i T_i(x,y), w_i = \frac{1}{n}, i = 1,2,\dots,n \quad (20)$$

式中: c_i 表示不同的尺度, w_i 表示各个对比度增强输出结果的相应权重, n 表示卷积尺度个数。根据经验一般选用 3 个尺度,即 $w_i = 1/3$, 所得到增强图像的视觉效果比较自然。小卷积尺度的取值范围选用图像宽度和高度最小值的 1%~5%, 中等尺度和大尺度分别选用 10%~15% 和 25%~45% 较为合理。

3) 彩色恢复

RGB 三通道图像经过自适应亮度和对比度增强之后,需要对色调信息进行彩色线性恢复,如式(21)所示。

$$T_j(x,y) = T(x,y) \frac{I_j(x,y)}{I(x,y)} \gamma_j, j = R,G,B \quad (21)$$

式中: $j = R,G,B$ 分别表示 RGB 3 个通道,参数 γ_j 表示彩色增强因子,该参数小于且非常接近 1。当所有的 γ_j 都为 1 时,可以较大程度地保持原图像的色彩信息,图像色偏现象最小。

2.3 侧窗滤波

窗口滤波是图像降噪和去模糊的常用技术,例如均值滤波、中值滤波和高斯滤波等。传统做法是将滤波窗口的中心点和被滤波像素点重合,滤波半径决定窗口尺寸的大小。像素的滤波值由窗口内所有的像素点通过线性加权得到,如式(22)所示。

$$I'_i = \sum_{j \in \Omega_i} \omega_{ij} q_j \quad (22)$$

式中: Ω 表示以像素点 i 为中心的滤波窗口内像素点的集合, j 表示滤波窗口中的像素点。 ω_{ij} 和 q_j 分别表示滤波权重值和像素值。滤波算法本质是估计出最优的核权重 ω_{ij} ,

旨在滤波效果和损失二者之间取得平衡。图像滤波的损失函数采用二范计算,如式(23)所示。

$$S = \|I_i - I'_i\|_2^2 = (I_i - \sum_{j \in \Omega_i} \omega_{ij} q_j)^2 \quad (23)$$

图像在位置 (x,y) 处的像素值用 $f(x,y)$ 表示,位置 m 左右两侧区域 ' $m-$ ' 和 ' $m+$ ' 分别用左极限 $(x-\epsilon,y)$ 和右极限 $(x+\epsilon,y)$ 表示,其中 $\epsilon > 0$ 。像素值在边缘处由于发生跳跃而不连续并且不可导,边缘左右两侧区域的像素值和其导数均不相等,即 $f(x-\epsilon,y) \neq f(x+\epsilon,y)$ 且 $f'(x-\epsilon,y) \neq f'(x+\epsilon,y)$ 。对边缘 m 左右两侧区域的像素点进行泰勒公式展开,如式(24)~(25)所示。

$$f(x-2\epsilon,y) \approx f(x-\epsilon,y) + f'(x-\epsilon,y)(-\epsilon) \quad (24)$$

$$f(x+2\epsilon,y) \approx f(x+\epsilon,y) + f'(x+\epsilon,y)\epsilon \quad (25)$$

发现 ' $m-$ ' 和 ' $m+$ ' 区域的滤波估计分别来自边缘 m 的左边和右边的。传统滤波处理之后,边缘必定在一定程度上被平滑,导致边缘信息丢失,需要重新考虑滤波窗口的样式设计。文献[21]提出了一种新型侧窗滤波(side window filtering, SWF)策略,首先将每一个像素点都当作是潜在的边缘点,然后在其周边生成几种不同的滤波侧窗口,侧窗的定义如图 2 所示,参数 θ 表示窗口与水平线的夹角, r 为窗口半径, ρ 为控制窗口长度的另一参数且 $\rho \in [0,r]$ 。通过比较和评估不同侧窗口处理后的效果最终选出最佳的侧窗。

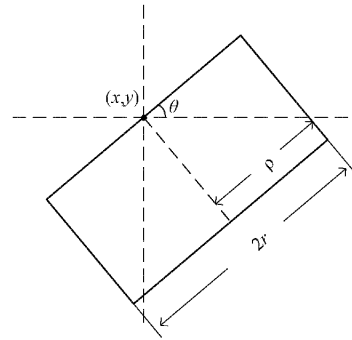


图 2 侧窗示意图

为了保证各个滤波侧窗口的边缘和被滤波的像素点对齐,同时计算快速且简单,本文定义了 4 个滤波侧窗口,方向为 $\theta = k \times \pi/2, k \in [0,3]$,半径 $\rho = r$,按照逆时针方向共有下、右、上和左 4 个侧窗口,分别标记为 $\Theta_i^D, \Theta_i^R, \Theta_i^U$ 和 Θ_i^L ,如图 3 所示。

侧窗滤波的算法详细步骤如下:

1) 在被滤波的目标像素点 i 的侧窗中,计算单个侧窗基于核函数 F 的权重值之和:

$$W_n = \sum_{j \in \Theta_i^n} \omega_{ij}, n \in S \quad (26)$$

式中: $S = \{D,R,U,L\}$ 表示侧窗索引的集合, Θ_i^n 表示像素点 i 周围所有的侧窗, ω_{ij} 表示目标像素点 i 的领域像素点 j 所对应的基于核函数 F 的权重值。

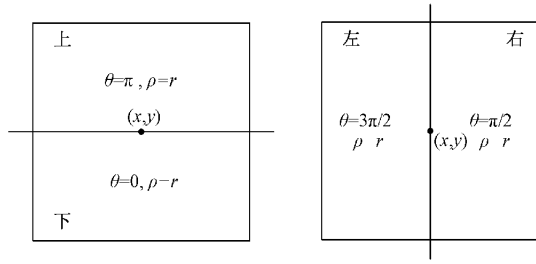


图 3 像素点的 4 个滤波侧窗口

2) 在每一个侧窗中, 计算目标像素点 i 的滤波像素值:

$$I_i^{\theta, \rho, r} = \frac{1}{W_n} \sum_{j \in \Theta_i^{\theta, \rho, r}} \omega_{ij} q_j \quad (27)$$

$$n \in S, \theta = k \times \frac{\pi}{2}, k \in [0, 3], \rho = r$$

式中: q_j 表示目标像素点的领域像素 j 的像素值, θ 表示滤波侧窗的方向角度, ρ 表示滤波侧窗的半径。

3) 利用 2 范数定义像素点滤波效果的代价函数, 遍历所有的侧窗求其最小化的 \hat{I}_i , 优化函数如下:

$$\hat{I}_i = \operatorname{argmin}_{I_i^{\theta, \rho, r}} \| q_i - I_i^{\theta, \rho, r} \|_2^2 \quad (28)$$

式中: \hat{I}_i 表示像素点 i 的最优滤波像素值, $\| \cdot \|_2^2$ 表示 2 范的平方。

将侧窗滤波算法和自适应非线性彩色增强算法相融合, 即得到本文所提出的基于侧窗滤波的自适应非线性彩色增强算法。本文所提算法在亮度增强、对比度增强、色彩恢复 3 个方面, 关注图像全局和局部的信息, 保留了图像边

缘的细节信息, 可以有效地提高图像质量。

3 基于 YOLOv3 的司机行为识别模型

基于候选区域 Region Proposals 的 R-CNN 系列目标检测算法^[22-24]首先产生潜在的目标边框, 然后对边框进行分类, 最后再对边框精修消除重复检测的问题。这种二阶段的检测算法会产生大量假阳的目标检测框, 消除假阳检测时消耗了巨大的计算资源, 导致算法的耗时比较长, 不利于在工业检测领域中部署应用。YOLO 算法^[25]首次实现单阶段目标检测, 它将目标检测问题转化为回归问题, 直接在输出层回归出目标包围框 Bounding box 的类别和位置, 实时性比较好, 具有较强的工程应用价值, 因此被广泛运用于工程领域。YOLOv3 采用全卷积网络结构, 包括骨干网络 Darknet-53 和检测网络, 是一种基于端到端的检测算法, 如图 4 所示。Darknet-53 网络引入了残差结构模块 Residual Block^[26], 按照相应的维度将输入和输出特征图进行相加, 即 $y = f(x) + x$, 能够更好地解决梯度消失的问题。另外还借鉴了 DenseNet 网络^[27]设计的思路, 按照通道的维度将不同的特征图直接进行拼接, 具有更为强大的特征提取能力。检测网络借鉴特征金字塔网络 FPN^[28]的思想, 小尺寸特征图用于检测大尺寸物体, 大尺寸的特征图用于检测小尺寸物体, 共计输出 3 个尺度的特征图, 然后通过多尺度特征融合的方式对目标进行预测。YOLOv3 利用多个简单的逻辑回归 sigmoid 函数输出类别得分, 用二元交叉熵损失函数计算类别损失, 每个分类器对目标框中出现的物体判断是否属于当前标签。

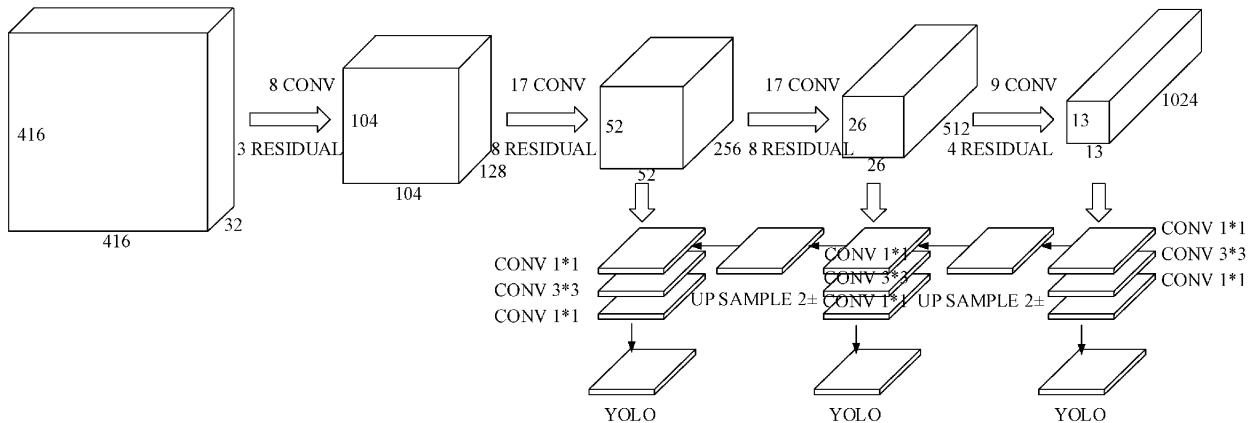


图 4 YOLOv3 网络结构

本文利用 YOLOv3 建立司机行为识别模型, 将每个尺度特征图划分的网格数为 $S \times S$ 个, 每个网格预测出数量为 B 、置信度为 C 的包围框 Bounding-box。包括框的参数用 (x, y, w, h) 表示, 其中 x, y 分别表示包围框的横纵坐标, w, h 分别表示宽度和高度。每个网格同时预测出 N 个目标对象的概率 P , 每个预测张量输出的维度为 $S \times S \times [B \times (5 + N)]$ 。损失函数 $Loss$ 为中心坐标误差 $dist_{xy}$ 、

宽高坐标误差 $dist_{wh}$ 、置信度误差值 $dist_C$ 与分类误差 $dist_P$ 总和的均值, 其中 $dist_{wh}$ 为平方差函数, $dist_{xy}, dist_C$ 和 $dist_P$ 为交叉熵函数。Loss 损失函数如式 (29) 所示。

$$Loss = \frac{1}{N} (dist_{xy} + dist_{wh} + dist_C + dist_P) \quad (29)$$

$$dist_{xy} = \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{obj} \{ [x_i^j \log(\hat{x}_i^j)] + (1 - x_i^j) \log(1 - \hat{x}_i^j) + [y_i^j \log(\hat{y}_i^j)] + (1 - y_i^j) \log(1 - \hat{y}_i^j) \} \quad (30)$$

$$dist_{wh} = \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{obj} [(w_i^j - \hat{w}_i^j)^2 + (h_i^j - \hat{h}_i^j)^2] \quad (31)$$

$$dist_C = \sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{obj} [C_i^j \log(\hat{C}_i^j) + (1 - C_i^j) \log(1 - \hat{C}_i^j)] + \lambda_{\text{noobj}} \sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{noobj} [C_i^j \log(\hat{C}_i^j) + (1 - C_i^j) \log(1 - \hat{C}_i^j)] \quad (32)$$

$$dist_P = \sum_{i=0}^{S^2} I_{ij}^{obj} \sum_{c \in \text{classes}} [P_i^j \log(\hat{P}_i^j) + (1 - P_i^j) \log(1 - \hat{P}_i^j)] \quad (33)$$

式中： \hat{x}_i^j 、 \hat{y}_i^j 、 \hat{w}_i^j 、 \hat{h}_i^j 、 \hat{C}_i^j 和 \hat{P}_i^j 分别为 YOLOv3 模型输出的第 i 个网格第 j 个包络框的中心横坐标、中心纵坐标、宽度、高度、置信度和分类概率的预测值， x_i^j 、 y_i^j 、 w_i^j 、 h_i^j 、 C_i^j 和 P_i^j 分别为相应的真实值。 I_{ij}^{obj} 表示第 i 个网格中的第 j 个包络框是否负责这个目标对象，包络框是否负责通过包络框和真实方框的交并比 (intersection over union, IoU) 来衡量计算。如果负责， I_{ij}^{obj} 取值则为 1，否则取值为 0。 λ 表示权重，为了扩大含有目标对象的网格损失值，本文中设置 $\lambda_{\text{coord}} = 5$ ，同时减小不含有目标对象的网格损失值，设置 $\lambda_{\text{noobj}} = 0.5$ ，这样便于网络模型的收敛。

4 试验分析

4.1 试验环境

实验室视频分析服务器司机行为识别系统搭载在实验室视频分析服务器平台上，硬件方面选用 125 G RAM、NVIDIA Tesla T4 GPU 和 2.10 GHz Intel(R) Xeon(R) Silver 4110，软件方面选用 Ubuntu16.04(64 位)操作系统，CUDA 版本为 10.0，CuDNN 版本为 10.0，利用 Pytorch 深度学习框架搭建深度神经网络模型。

从铁路局机务段现场部署的机车联动分析系统下载 6A 司机室视频，每个视频时长约为 15 min，分辨率为 352×288 ，大小约为 40 M。然后视频抽帧，并根据相关 LKJ 信息制作顶点的图像训练和测试数据集。用于深度神经网络模型训练的图像训练集涵盖了白天、夜晚和进出隧道等场景，共计 66 196 张，可以更好地提高深度神经网络模型的准确率和鲁棒性。根据主聚类推定照度分类算法^[16]计算图像的亮度值，得出图像训练数据集的亮度范围分布在 0.014~0.816。根据经验和统计学规律，照度分类的阈值设置为 0.17 和 0.82，即低于 0.17 为低光照场景，高于 0.82 为曝光图像，其它的为正常光照图像。制作低光照图像测试集，样本数量共计 1 073 张，用于评价本文所提

出的低照度图像增强算法的性能优劣。

在网络模型的训练过程中，将输入图片的 batch 和 subdivisions 分别设置为 64 和 16。使用随机梯度下降 (stochastic gradient descent, SGD) 优化算法训练模型，动量 (momentum) 参数设置为 0.9，保证网络模型有更好的最优解同时收敛的速度也比较快。其他的超参数学习率 (learning rate) 设置为 0.001，权值衰减 (weight decay) 设置为 0.000 5。在搭载 4 块 NVIDIA Tesla T4 的视频分析服务器平台上进行多卡并行训练，训练的迭代次数设置为 30 000 次。

4.2 图像增强试验分析

峰值信噪比 (peak signal to noise ratio, PSNR) 基于像素点之间的误差敏感度进行评价，具有一定的局限性。结构相似性 (structural similarity index measure, SSIM) 综合考量图像亮度、对比度和结构的 3 个关键信息，更加符合人眼的视觉特性^[29]。本文采用 PSNR 和 SSIM 指标，同时兼顾人眼直接的主观感受，对图像的增强效果进行综合评价。

将本文所提算法与局部自适应直方图增强 (adaptive histogram equalization, AHE)、同态滤波 (homomorphic filtering, HF)、多尺度 Retinex (multi-scale retinex, MSR)、基于 GAN 的低照度增强算法 (enlighten GAN, EGAN)、ANCE 算法进行对比实验，图像增强的结果如图 5 所示。本文所提算法的增强效果如图 5(g) 所示，在左侧的小矩形注释框内图像色彩得到了恢复，在右侧的大矩形注释框内图像的亮度和对比度都得到了增强。PSNR 和 SSIM 指标评价的平均值如表 1 所示。

PSNR 和 SSIM 指标值越大，说明增强前后的差异性越小，二者之间越相似。在一定范围内指标值越小，说明对原始低照度图像的增强效果越好。AHE 算法针对图像的局部对比度进行提高图像质量，增强模式较为单一，可以看出该算法的 PSNR 和 SSIM 指标值是最大的。HF 算法在频域中通过压缩亮度范围和增强对比度来改善图像质量，可以更好地光照不均匀的问题，增强暗处细节信息。但是 HF 算法没有考虑图像色彩方面信息，所以 PSNR 和 SSIM 值只是比 AHE 低一些，均高于其它剩余的算法。MSR 算法将不同尺度的增强结果线性相加组合，同时考虑局部信息和全局信息，在动态范围压缩和色调恢复方面更加均衡。Retinex 理论假设图像光照变化是平滑的，因此对于区域亮度差异较大的增强图像产生光晕，颜色有时会发生扭曲和失真，纹理不清晰。虽然 MSR 的评价指标值最小，但是人眼观察其增强图像时，发现色彩发生严重失真，细节信息反而变得模糊。EGAN 增强算法利用生成网络和判别网络的不断博弈，生成尽可能接近现实的‘伪造’的正常光照图像，其增强效果优于 AHE、HF 和 MSR 传统算法。GAN 训练容易出现模型不收敛、模式崩溃、过拟合和高度敏感的超参选择等问题，此类无监督学习的方

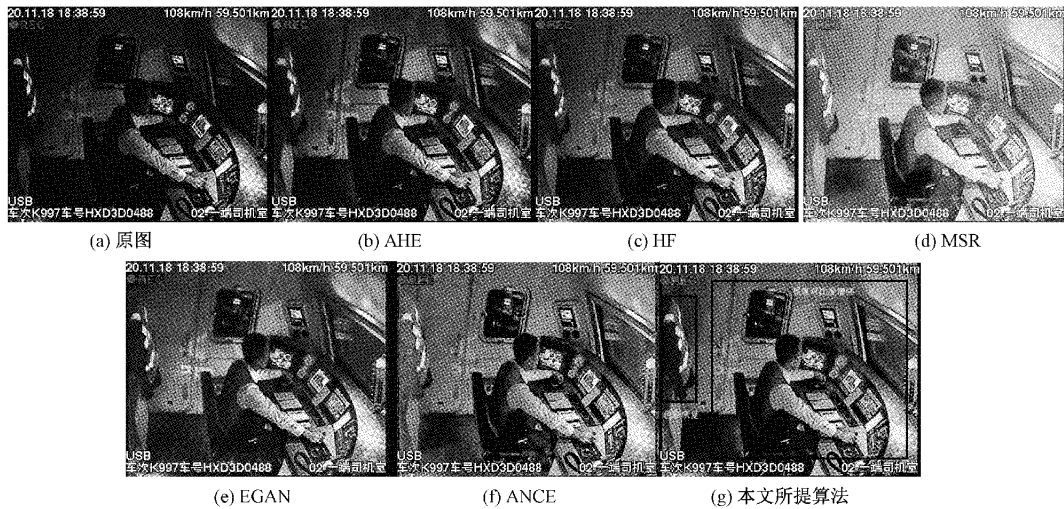


图 5 AHE、HF、MSR、EGAN、ANCE 和本文所提算法的处理结果对比

表 1 不同图像增强算法的 PSNR 和 SSIM 对比

| 算法 | AHE | HF | MSR | EGAN | ANCE | 本文所提算法 |
|------|--------|--------|-------|--------|--------|--------|
| PSNR | 19.125 | 16.381 | 6.725 | 13.248 | 13.012 | 12.473 |
| SSIM | 0.767 | 0.695 | 0.363 | 0.609 | 0.612 | 0.577 |

法研究还不太完善,导致图像增强的效果不能完全令人满意。ANCE 算法在图像亮度增强和对比度增强两个方面达到较好的平衡,能够获取并利用图像的全局和局部信息,但是输出图像存在色偏,色彩恢复不佳。ANCE 与侧

窗滤波算法相结合可以更好地保留边缘等细节信息。因此从指标和视觉感受综合来看,本文所提算法的增强效果达到了最优。

4.3 项点检测试验分析

项点检测的试验流程包括图像增强和目标检测两个阶段。第 1 阶段采用上述 6 种算法对图像进行增强处理,第 2 阶段利用 YOLOv3 对增强后的低照度图像进行目标检测,完成离岗项点的识别任务。通过在低光照图像测试集中做测试,对比试验的准确率结果如表 2 所示,‘/’表示对低照度图像未使用图像增强算法处理。

表 2 不同算法的离岗项点检测准确率对比

| 第 1 阶段图像增强 | / | AHE | HF | MSR | EGAN | ANCE | 本文所提算法 |
|------------|-------|-------|-------|--------|-------|-------|--------|
| 第 2 阶段目标检测 | | | | YOLOv3 | | | |
| 准确率/% | 90.87 | 93.67 | 94.48 | 86.34 | 96.46 | 95.92 | 97.20 |

由表 2 可以看出,采用 MSR 对低照度图像增强后,项点准确率与未增强相比反而下降了 4.53%。原因可能是 MSR 导致图像出现色偏和纹理不清晰,不利于 YOLOv3 的目标检测。AHE 和 HF 在图像对比度和亮度方面改善了低照度图像质量,与未增强相比项点准确率提高了 3% 左右。EGAN、ANCE 和本文所提算法可以从图像的亮度、对比度、纹理等方面,较为全面地提高低照度图像的质量,与未增强相比项点准确率提高了 6% 左右。然而 EGAN 算法受样本集的影响较大,训练过程有时出现模式崩溃的现象、模型不收敛以及过拟合,不利于后期的迭代优化,而且生成的增强图像有些细节信息会受到破坏,准确率与 ANCE 算法接近。ANCE 对低照度图像进行增强时,能够将图像全局和局部信息相结合,自适应地增强亮度和对比度,但输出图像存在明显色偏。本文所提算法在 ANCE 的基础上维持图像亮度、对比度增强效果的同时,

色彩恢复较好,有效增强图像暗区细节,有利于神经网络模型提取目标特征,因此利用该增强算法处理后项点的准确率最高,达到了 97.20%。

5 结 论

本文提出了一种新型基于侧窗的自适应非线性彩色增强算法,不仅可以有效地利用图像全局和局部相融合的信息,而且更好地保留了图像边缘等细节信息,在图像亮度和对比度方面能够自适应地增强,同时色彩饱和度未明显下降,避免了色偏现象,图像质量达到了最优。

针对在低光照场景下项点检测精度较低的问题,本文首先对低照度图像进行增强处理,然后使用 YOLOv3 算法对机车乘务员的驾驶行为进行检测,项点的检测精度与未增强相比提升 6.33%,达到了 97.20%。本文所设计的新

型司机行为识别系统具有良好的检测效果,满足铁路局机务段视频智能分析的需求,促进乘务员作业执行标准的规范化,提升铁路运输安全的保障能力。

本文所设计的图像增强算法可能存在色彩增益因子过大导致彩色恢复过程出现色觉异常的情况,后期可以对该参数进行定量分析和评估,尝试对色彩增益因子进行自适应调整。而对于本文提出的司机行为识别系统整体方案,未来需进一步提高系统运行速度和效率,以降低机务段视频分析服务器的部署成本。可考虑利用 CUDA 并行运算平台对图像增强算法进行加速优化,采用高性能 TensorRT 推理器提高基于 YOLOv3 的司机行为识别模型的推理速度。

参考文献

- [1] 徐宏. 铁路职工工作压力测评与应用研究[D]. 北京: 北京交通大学, 2017.
- [2] 谢小军. 机车车载安全防护系统(6A 系统)数据分析及故障判断处理方法[J]. 科技创新与应用, 2017(5): 56-57.
- [3] 翁建辉. LKJ 设备运行监测系统的技术实现及运用[J]. 铁道通信信号, 2019, 55(1):6-9.
- [4] 张索非, 冯焯, 吴晓富. 基于深度卷积神经网络的目标检测算法进展[J]. 南京邮电大学学报(自然科学版), 2019, 39(5):72-80.
- [5] DHILLON A, VERMA G K. Convolutional neural network: A review of models, methodologies and applications to object detection [J]. Progress in Artificial Intelligence, 2020, 9(2):85-112.
- [6] 姚巍巍, 田野, 李博. 基于改进 SOLO 的列车主动避障视觉算法研究[J]. 电子测量技术, 2022, 45(9):133-139.
- [7] 董雅文. 基于深度学习的地铁司机操作行为识别技术研究[D]. 大连:大连交通大学, 2020.
- [8] 胡超. 时空增强网络研究及其在列车司机行为识别的应用[D]. 成都:西南交通大学, 2019.
- [9] 姚巍巍. 基于深度卷积神经网络的铁路司机行为识别技术研究[D]. 成都:西南交通大学, 2020.
- [10] DENG J, DONG W, SOCHER R, et al. ImageNet: A large-scale hierarchical image database [C]. 2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2009), 20-25 June 2009, Miami, Florida, USA. IEEE, 2009.
- [11] GONZALEZ R, WOODS R. Digital image processing[M]. New Jersey: Prentice Hall, 2012:144-150.
- [12] GONZALEZ R, WOODS R. Digital image processing[M]. New Jersey: Prentice Hall, 2012:311-315.
- [13] FUNT B, CIUREA F, MCCANN J. Retinex in Matlab[C]. Color & Imaging Conference, 2000.
- [14] 郭春乐. 基于深度学习的低质图像增强和复原算法研究[D]. 天津:天津大学, 2020.
- [15] ZHU J Y, PARK T, ISOLA P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks [C]. Proceedings of the IEEE International Conference on Computer Vision, 2017: 2223-2232.
- [16] LI C Y, GUO J H, GUO C L, et al. Emerging from water: Underwater image color correction based on weakly supervised color transfer [J]. IEEE Signal Processing Letters, 2018, 25(3): 323-327.
- [17] REDMON J, FARHADI A. YOLOv3: An incremental improvement[J]. ArXiv Preprint, 2018, ArXiv:1804.02767.
- [18] TIAN Y, YOKOTA Y. Estimating the major cluster by mean-shift with updating kernel[J]. Mathematics, 2019, 7(9), DOI:10.3390/math7090771.
- [19] NEWMAN, NANCY J. The visual neurosciences[J]. Journal of Neuro Ophthalmology, 2005, 25(4):337.
- [20] LI T, ASARI V. Adaptive and integrated neighborhood-dependent approach for nonlinear enhancement of color images [C]. International Conference on Information Technology: Coding & Computing. IEEE Computer Society, 2004.
- [21] YIN H, GONG Y H, QIU G P. Side window filtering[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019; 8758-8766.
- [22] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014; 580-587.
- [23] GIRSHICK R. Fast r-cnn [C]. Proceedings of the IEEE International Conference on Computer Vision, 2015: 1440-1448.
- [24] REN S Q, HE K M, GIRSHICK R, et al. Faster r-cnn: Towards real-time object detection with region proposal networks [C]. Advances in Neural Information Processing Systems, 2015; 91-99.
- [25] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection[J]. CoRR, 2015, DOI:abs/1506.02640.
- [26] HE K M, ZHANG X Y, REN S P, et al. Deep residual learning for image recognition [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016; 770-778.
- [27] GAO H, LIU Z, MAATEN L, et al. Densely connected convolutional networks[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 4700-4708.
- [28] LIN T G, DOLLAR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017:2117-2125.
- [29] HORE A, ZIOU D. Image quality metrics: PSNR vs SSIM [C]. 20th International Conference on Pattern Recognition, ICPR 2010 IEEE Computer Society, 2010.

作者简介

崔宵洋, 硕士, 助理工程师, 主要研究方向为图像信号处理、计算机视觉和铁路信号检测技术, 以及轨道交通领域信息与智能化等。
E-mail: shiaoyoungcui@outlook.com