

基于改进 YOLO v3 的相似外部特征人员检测算法

梁思源, 王 平, 罗凡波, 徐桂菲, 王 伟
(西华大学 电气与电子信息学院, 四川 成都 610039)

摘 要:针对目前异常行为检测中相似特征人员检测方法稀缺、人员和特征检测准确度低、特征量少,而相似外部特征往往意味着团队行动及潜在异常行为等问题,提出一种基于深度学习的相似外部特征人员检测算法。首先采用加入 Fast Guided Filter 的暗通道去雾算法对 INRIA 数据库图像进行前期处理,得到质量更佳训练样本;然后用得到的样本对改进的 YOLO v3 进行训练;最后将提取出来的行人进行颜色特征和几种纹理特征提取,组合之后用 ELM 进行分类。仿真结果表明:加入 Fast Guided Filter 的暗通道去雾算法明显优于单纯的暗通道去雾算法,保留了更多的边缘和纹理特征,在雾天和强曝光下效果尤为明显。相比 HOG + SVM 方法,该算法对人员检测的误检率和漏检率都大大降低,且具有较好的实时性。最后 ELM 分类的准确性能达到 96.104%。

关键词:暗通道去雾 YOLO 网络;颜色特征;纹理特征;极限学习机

中图分类号:TP391.41

文献标识码:A

文章编号:1673-1670(2020)02-0047-08

0 引言

计算机视觉是指用计算机实现人的视觉功能,希望能根据感知到的图像(视频)对实际的目标和场景内容做出有意义的判断^[1]。随着计算机视觉技术的发展,智能视频监控已经成为一个热门的研究领域。近年来计算机硬件技术尤其是 GPU 技术的飞速发展,使得深度学习已经在视频监控领域占据了不可替代的位置。但是现在针对相似外部特征人员的检测算法还相对较少,而公共场所多个相似特征人员的聚集往往预示着异常情况的发生。就现状笔者针对大雾和强曝光天气下的外部特征相似人员提出一种基于深度学习的检测算法。

深度学习不需要人工对目标的特征进行设计和提取,而是由机器自动地学习获得。这种自动学得特征具有很好的泛化能力与鲁棒性^[2]。但是深度学习的分类结果受训练样本的影响较大,尤其是在大雾和强曝光之下,图像的特征丢失严重,会导致最后的分类效果并不理想。因此笔者对自建数据库中质量不佳的训练样本进行了暗通道去雾算法处理,得到更优质的样本模型进行训练^[3]。该算法是一种实用性很强的算法,该算法作者还在

2015 年提出了 Fast Guided Filter,笔者采用结合了 Fast Guided Filter 的暗通道去雾算法,实验表明该算法在去雾的同时保留了更多的目标人物的边缘细节,为后期的图像特征提取也保留了很多有价值的特征信息。

行人检测作为目标检测的一个重要分支,经过几十年的发展已经在检测的速度和精度上取得了很大的进步,很多方法已经完全能够达到视屏级的实时性。传统方法采用人工设计的特征提取器来提取特征并对分类器进行训练。例如通过提取 Haar 特征、HOG(Histogram of Oriented Gradient)特征、LBP(Local Binary Pattern)特征等进行训练。其中文献[4]采用 HOG + LBP 特征处理行人的遮挡问题,提高了检测的准确率。但由于人体的柔性特征和人群中的遮挡问题,人工设计的行人特征很难适应行人的大幅度变化^[5]。笔者采用经过去雾处理的自建数据集对改进的 YOLO v3 网络进行训练,并将得到的结果与 HOG + SVM 方法进行对比,结果证明在实时性和准确性方面该方法均优于传统的 HOG + SVM 方法。

提取出目标人员之后再对其进行特征的提取。很多相关研究已经在颜色、形状和纹理特征这几方

收稿日期:2019-06-17

基金项目:教育部“春晖计划”资助项目(Z2012029);四川省信号与信息处理重点实验室开放基金资助项目(szjj2012-015);西华大学研究生创新基金资助项目(ycyj2018025)

作者简介:梁思源(1993—),男,四川省广安市人,西华大学电气与电子信息学院硕士研究生,主要从事计算机视觉、模式识别研究。



面展开,其中效果相对较好的特征主要为颜色与纹理的特征.常见的颜色空间有 RGB 颜色空间、HSV 空间和 YUV 空间.目前已有方法对 RGB 空间分量和 YUV 空间分量的特点和优势进行综合并进行特征提取^[6].纹理特征经过多年的发展,已经发展出灰度共生矩阵(CLCM)、灰度行程长度法(gray level run length)、自相关函数法等,同时随着应用领域的不断扩大和新理论如分形理论、马尔科夫随机场(MRF)理论、小波理论等的引入,又不断壮大了纹理特征提取的研究^[7].但是由于提取出的特征往往具有较高的维度,因此处理这些数据所消耗的时间可能会较长,对于视屏监控的实时性不能满足.但如果为了提高处理速度而降低特征维度又会导致错误率的增加,不能适应复杂的场景,抗噪能力不足.

笔者基于人类视觉特征对图像感知的 3 个特性提出自适应颜色特征提取方法,颜色特征上使用 RGB 和 HSV 分别描述,因此在降低特征维度的情况下依然能获得较好的效果,并在此基础上使用灰度直方图、灰度共生矩阵、小波变换来获取多个新的纹理特征,用于更好地检测外部相似特征.最后将获取的特征放入极限学习机(ELM)中进行分类.实验结果表明本文算法能在个体目标相似度上取得较好效果,优于传统的相似度匹配算法.

1 暗通道去雾处理

为了提高图像的质量,减小大雾天气和强曝光天气对训练所造成的影响,提高检测的准确度和鲁棒性,首先对得到的数据集进行暗通道去雾处理.

1.1 暗通道定义

暗通道是指在图片中,除去天空、海洋等大面积单色区域,一些像素会出现一个或多个颜色通道的值很低的现象.对于一幅输入图像 J , 它的暗通道可以用式(1)表达:

$$J^{dark}(x) = \min_{y \in \Omega(x)} (\min_{c \in \{r, g, b\}} J^c(y)). \quad (1)$$

其中 J^c 表示彩色图像的每个通道, $\Omega(x)$ 表示以像素 x 为中心的一个窗口. r, g, b 表示图像的 3 个通道^[8].暗通道先验的理论指出:

$$J^{dark} \rightarrow 0. \quad (2)$$

1.2 暗通道去雾原理

在计算机视觉中常用公式(3)来表示雾图形

成模型.

$$I(x) = J(x)t(x) + [A(1 - t(x))]. \quad (3)$$

式中, $I(x)$ 表示现有未去雾图像, $J(x)$ 表示要得到的去雾后图像, A 表示全球大气光成分, $t(x)$ 表示透射率.为得到方程的解,将(3)式变化为(4)式.

$$\frac{I^c(x)}{A^c} = t(x) \frac{J^c(x)}{A^c} + 1 - t(x). \quad (4)$$

这里的 c 表示 RGB 三个通道,为了得到我们定义的暗通道,求 c 的最小值得到式(5):

$$\min_{y \in \Omega(x)} \left(\min_c \frac{J^c(y)}{A^c} \right) =$$

$$\tilde{t}(x) \min_{y \in \Omega(x)} \left(\min_c \frac{J^c(y)}{A^c} \right) + 1 - \tilde{t}(x). \quad (5)$$

由前述暗原色先验理论可得

$$J^{dark}(x) = \min_{y \in \Omega(x)} (\min_c J^c(y)) = 0. \quad (6)$$

由此可得

$$\min_{y \in \Omega(x)} \left(\min_c \frac{J^c(y)}{A^c} \right) = 0. \quad (7)$$

将(7)代入(5)中可得

$$\tilde{t}(x) = 1 - \min_{y \in \Omega(x)} \left(\min_c \frac{J^c(y)}{A^c} \right). \quad (8)$$

空气中的颗粒和雾感能给人带来一定的景深效果,这里笔者在(8)中引入一个 $[0, 1]$ 之间的数来保留一定的雾,本文仿真取值为 0.95.

$$\tilde{t}(x) = 1 - \omega \min_{y \in \Omega(x)} \left(\min_c \frac{J^c(y)}{A^c} \right). \quad (9)$$

而全球大气光 A 的求取,笔者借助于暗通道图来从有雾图像中获取该值.目前 A 值的求取主要分为以下两种方法:1、从暗通道图中按照亮度的大小取前 0.1% 的像素^[8];2、在这些位置中,在原始有雾图像 I 中寻找对应的具有最高亮度的点的值,作为 A 值.当投射图的 t 值很小时,会导致 J 值偏大,从而使图像整体向白场过渡,因此一般可设置一阈值 t_0 , 当 t 值小于 t_0 时,令 $t = t_0$,本文中所有效果图均以 $t_0 = 0.1$ 为标准计算.因此,最终去雾公式如下:

$$J(x) = \frac{I(x) - A}{\max(t(x), t_0)} + A. \quad (10)$$

1.3 导向滤波算法

导向滤波算法是通过一个引导图像 I 来对 p 进行滤波而得到 q ,其数学公式为

$$q_i = \sum_{j \in W_i} W_{ij}(I) \cdot p_j. \quad (11)$$



$W_{ij}(I)$ 表示由引导图像 I 来确定加权平均运算中所采用的权值。

噪声的特点通常是以其为中心的各个方向上梯度都较大而且相差不多。边缘则不同,边缘相比于区域也会出现梯度的跃变,但是边缘只有在其法向方向上才会出现较大的梯度,而在切向方向上梯度较小。相比普通平滑和高斯平滑,导向滤波的加入能保护更多的边缘信息,使后期算法能学习到更



图1 原图



图2 未加入 Fast Guided Filter 的暗通道去雾效果图



图3 加入 Fast Guided Filter 后的暗通道去雾效果图

好的模型,同时也保留了更多的纹理特征。

Fast Guided Filter^[9]是何凯明博士于2015年提出的一种很实用的更快速的导向滤波流程。其本质是通过下采样减少像素点,计算 $mean_a$ 和 $mean_b$ 后进行上采样恢复到原有的尺寸大小。导向滤波的时间复杂度为 $O(N)$,其中 N 为像素点的个数。假设缩放比例为 s ,那么缩小后像素点的个数为 N/s^2 ,那么时间复杂度变为 $O(N/s^2)$ 。去雾效果如图1~3所示。

2 目标检测

2.1 YOLO 算法

YOLO v3^[10] 作为目前最优秀的目标检测框架之一,其实时性相比之前的 RCNN^[11]、SPP-Net^[12]、Fast-RCNN^[13]、Faster-RCNN^[14] 都有很大的进步。YOLO 的核心思想是利用整张图片作为输出,直接在输出层回归 bounding box(边界框)的位置及其所属的类别。首先将一幅图片分割成 7×7 个网格(grid cell),再对目标进行检测,如果目标中心落在某个网格中,网格就负责对该目标进行预测。预测的候选框还需要预测一个 confidence 值来判别目标是否存在^[5]。

$$confidence = \Pr(\text{Object}) \times IOU_{\text{Pred}}^{\text{Truth}}. \quad (12)$$

如果有目标落在一个网格里,那么 $\Pr(\text{Object})$ 值取1,否则取0。 $IOU_{\text{Pred}}^{\text{Truth}}$ 是预测的 bounding box 和实际的 groundtruth 之间的交并比。

$$IOU_{\text{Pred}}^{\text{Truth}} = \frac{\text{area}(\text{box}(\text{Truth}) \cap \text{box}(\text{Pred}))}{\text{area}(\text{box}(\text{Truth}) \cup \text{box}(\text{Pred}))}. \quad (13)$$

在预测时,每个候选框的分类置信分数为预测目标的类别概率和 bounding box 预测的 confidence 信息相乘。

$$\text{Conf} = \Pr(\text{Class}_i | \text{Object}) \times \Pr(\text{Object}) \times IOU_{\text{Pred}}^{\text{Truth}}. \quad (14)$$

得到每个分类置信分数之后,设置阈值将得分低的 boxes 筛除掉,就得到最终的检测结果。

但是由于 YOLO 的全连接层输入输出的特征向量必须固定长度,所以 YOLO 只能接收固定尺寸的图片。当物体占画面比例较小的时候,有可能会出一个格子包含多个目标物体,但是 YOLO 这时只能检测出其中一个。

2.2 YOLO 算法的改进

目前大多数神经网络的改进趋势都是往更大更深的网络发展。通常情况下层级更多的神经网络往往能在实际情况中学习到更多的细节从而达到更好的效果。YOLO v2^[15] 并没有致力于加深神经网络的层数,而是通过 Batch Normalization、High resolution classifier、Multi-Scale Training 等方法使其在保持原有速度的同时提高了精度。整个网络应用了 Darknet-19 的网络架构,包含 19 个卷积层和 5 个 max pooling 层,卷积的计算量相较于 YOLO 大大减少。YOLO v3^[10] 采用了 Darknet-53 的基础网络结构,在保证速度的同时又再一次大大提高了精确度。

Darknet-53 网络结构借鉴了残差网络 residual network 的做法,在一些层之间设置了快捷链路(shortcut connections)。图4左侧的1、2、8等数字表示残差组件的个数,每个残差组件有两个卷积层和一个快捷链路。同时还采用多尺度预测方案,针对 COCO 数据集,YOLO v3 设计了9个不同大小的预测框,将其按照大小均分给3个尺度。

但是 YOLO 算法的预测框均是建立在 COCO

数据集的基础之上的. 种类较多, 尺寸比例差异较大, 笔者仅在 INRIA 数据集上对行人进行检测和识别. 因此笔者在设计预测框时对 INRIA 数据集进行了 K-means 方法聚类. 最后对 INRIA 数据集的聚类结果, 取消了 YOLO v3 原来在三个维度上的输出检测, 而直接将 anchor boxes 的个数和宽高维度设定为聚类所得到的结果.

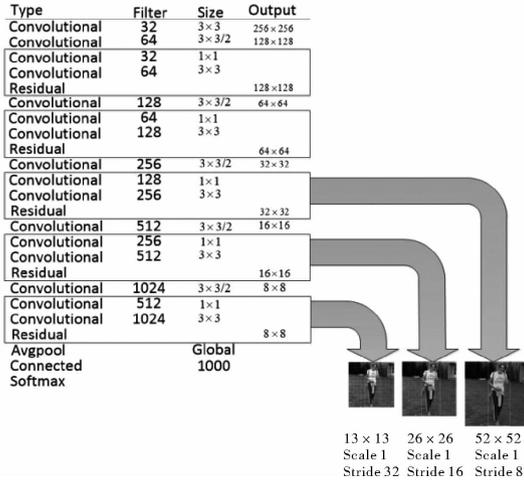


图4 Darknet-53 网络结构

从图5中可以看出簇 K 值在等于 4 之后逐渐趋于平稳. 所以笔者将变化的拐点定为最佳 anchor boxes 的数量. 实验证明聚类后的预测框大大缩短了训练的收敛时间, 同时还可以消除候选框带来的误差.

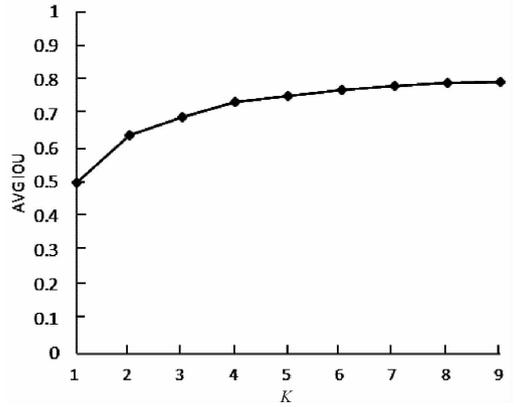


图5 K-means 聚类分析结果

2.3 目标检测结果分析

笔者采用深度学习架构 YOLO - Darknet53 搭建仿真的运行环境, 并配置 NVIDIA 的 CUDA 环境进行 GPU 并行加速计算. 图6是训练过程平均损失变化的对比.

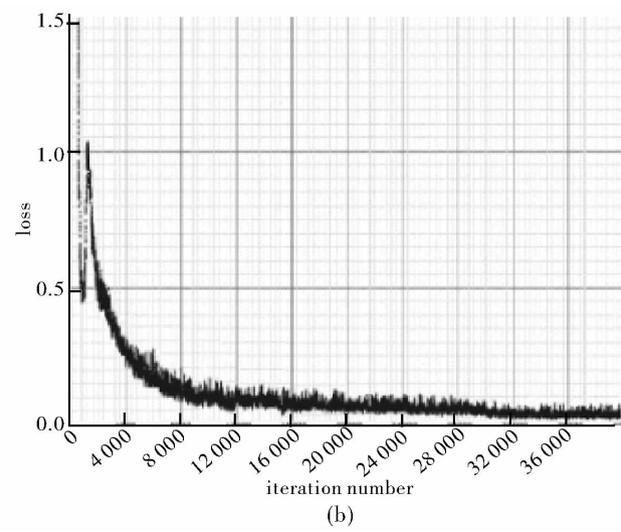
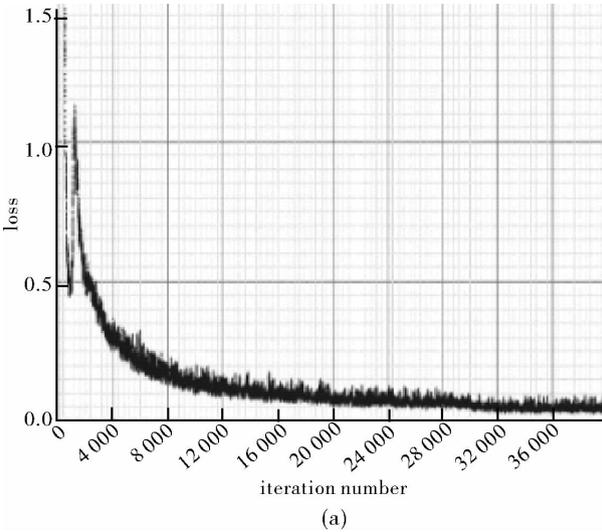


图6 平均损失值变化比较

仿真实验中每轮迭代从处理后的训练集里随机抽 batch = 64 个样本进行训练, 由于本次实验采用的显卡为 2G 显存, 为了减轻内存占用压力, 每轮的 64 个样本又被均分为 subdivision = 16 次送入网络参与训练. 同时采用分步学习策略, 为了提高学习效率, 又避免过拟合现象, 笔者将最初的学习率定为 10×0.001 , 在 15 000 次迭代后, 学习率更改为 0.1×0.001 , 30 000 次后再更改为上次更改后的 0.1 倍, 最终, 经过 40 000 次迭代, YOLO v3 平均损失值最后保持在 0.042 2. 而经改进后的 YOLO v3 平均损失值可以达到 0.36. 并且从图中可以明显地看出经

K-means方法聚类出预测框的 YOLO v3 算法收敛速度更快. 算法对比效果见图7~9.



图7 HOG + SVM 目标识别效果





图8 YOLO v3 目标识别效果



图9 本文算法目标识别效果

3 图像特征提取

3.1 颜色特征的提取

颜色通常是用3个独立的属性来描述,常见的颜色空间有RGB、HSV和YUV.3种颜色空间各有自己的特点和优势,RGB颜色空间是通过3种不同颜色的分量叠加形成任何颜色的光,RGB表示的物理意义很清晰,但是不能很好地适应人的视觉特点.而HSV空间则能够描述人看到物体时所使用的色调、饱和度和明度.YUV主要用于优化彩色视频信号的传输.笔者使用RGB与HSV分别描述前景个体颜色特征.

在RGB颜色空间提取R分量平均值 μ_R 、G分量平均值 μ_G 、B分量平均值 μ_B 、RGB平均值 μ_{RGB} 、RGB标准差 σ_{RGB} 5个颜色特征值.

分别提取目标对象的R、G、B3个分量各自的平均值 μ_λ :

$$\mu_\lambda = \frac{1}{N_{obj(i,j) \in obj}} \sum I_\lambda(i,j). \quad (15)$$

式中分别取R、G、B, N_{obj} 为提取目标对象后图像的像素数, $I_\lambda(i,j)$ 为像素点 (i,j) 的 λ 分量,得 μ_R 、 μ_G 、 μ_B 三值.

RGB平均值:

$$\mu_{RGB} = \frac{1}{N_{obj(i,j) \in obj}} \sum I_{RGB}(i,j). \quad (16)$$

式中 $I_{RGB}(i,j)$ 为像素点 (i,j) 的RGB平均值,得到 μ_{RGB} .

RGB标准差:

$$\begin{aligned} A &= \sum_{(i,j) \in obj} (I_R(i,j) - I_{RGB}(i,j))^2; \\ B &= \sum_{(i,j) \in obj} (I_G(i,j) - I_{RGB}(i,j))^2; \\ C &= \sum_{(i,j) \in obj} (I_B(i,j) - I_{RGB}(i,j))^2; \\ \sigma_{RGB} &= \sqrt{\frac{1}{N_{ROI}}(A + B + C)}. \end{aligned} \quad (17)$$

得到 σ_{RGB} .

在HSV空间用同样方法提取H分量平均值 μ_H 、S分量平均值 μ_S 、V分量平均值 μ_V 、HSV平均值 μ_{HSV} 、HSV标准差 σ_{HSV} 5个颜色特征值.

由RGB颜色特征组成的五维向量标记为:

$$V_{rgb} = (\mu_R, \mu_G, \mu_B, \mu_{RGB}, \sigma_{RGB}).$$

由HSV颜色特征组成的五维向量标记为:

$$V_{hsv} = (\mu_H, \mu_S, \mu_V, \mu_{HSV}, \sigma_{HSV}).$$

3.2 纹理特征的提取

笔者采用统计理论与小波变换2种方法对纹理特征进行描述.一副图像的灰度图像由无数灰度值的像素点组成,灰度直方图能够直观地表现图像中灰度分布情况,包含了大量的图像特征.笔者基于灰度直方图提取了均值 m 、标准方差 σ 、平滑度 R 、三阶矩 μ_3 、一致性 U 、熵 e .

$$m = \sum_{i=0}^{L-1} z_i p(z_i), \quad (18)$$

$$\sigma = \sqrt{\sum_{i=0}^{L-1} (z_i - m)^2 p(z_i)}, \quad (19)$$

$$R = \frac{1}{(1 + \sigma^2)}, \quad (20)$$

$$\mu_3 = \sum_{i=0}^{L-1} (z_i - m)^3 p(z_i), \quad (21)$$

$$U = \sum_{i=0}^{L-1} p^2(z_i). \quad (22)$$

式中 L 是灰度级数, z_i 表示第 i 个灰度级, $p(z_i)$ 是归一化直方图灰度级分布中灰度为 z_i 的概率.将6个特征组合成六维特征向量 V_{gh} :

$$V_{gh}(m, \sigma, R, \mu_3, U, e). \quad (23)$$

灰度共生矩阵是相距一定距离的两个特定灰度值的像素同时出现的矩阵表达形式.它描述了空间位置上特定两个像素点之间的联合分布,包含了大量的图片纹理特征信息.但是由于数据量较大,一般会基于其构建一些统计量作为分类特征.笔者基于灰度共生矩阵的特征,提取了能量ASM、对比度CON、相关COR和熵ENT,并求取四个特征的均值和标准差组成八维向量 V_{cm} .

$$ASM = \sum_{i=0}^{l-1} \sum_{j=0}^{l-1} P^2(i, j, d, \theta), \quad (24)$$



$$CON = \sum_{i=0}^{l-1} \sum_{j=0}^{l-1} [(i-j)^2 P^2(i,j,d,\theta)], \quad (25)$$

$$COR = \sum_{i=0}^{l-1} \sum_{j=0}^{l-1} i \times j \times P(i,j,d,\theta), \quad (26)$$

$$ENT = \sum_{i=0}^{l-1} \sum_{j=0}^{l-1} P(i,j,d,\theta) \log_2 P(i,j,d,\theta). \quad (27)$$

式中:

$$P(i,j,d,\theta) = \{P[(x,y), (x+D_x, y+D_y)] | f(x,y) = i; f(x+D_x, y+D_y) = j\}, \quad (28)$$

D_x, D_y 是位置偏移量, d 为共生矩阵的生成步长, θ 为共生矩阵的生成方向. 由以上 4 个特征: 能量、对比度、相关和熵的均值、标准差组成八维向量, 记为 V_{cm} .

$$V_{cm} = (ASM_m, ASM_\sigma, CON_m, CON_\sigma, COR_m, COR_\sigma, ENT_m, ENT_\sigma). \quad (29)$$

本文在小波变换的描述上基于小波系数能量 WE 对目标灰度通道进行两层小波分解, 得到 6 个子细节图的平均能量向量 WE_{grey} , 其中:

$$WE = \frac{1}{m \times n} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} w^2(i,j), \quad (30)$$

$$WE_{grey} = (WE_{hl1}^{grey}, WE_{hl2}^{grey}, WE_{hh1}^{grey}, WE_{hl1}^{grey}, WE_{hl1}^{grey}, WE_{hh1}^{grey}). \quad (31)$$

式中: $m \times n$ 为图像像素大小; $w(i,j)$ 为对应点的小波系数; $WE_{hl1}^{grey}, WE_{hl2}^{grey}, WE_{hh1}^{grey}$ 分别为灰度通道小波分解的第 1 层垂直分量、水平分量和对角系数能量; $WE_{hl2}^{grey}, WE_{hl2}^{grey}, WE_{hh2}^{grey}$ 分别为灰度通道小波分解的第 2 层垂直分量、水平分量和对角系数能量.

3.3 特征向量组合

为了对比不同特征向量组合对具有相似外部特征人员检测的效果, 笔者对提取到的特征进行不同的编组, 得到不同的特征组如下:

$$F_1 = (V_{rgb}, V_{hsv}), \quad (32)$$

$$F_2 = (V_{rgb}, V_{hsv}, V_{gh}), \quad (33)$$

$$F_3 = (V_{rgb}, V_{hsv}, V_{cm}), \quad (34)$$

$$F_4 = (V_{rgb}, V_{hsv}, WE_{grey}), \quad (35)$$

$$F_5 = (V_{rgb}, V_{hsv}, V_{gh}, V_{cm}, WE_{grey}). \quad (36)$$

4 基于极限学习机的特征分类

极限学习机 (Extreme Learning Machine)^[16] 是由黄广斌教授提出的求解单隐层前馈神经网络的算法. 其核心内容是输入层与隐藏层之间的权值参数求解, 以及隐藏层的偏置参数的求解, 其求解过程也较简单, 可划归为求解一个矩阵的广义逆的问题. 在保证相同学习精度的情况下, 与传统的神

经网络相比也能拥有较快的速度. 针对隐含层与连接层初始值的选取, 本文通过 *rand* 生成. 笔者针对两组测试数据, 选取的隐含层神经元个数为 45 和 50 个.

4.1 极限学习机分类

1) 训练集: 给定 Q 个不同样本 (x_i, t_i) . 其中 $x_i = [x_{i1}, x_{i2}, \dots, x_{in}]^T$; $t_i = [t_{i1}]$. x_i 为上述几种特征组合; t_i 为一个标签, 表示对应目标属于哪一类人群.

2) ELM 训练: W 为输入层与隐含层的连接权值矩阵, β 为隐含层与输出层连接权值矩阵, b 为隐含层神经元阈值, 激活函数为 $g(x)$. H 为神经网络隐含层的输出矩阵, T^T 为网络的输出. 于是可得 $H\beta = T^T$, 由 ELM 的两个定理, 根据最小二乘法求解 $\min \|H\beta\|$, 其解为 $\hat{\beta} = H^+ T^T$, 其中 H^+ 为隐含层输出矩阵 H 的广义逆.

3) 对经过训练的 ELM 输入测试视频, 即可对当前视频中人群类别进行分类, 将具有共同颜色和纹理特征的人分离出来.

4.2 算法步骤

1) 针对两组不同的测试集, 笔者选取的隐含层神经元个数分别为 45 和 50 个, 对 ELM 在 $(0, 1)$ 区间采用均匀分布随机设定输入层与隐含层的链接权值 ω 与隐含层神经元的阈值 b .

2) 选择 S 函数作为隐含层神经元的激活函数, 进而计算隐含层输出矩阵 H ;

3) 计算输出层权值 $\hat{\beta} = H^+ T^T$.

使用 ELM 算法对提取的颜色特征、纹理特征参数进行分析, 对人群进行相似度分类, 把人群分为一般人群和具有相似外部特征人群.

相关实验表明, 在 ELM 中许多非线性函数都可以使用, 甚至还可线性激活函数分类非线性样本. 针对本文数据集, 根据相关研究^[17], 选取 Sigmoid 函数作为激活函数的分类效果优于其他函数, 隐含层神经元个数根据样本本数的变化而做出改变能获得较好的分类精度.

5 实验分析

本次仿真基于普通 PC 机平台 (CPU 为 6500K, 3.2 GHz, 6.00 GB 内存, 显存为 2 G) 完成. YOLO 采用深度学习架构 YOLO - Darknet53 搭建仿真的运行环境, 并配置 NVIDIA 的 CUDA 环境进行 GPU 并行加速计算. 特征提取和分类部分通过 MATLAB 2014a 编程仿真环境编写代码并进行仿真^[18]. 笔者在拍摄的视频数据集里提取了 248 个目标个体与纹理特征数据作为训练集并做了大量

仿真实验, 最后选取其中两帧的仿真结果对比图, 并对仿真整体做了平均准确率的计算.

本文特征分类部分训练集为 248 个目标个体, 具有相似外部特征的样本有 4 组, 其中每个类别含有 6 个目标个体. 测试集数据为 77 个目标个体, 具有相似外部特征的个体有 2 类, 每类中有 4 个目标个体, 其余均为普通个体. 表 1 为不同的特征组合的人群类别分类准确率; 图 10 ~ 14 为测试集预测分类的结果与真实值的对比. 在表 1 中可以看到 F_1 特征组合的识别率为 84.415%, 相比于其他组合识别是最低的, 说明在复杂条件下仅仅依靠颜色特征无法可靠识别出目标个体, 而在一定程度上增加特征数量可以提高识别的准确率. 最后一组 F_5 特征组合在实验条件下达到了 96.104%, 该组融合了文中提到的全部特征向量, 各种特征之间相互弥补了各自的不足, 呈现出较好的仿真效果. ELM 作为一种机器学习分类算法, 学习效率高、泛化能力强, 相比于传统单隐层前馈神经网络, 在保证精度的前提下速度更快.

表 1 5 种特征组合在 ELM 下的准确率

特征组合类别	总检测人数	识别率 / %
F_1	77	84.415
F_2	77	90.909
F_3	77	93.506
F_4	77	94.805
F_5	77	96.104

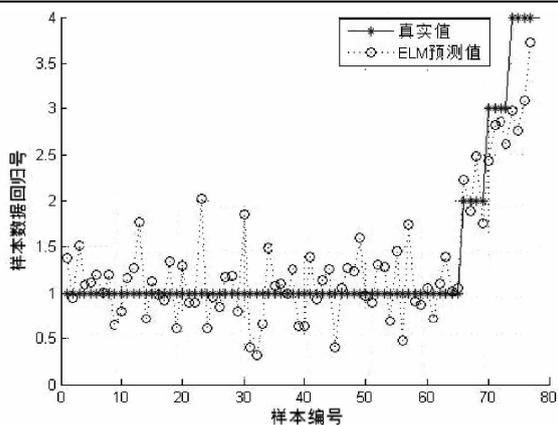


图 10 F_1 特征组合

6 结论

笔者提出的基于深度学习的相似外部特征人员检测算法利用了加入 Fast Guided Filter 的改进暗通道去雾算法对拍摄的视频数据集进行优化, 再将经过优化增强后的样本图片投入到经 K-means 方法聚类出预测框的 YOLO v3 网络框架下进行训练. 然后用训练出的深度模型提取出需要检测相似

性特征的目标人物. 最后将目标任务提取多种颜色及纹理特征并结合 ELM 对其进行分类. 深度学习的加入大大降低了由于遮挡和弯曲所造成的目标漏检问题, 多特征的应用也让分类准确率有了极大的保障. 目前由于采集到的目标人物有限, 实际的目标类别和特征都会远大于本文数量, 故本研究成果对于实际的精准分类来讲还需要继续完善. 另外 YOLO 网络对于图像中过小的目标依然存在漏检的现象, 下一步的工作将考虑根据特定的环境改进现有神经网络的结构, 使之更能适应当前环境的应用.

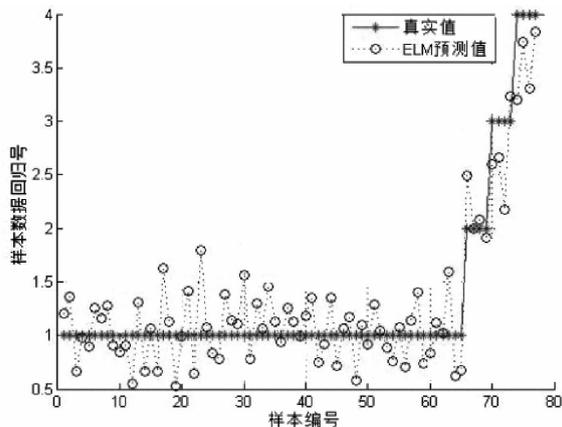


图 11 F_2 特征组合

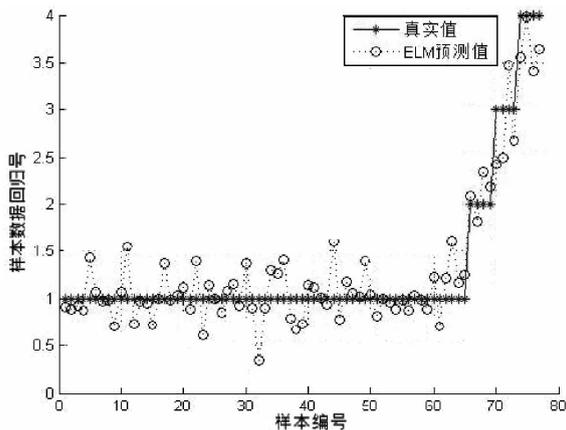


图 12 F_3 特征组合

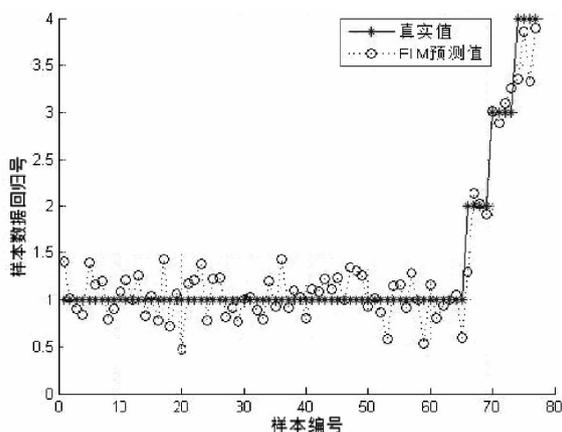
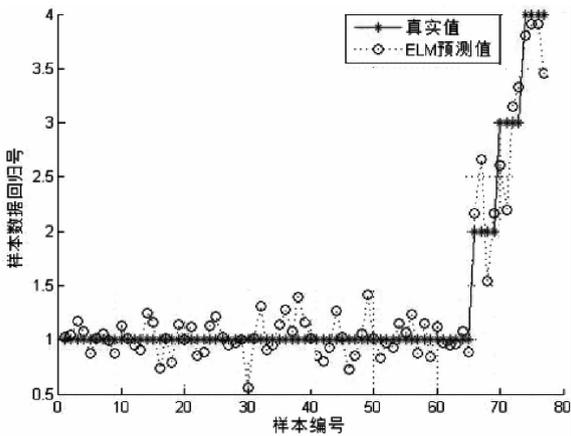


图 13 F_4 特征组合



图 14 F_5 特征组合

参考文献:

- [1] 郑胤, 陈权崎, 章毓晋. 深度学习及其在目标和行为识别中的新进展[J]. 中国图象图形学报, 2014, 19(2): 175 - 184.
- [2] LECUN Y, BENGIO Y, HINTON G. Deep learning[J]. Nature, 2015, 521: 436 - 444.
- [3] 田青, 袁瞳阳, 杨丹, 等. 基于暗通道去雾和深度学习的行人检测方法[J]. 激光与光电子学进展, 2018, 55(11): 111007.
- [4] WANG X Y, HAN T X, YAN S C. An HOG - LBP human detector with partial occlusion handling[C]. Washington: Proceedings of IEEE International Conference on Computer Vision, 2009: 32 - 39.
- [5] 高宗, 李少波, 陈济楠, 等. 基于 YOLO 网络的行人检测方法[J]. 计算机工程, 2018, 44(5): 215 - 219, 226.
- [6] 李文勇, 李明, 陈梅香, 等. 基于机器视觉的作物多姿态害虫特征提取与分类方法[J]. 农业工程学报, 2014, 51(9): 154 - 162.
- [7] 刘丽, 匡纲要. 图像纹理特征提取方法综述[J]. 中国图象图形学报, 2009, 14(4): 622 - 635.
- [8] HE K, SUN J, TANG X. Single Image Haze Removal Using Dark Channel Prior[J]. IEEE Trans Pattern Anal Mach Intell, 2011, 33(12): 2341 - 2353.
- [9] HE K M, SUN J. Fast Guided Filter[M/OL]. <https://arxiv.org/pdf/1505.00996.pdf>.
- [10] REDMON J, FARHADI A. Yolov3: An incremental improvement[M/OL]. <https://arxiv.org/pdf/1804.02767.pdf>.
- [11] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [EB/OL]. <https://arxiv.org/abs/1311.2524>.
- [12] 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 & Machine Intelligence, 2014, 37(9): 1904 - 1916.
- [13] GIRSHICK R. Fast R - CNN[C]. Washington: Proceedings of IEEE International Conference on Computer Vision, 2015: 1440 - 1448.
- [14] REN S Q, HE K M, GIRSHICK R, et al. Faster R - CNN: towards real - time object detection with region proposal networks[M/OL]. <https://arxiv.org/pdf/1506.01497.pdf>.
- [15] REDMON J, FARHADI A. YOLO 9000: better, faster, stronger[C]. Washington: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2017: 6517 - 6525.
- [16] HUANG G B, ZHOU Q Y, SIEW C K, et al. Extreme learning machine: a new learning scheme of feedforward networks with random hidden nodes[J]. IEEE Transactions on Neural Networks, 2006, 17(4): 879 - 892.
- [17] 张文博, 姬红兵, 王磊, 等. 多隐层输出矩阵极限学习机[J]. 系统工程与电子技术, 2014, 36(8): 1656 - 1659.
- [18] 王小川, 史峰. MATLAB 神经网络 43 个案例分析[M]. 北京: 北京航空航天大学出版社, 2013: 8.

(责任编辑:王彦江)

Similar External Features Personnel Detection Algorithm Based on Improved YOLO v3

LIANG Siyuan, WANG Ping, LUO Fanbo, XU Guifei, WANG Wei

(School of Electrical & Electronic Information, Xihua University, Chengdu, Sichuan 610039, China)

Abstract: Similar external characteristics often mean problems such as team action and potential abnormal behavior. To solve the problems of fewer detection methods for similar features, low accuracy and fewer features in abnormal behavior detection, an algorithm for detecting similar external features based on in-depth learning is proposed. And similar external features often mean team action and potential abnormal behavior. Firstly, the dark channel defogging algorithm with Fast Guided Filter is used to pretreat the image of the INRIA database, and the training samples with better quality are obtained. Then, an improved detection algorithm of YOLO v3 is trained with the obtained samples. Finally, pedestrians will be extracted from the extraction of color features and texture features of several combinations, after using ELM classification. The simulation results show that the dark channel defogging algorithm with Fast Guided Filter is superior to the simple dark channel defogging algorithm, retaining more edge features and texture features, especially in foggy weather and strong exposure environment. Compared with the HOG + SVM method, the false and missed detection rate of pedestrian detection in this algorithm are greatly reduced and have better real-time performance. Finally, the accuracy of ELM classification can reach 96.104%.

Key words: dark channel defogging YOLO network, color characteristics, texture feature, ELM

