文章编号:1671-4598(2018)11-0277-04 DOI:10.16526/j.cnki.11—4762/tp.2018.11.060 中图分类号:TP3 文献标识码:A

油田安防领域基于改进的深度残差网络行人检测模型

杨其睿
(中国石油工程建设有限公司 西南分公司,成都 610041)

摘要:油田安防中行人目标检测是是当今前沿的一个热门研究课题;针对野外场景采集的图像视频分辨率低,背景复杂等问题,提出了一种基于单次多目标检测器(Single Shot MultiBox Detector,SSD)模型的改进行人检测算法,该算法首先利用聚合通道特征模型对图像或者视频序列进行进行预处理,获得疑似目标区域,大大降低单帧图像检测的时间;然后对SSD的基本网络VGG—16替换为Resnet—50,通过增加恒等映射解决网络层数加深但检测精度下降的问题;最后采用强大而灵活的双参数损失函数来优化训练深度网络,提高网路模型的泛化能力;定性定量实验结果表明本文所提检测算法的性能超过现有的检测算法,在保证行人检测准确率的同时提高检测效率。

关键词:行人检测;深度学习;损失函数;恒等映射;聚合通道特征

Pedestrian Detection Model based on Improved Deep Residual SSD Network in Oilfield Security Field

Yang Qirui
(China Petroleum Engineerin&Construction Corp. Southwest Company, Chengdu 610041, China)

Abstract: Pedestrian detection in oilfield security is a hot research topic. For the low resolution and complex background in image, an improved pedestrian detection algorithm based on the Single Shot MultiBox Detector (SSD) is proposed. The algorithm firstly uses the aggregate channel feature model to preprocess the image or video sequence to obtain the suspected pedestrian area, which greatly reduces the time of single—frame image detection. The basic network VGG—16 is replaced by Resnet—50, which introduces the identity mapping to solve the problem of reducing the detection accuracy when the number of network layers are increased. Finally, the powerful and flexible two—parameter loss function is used to optimize the training deep network and improve the network model generalization ability. Qualitative and quantitative experiments show that the performance of the proposed detection algorithm exceeds the existing detection algorithm, and the detection efficiency is improved while ensuring the accuracy of pedestrian detection.

Keywords: pedestrian detection; deep learning; loss function; Identity mapping; aggregate channel feature

0 引言

油田领域的安防系统,就是对探测器获取的图像序列进行图像信号处理分析,从视频背景中检测到目标信息,进而利用跟踪算法在随后的序列图像中定位出目标的位置坐标信息,为后续深层次的行人行为理解与决策奠定基础[1]。因此,智能视频监控中的行人目标检测是当今前沿的一个热门研究课题,有着重要的实际应用价值和研究意义,其算法的好坏直接影响着行人目标识别与行为理解的稳定性和准确性[2]。本文主要结合油气田地面建设工程通信网络与安防系统设计的需求,对现有的行人检测模型进行分析,为油气田安全、高效生产运行提供技防保障。

传统的行人检测方法主要是采用低级特征对行人进行表征,如HOG特征,Haar特征,LBP特征,LUV特征,ICF特征,ChnFtrs特征,以及LDCF特征,然后利用支持向量机,决策树等分类器进行分类,实现行人目标检测[3]。由于监控环境复杂,目标外观差异巨大,直接采用单一的人工特征表征能力不足,泛化效果较差,不能解决复杂干扰条件下行人检测精度不高的问题。同时,现在行人检测算法大都采用滑动窗口进行穷举搜索,通过采取一定的策略生成候选窗口,然后利用分类算法对对候选区域分类。然而,由于行人目标外形差异性太大,通过多尺度筛选候选窗口的复杂度太大,分类处理十分低效的。

随着近两年以卷神经网络为代表的深度学习算法在目标识别、医学诊断等领域所取得的突破,国内外研究人员也将深度学习模型扩展并应用到安防领域行人检测应用领域中,并取得了有效的研究成果[4]。在深度学习方法中,特征提取网络层即为整个构架的核心技术,其功能等同梯度方向直方图或DPM等提取特征的方法,唯独不同的地方是CNN所提取的特征并非人工设计的,而是藉由训练网络而学习到的,也因此CNN能提取到更为关键的特征进而达到更好的效果。分类网络层则等同于SVM等分类器,透过学习利用前一个网络所提供的信息来分类图像是否为行人目标。一般来说,基于深度学习的行人检测方法可以分为两类:基于区域选择[5],即提取候选框,如R—CNN,SPP

收稿日期:2018-08-09; **修回日期**:2018-09-07。

作者简介:杨其睿(1985-),男,四川江油人,工程师,主要从事油气田地面建设工程通信网络与安防系统设计等方向的研究。

－Net，Faster－RCNN，R－FCN 等；基于端到端的回归网络[6]，如 YOLO，SSD 等。前者将目标检测转化为分类为题和回归问题，首先采用独立的区域提取网络求取疑似目标区域，然后利用边界框回归对提取的区域进行位置修正，最后采用 softmax 进行分类；后者是将目标检测作为一个回归问题进行求解，通过将图像经深度网络处理，得到图像中所有物体的位置和其所属类别及相应的置信概率[7]。可以看出，后者将检测过程整合为一个回归问题，使得网络结构简单，检测速度大大加快；同时，由于网络没有分支，所以训练也只需要一次即可完成。因此，通过将目标检测模型转化为回归问题的思路非常有效。虽然基于深度学习的行人检测正在迅速发展，但无论在准确性还是速度方面，仍有很大的改进空间。该方法的目标是在不影响检测精度的情况下降低计算成本。

笔者结合多年在安防监控领域图像处理算法的经验，以“XXX 油田地面建设工程通信网络与安防系统设计项目”为契机，通过对传统模型与深度网络的分析，提出了一种基于改进 SSD 模型的深度网络行人检测算法。该算法首先利用聚合通道特征模型对图像或者视频序列进行进行预处理，获得疑似目标区域，大大降低单帧图像检测的时间；然后对 SSD 的基本网络 VGG－16 替换为 Resnet－50，通过增加恒等映射解决网络层数加深但检测精度下降的问题；最后采用强大而灵活的双参数损失函数来优化训练深度网络，提高网路模型的泛化能力。

1 相关技术

1.1 SSD 结构

SSD 模型的核心是使用卷积滤波在特征图中很多给定的边界框中得到目标所属类别及对应位置偏差[8]。相比于 YOLO 模型，SSD 模型采用卷积层替代 YOLO 的全连接层，实现多尺度目标位置预测，其流程框架如图 1 所示。SSD 模型主要基于 VGG16 网络进行改进，将全连接层替换成卷积层，同时再网络中还增加了最大池化层，并为卷基层输出一个对应的特征图，并以此作为预测的输入[9]。换句话说，这些不同尺度的特征图能够作为预测的输入，以此来获得不同特征尺度。为了实现多层次映射，采用了 3×3 的卷积核进行特征描述，而不是 VGG16 网络采用的 5×5、7×7、11×11 等尺度的卷积核。因此，该模型能够同时提取图像序列的低层次特征与高层次深度特征。

在不同尺度下获取每个位置在特征图上对应网的一系列固定大小的框，其中每个网格具有 k 个矩阵框，每个矩形框能够预测 c 个目标类别的评分及其对应的位置偏差[9]。若获取的特征图的大小是 $m\times n$，则可以获得 $m\times n$ 个特征图网格，那么该特征图共有 $(c+4)\times k\times m\times n$ 个输出。在训练阶段，一旦矩形框与基准框匹配成功，那么该矩形框就是为正样本，反之则为负样本。SSD 模型只有正样本参与代价函数的计算，先学习重合度高的框的位置信息，再学习所属类别信息。因此，训练阶段获取的正样本的数目较少，使得最终网络训练开销不大。

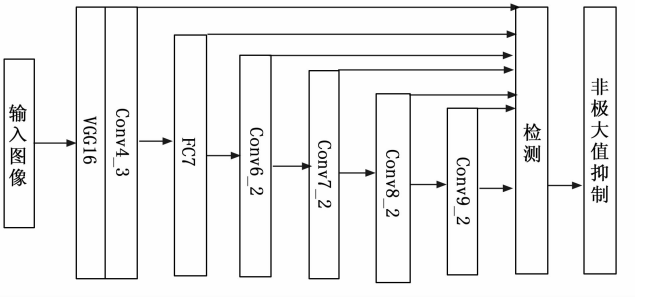

图 1 SSD 模型检测框架

SSD 模型的目标代价损失函数是由位置损失项与置信度损失项加权的结果，其等式如下所示：

$$L(x,c,l,g)=(L_{conf}(x,c)+\alpha L_{loc}(x,l,g))/N \quad (1)$$

其中：c 表示多类别目标属性的置信度；l 和 g 分别表示获取的预测框与基准框；N 则是匹配到的矩形框的个数；α 是交叉验证的正则化权值系数。位置损失项 $L_{loc}(x,l,g)$ 是 l 和 g 之间的 $smooth_{L1}$ 正则项获取，通过对不同位置中心的多尺度矩形框进行偏移回归，获取最优的匹配点位置，其计算表达式如下：

$$L_{loc}(x,l,g)=\sum_i\sum_m x_{ij}^k smooth_{L1}(L_i^m-\hat{g}_j^m) \quad (2)$$

其中：x_{ij}^k 表示类别属性 k 中第 i 个矩形框与第 j 个基准框的匹配结果，若 $x_{ij}^k=1$ 表示结果保持一致；反之则不一致。置信度损失项 $L_{conf}(x,c)$ 是通过计算多类别属性的 Softmax 损失函数得到，其表达式如下：

$$L_{conf}(x,c)=-\sum_{i\in pos}x_{ij}^p\log(\hat{c}_i^p)-\sum_{i\in neg}\log(\hat{c}_i^0) \quad (3)$$

SSD 模型添加逐层递减的卷积层，从而得到多尺度的特征集，有助于实现多尺度目标检测。多个不同层次的网络能够捕获图像不同级别的特征，实验结果已经表明采用低层次特征可以因为较低层捕获更精细的细节，增强目标的语义表征能力。

1.2 深度残差网络

深度残差网络（Resnet）的关键核心点是引入了残差块，在一个浅层网络基础上叠加恒等映射层进行残差学习，提升深度特征提取的精度，解决了梯度消失的问题[10]。假定 Resnet 网络的原始输入样本为 x，经过多层网络映射后能够得到 $F(x)$，因此，残差函数 $H(x)=F(x)-x$，如图 1 所示。可以看出，经过一个恒等映射，将输入叠加到卷积输出上，形成能够跳过一层或多层的跳跃式连接，消除梯度消失现象，可以使网络深度做到成百上千层。

恒等映射简单地叠加在网络中，即便增加网络的层数也不会降低网络的性能。图 1 的结构能够可以简单地使多个非线性层的权重趋向零以近似恒等映射，其输出可以表示为：

$$y=H(x,W_i)+x \quad (4)$$

其中：x 和 y 分别表示子块的输入与输出结果，$H(x,W_i)$ 是残差映射。上式中的引入的 x 通路，既没有引入额外参数也没有增加计算复杂度。仿真实验结果表明 Resnet 网络比相同规模的简单网络更容易收敛，能获得较好输出结

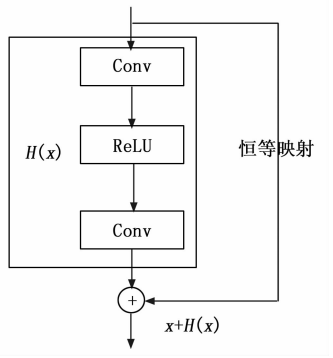

图 2　Resnet 网络中子块构造的示意图

果，不受网络深度影响。

2　本文改进的深度检测网络

现有的深度网络目标检测算法更注重提取图像的高层次特征，包括全局特征，而忽略低层次特征。低层次特征包含丰富的局部细节，可获得令人满意的人类视觉效果，这对行人检测任务具有实质性影响；高层次特征主要关注于大范围的感受野，确保超分辨重建的图像准确性。本文在 SSD 模型的基础上，提出了一种改进的行人检测算法，该算法首先利用聚合通道特征模型对图像或者视频序列进行进行预处理，获得疑似目标区域，大大降低单帧图像检测的时间；然后对 SSD 的基本网络 VGG－16 替换为 Resnet－50，通过增加恒等映射解决网络层数加深但检测精度下降的问题；最后采用强大而灵活的双参数损失函数来优化训练深度网络，提高网路模型的泛化能力。

2.1　基于聚合信道特征（ACF）模型的快速检测

现在行人检测算法大都采用滑动窗口进行穷举搜索实现分类[11]。然而，行人目标外形差异性太大，通过多尺度筛选候选窗口的复杂度太大，分类处理十分低效的。为了提高检测的效率，本文首先采用基于聚合信道特征（ACF）模型的目标检测算法对图像进行预处理，生成候选窗口。实验仿真表明，经过预处理后的候选区域窗口几乎囊括了图像中所有可能的目标区域，大大降低疑似目标数量，对图像中任何行人基本具有低漏检的检测效果。

假定处理后的图像具有 M 疑似区域个数，可以表示为 $\{B^i \in R^{m_i \times n_i} \mid i=1,2,\ldots,M\}$。由于残差网络与分类器的输入参数大小必须相同，所有的候选区域必须经归一化处理到统一的尺度，$\{D^i \in R^{m \times n} \mid i=1,2,\ldots,M\}$。

2.2　基于残差网络的 SSD 模型改进

众所周知，残差网络解决了网络层数加深，但检测精度下降的问题。与传统 VGG 相比，残差网络具有更少的滤波器和更低的计算代价，因此本文考虑将 SSD 的基准 VGG 网络替换成 Resnet－50 网络[12]。SSD 模型采用的多层次特征提取，导致获取的特征信息冗余太多，直接影响训练的复杂度，且存在维数灾难问题。因此通过前置网络对输入数据进行特征提取，为后续网络层提供输入信息，可加快后续训练速度，提高网络的泛化能力。

Resnet－50 结构采取了多个不同层次的残差结构，其中每个卷积块都包含不同数目的残差单元，每个残差单元进行三次卷积操作。由于残差结构中采用多个恒等映射进行叠加，即便增加网络的层数也不会降低网络的性能。等式 4 所示，恒等映射的是残差映射 $H(x,W_i)$ 表示为 $W_2\sigma(W_1x)$）是 σ 为 Relu 激活函数，若残差结构块的输入与输出维度不一致，则需要对输入 x 进行投影变换，消除维度间的差异，其等式如下：

$$y = H(x,W_i) + W_s x \tag{5}$$

在油气田野外监控环境下，其数据处理设备存在低功耗的要求，因此本文选用采取了裁剪的 Resnet－50 残差结构进行网络替换，其中第二层网络对获取的疑似区域进行特征提取，分别采用大小为 1×1×16，3×3×64，1×1×256 的卷积核进行描述；第三层卷积网络分别使用大小为 1×1×128，3×3×128，1×1×512 的卷积核；第四层卷积网络结构分别使用大小为 1×1×256，3×3×256，1×1×1024 的卷积核；第五层卷积网络结构分别使用大小为 1×1×512，3×3×512，1×1×2048 的卷积核）；剩下的网络结构与 VGG16 的结构保持一致。

采取多个不同尺寸的特征图对目标进行预测，可以对不同尺度的目标进行正确的检侧。本文提出的改进模型，不仅是端对端的检测网络，还可以进行参数的共享传递，增强特征提取效率。不同尺度获得的特征映射图中，不同特征位置输出的结果，对应到输入图像中的待检测区域。由于不同尺度下特征图与检测框进行配准存在差异，本文采用自适应矩形框提取策略。假定特征映射图中存在 m 个目标，其对应的尺寸可以表示为：

$$s_i = s_{\min} + \frac{s_{\max} - s_{\min}}{m-1}(i-1) \tag{6}$$

其中：$i \in [1,m]$，$s_{\min}$，$s_{\max}$ 分别表示深度结果中对应尺寸下最小与最大矩阵框尺度。文献［5］提出，生成的矩形框的长宽比 $r \in \{1,2,3,\frac{1}{2},\frac{1}{3}\}$ 是最优的搜索策略，其对应的长与宽分别为 $w_i^a = s_i\sqrt{r}$，$h_i^a = s_i\sqrt{r}$。通过以上策略对矩形框进行选择，可以增强最终行人目标覆盖精度，加快后续模型代价损失函数的计算效率。

2.3　损失函数

本文采用了一种广义的双参数损失函数，该函数可以推广到目前许多流行的鲁棒损失函数[13]。假定迭代前后之间的误差可以表示为 e，因此本文采用的损失函数可以表示为：

$$L(e,\alpha,\beta) = \begin{cases} \log(0.5(\frac{e}{\beta})^2+1) & \alpha = 0 \\ 1-\exp(1-0.5(\frac{e}{\beta})^2) & \alpha = -\infty \\ \frac{\rho(\alpha)}{\alpha}((\frac{1}{\rho(\alpha)}(\frac{e}{\beta})^2+1)^{\frac{\alpha}{2}}-1) & otherwise \end{cases} \tag{7}$$

其中：$\rho(\alpha) = \max(1,2-\alpha)$；$\alpha,\beta$ 是具有连续值属性的参数，可以通过不同的参数设置，模拟出任意的损失函数，如均方误差损失函数（l_2），绝对值误差损失函数（l_1）等。相比于传统固定参数损失函数，本文采用双参数损失函数

可以通过微调 α 和 β 获得更优的损失函数，具有很大的灵活性，可以适应更复杂的场景。

3 实验结果与分析

3.1 实验数据集及参数设置

为了评估本文提出的改进深度网络行人检测算法性能，训练集使用 NICAT，MIT 以及油田监控采集的数据集，训练集中正样本数量 23 589，负样本数量 23 991。测试集选择国际上广泛使用且非常具有挑战的公共测试数据集 INRIA 数据库以及油田监控序列，总共选择有 820 张图像。

本文采用 Resnet－50 残差结构模块进行构造深度网络。在训练阶段，最小批量设置为 16；采用梯度下降算法进行优化时，权重的更新规则中，学习速率初始化设置为设置为 0.25，然后在训练到第 30 个 Epoch（Epoch 是指对所有训练数据的一轮遍历）时，学习速率改为 0.025；若 100 个 Epoch 后，我们提出的损失函数没有改变则停止训练。为了提高优化效率，本文采用 ADAM 优化算法，Adam 优化算法是随机梯度下降算法的扩展式，可以基于训练数据迭代地更新神经网络权重，其参数设定为：alpha＝0.001，beta1＝0.9，beta2＝0.999 和 epsilon＝10－8。最小与最大矩阵框尺度分别设置为 $s_{\min}=0.15$，$s_{\max}=0.95$。损失函数中，α 和 β 分别经验设置为 1.12 和 0.05。

3.2 对比算法及评价指标

为了验证本文提出的行人检测算法的有效性，结合本文算法的特性，所选择的对比算法分别是 ConvNet[14]，YOLO－v2[15]，SSD[8]，R－CNN[16]。本文选择检测错误权衡图（DET）曲线评价行人目标的检测效果，其中 DET 表征检测率与每个图像误检率（FPPI）的关系。本文实验环境为：Intel Xeon CPU ES 1620 V33.5 GHz，16 GB 内存，Nvidia Geforce GTX 1080，Ubuntul6.04，64 位操作系统。

3.3 行人检测定性定量对比

为了定量分析本文所选择对比算法的行人检测性能，图 3 展示了对比算法的行人检测率与 FPPI 的关系曲线，其中 FPPI 是指平均每张图中能正确检测到的行人目标数目。从基本数据可以看出，本文所提检测算法相比于 ConvNet 算法、YOLO－v2 算法、SSD 算法和 R－CNN 算法有更高的检测率，尤其是针对 INRIA 数据集中 Street 中的图像，具有相当高的检测准确率与效率。由于 FPPI 是在不同检测率下的统计结果，为了便于对比深度网路的实际检测情况，本部分主要讨论当 FPPI＝1 时各算法的检测结果以便直观分析。在 INRIA 数据库中，当 FPPI＝1 时，本文提出算法的检测率是 77.21%，而对比算法的最好结果是 YOLO－v2 算法，其结果为 75.44%，而 ConvNet 算法的检测率是 69.1%，SSD 算法的检测率是 71.88%，R－CNN 算法的检测率只有 67.1%。其原因在于大多数深度特征的行人检测方法对行人仅仅是直接利用多层次深度信息，忽略了梯度消失的问题，而本文提出的改进 SDD 网络是将基本网络 VGG－16 替换为 Resnet－50，通过增加恒等映射解决网络层数加深但检测精度下降的问题，并采用强大而灵活的双参数损失函数来优化训练深度网络，提高网路模型的泛化能力。因此，本文提出的深度网络能够更好提取行人图像特征的深度残差网络，并通过改进的预测矩形框方法加强训练，进而进一步降低每张图片误检率。

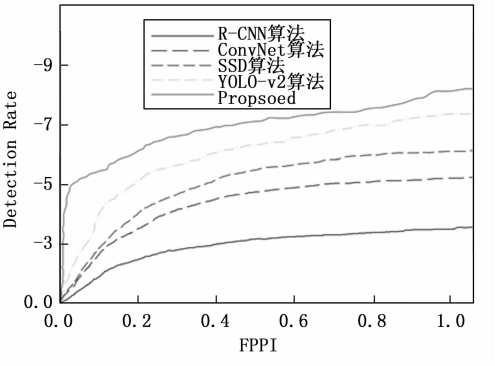

图 3 检测率与 FPPI 的关系曲线

图 4 展示了两幅比较有代表性的检测结果，其中一幅拍摄于昏暗的黄昏，导致视觉质量较差，严重影响行人目标的检测，非常符合油田野外环境下的成像实际。由于本文选择的对比算法较多，全部在一张图像显示不便于读者阅读，因此受限于篇幅，本文只对代表性的算法进行标注对比。如图 4 所示，红色的矩形框是本文算法检测到的行人结果，可以看出本文所提出的深度残差 SSD 训练模型能够比较准确的检测出复杂背景下的行人，尤其是针对模糊的行人目标，而直接采用 SSD 模型的检测结果存在一些弱小目标的漏检，主要归功于深度残差 SSD 的表征能力。但是，若图像中含有大量的行人且存在大量遮挡的时候，本文的模型也不能完全标注出所有的行人区域。一方面是大量遮挡使得行人结构变得千差万别，若不经推理很难分辨行人所在的区域。由此也可以说明，本文的算法仍然存在可以改进的地方。但从对比算法的结果而言，本文的方法在行人检测方面的性能具有更高的准确率。

(a)本文算法 (b)R-CNN (c)YOLO-v2 (d)SSD (e)ConvNet

图 4 不同算法的检测效果定性对比

4 总结

本文在 SSD 深度检测模型的基础上，提出了一种改进的行人检测算法，该算法首先利用聚合通道特征模型对图像或者视频序列进行进行预处理，获得疑似目标区域，大大降低单帧图像检测的时间；然后对 SSD 的基本网络 VGG－16 替换为 Resnet－50，通过增加恒等映射解决网络层数加深但检测精度下降的问题；最后采用强大而灵活的双参数损失函数来优化训练深度网络，提高网路模型的泛化能力。

（下转第 284 页）