基于深度学习的视频全量分析方法

肖磊，陈鑫，黄检宝
（南华大学计算机学院，湖南衡阳，421001）

基金项目：湖南省大学生创新创业训练计划项目：基于深度学习的视频结构化解析项目（编号：S202010555205）。

摘要：在深度学习领域中有众多的模型，但是目前几乎没有能解决目标检测、语义分割和多目标追踪的端到端模型，为了解决这个痛点，本文提出一种非端到端的视频全量分析方法，可针对目标检测、语义分割和多目标追踪对视频进行多维度分析，并实现对视频的全量分析，得到具有分割和追踪效果的可视化结果。

关键词：深度学习；全景分割 ；多目标物体追踪

DOI:10.16589/j.cnki.cn11-3571/tn.2020.18.023

0 引言

人工智能结合视觉分析，极大推动各行各业视觉应用。人脸识别、车辆检测、车辆智能驾驶等。结合当前高清1080P、2K、4K等视频，利用人工智能进行视觉分析技术，具有广泛的应用场景。

本文提出针对1080P视频，视频内容为街景内容（行车记录仪、电影等拍摄），需要利用视觉分析技术[1]，对高分辨率视频进行视频图像处理。本算法是专门针对街景或高楼的高清视频进行目标检测和统计。

1 算法简介

本文为了解决跟踪视频中的所有出现目标，统计每一类目标出现的数量，并将他们截取下来按类别保存，所以根据待解问题，我们在deepsort的基础上进行一些修改，实现多类别多目标追踪。

算法主体思想分为追踪和分割两部分。先对视频中出现的人物进行目标检测，识别出人和物，接着对视频中出现的人和物的状态和轨迹进行处理，得到人和物在某个视频的轨迹和数量，最后利用全景分割技术将视频中的人和物提取出来并配以相应的信息如颜色，时间，位置等，得到可供使用的信息。我们采用Deepsort与YOLOv3算法的结合体，进行目标的跟踪实现。相当于目标检测中的two stages的结构，采用detection + track，我们可以根据实际项目中的跟踪效果分别对detection部分（yolov3）和track部分（deepsort）采取一些优化手段，以实现我们的业务效果。在性能方面yolov3+Deepsort多目标追踪整体效果不错，可达到实时检测，yolov3主要作于检测目标，deepsort采用级联匹配算法，在sort算法的基础上添加马氏距离和余弦距离并添加深度学习特征进行尺度衡量，且Deepsort速度快，其中ID Switch-Tracking算法可用来解决遮挡物体，ReID算法可跨摄像头发现和跟踪同一个物体，可重新将被遮挡目标找回，降低被遮挡然后在出现的目标发生的ID Switch次数。对于分割方面，我们采用Detectron2项目，并使用全景分割模型Mask-rcnn，它使用起来方便，效果显著，训练速度快。Mask-rcnn提出Mask Prediction，我们利用这个mask信息得到人、物体的轮廓图，并且改良ROI Pooling提出ROI Align，这样使得被检测到的目标更加精确。综合这两点改进和继承之前模型优秀的设计，Mask-rcnn可以很好的展现全景分割的效果。

视觉追踪与视频物体分割的共同目标是预测第一帧中选定的任意物体的位置，但是视觉追踪通过边框坐标来表现目标物体，视频分割中物体的呈现是由一个二元分割mask构成，也就是某个像素点是否属于目标物体。相比于视觉追踪来说，视频物体分割是像素级上的预测，更加关注的是准确的物体表现，实时性需强，而目标检测算法yolov3实时性好，反馈信息更加真实，全景分割既关注像素级正确率也考虑到实例正确性，很好适应things类别的分割任务，也解决实例不重叠问题。视觉追踪是怎么用更好的Bounding Box锁定目标，以适应物体的形变、遮挡、消失等跟踪不准确的问题，改进的deepsort对街景视频有着很好的追踪效果，有效解决物体追踪出现的问题。

本文将追踪与分割两者融合，实现实例感知，在事先给定物体的先验知识信息的前提下进行场景中的多目标识别，借助全景分割，能实时分配语义标签，进行动态追踪，在不同场景中的物体相互位置不断变化时仍能进行实时分割、重建以及语义标注[2]。

2 算法分析

本文结合目标检测、多目标跟踪模型和全景分割，对街景视频进行分析和提取信息。

■ 2.1 目标检测模型YOLOV3[3]

基本的图像特征提取方面，YOLOV3使用称为Darknet-53的网络结构对图像的特征进行提取，然后利用多尺度特征进行对象检测，在预测对象类别时不使用softmax，改成使用logistic的输出进行预测，最后YOLOV3将其映射到3个尺度的输出张量，代表图像各个位置存在各种对象的概率。YOLOV3借鉴了残差网络结构，形

成更深的网络层次，以及多尺度检测，提升了 mAP 及小物体检测效果。

2.1.1 新的网络结构 Darknet-53

darknet-53 借用了 resnet 的思想，在网络中加入了残差模块，这样有利于解决深层次网络的梯度问题，每个残差模块由两个卷积层和一个 shortcut connections,1,2,8,8,4 代表有几个重复的残差模块，整个 v3 结构里面，没有池化层和全连接层，网络的下采样是通过设置卷积的 stride 为 2 来达到的，每当通过这个卷积层之后图像的尺寸就会减小到一半。而每个卷积层的实现又是包含卷积 +BN+Leaky relu，每个残差模块之后又要加上一个 zero padding, 其网络结构如图 1 所示。

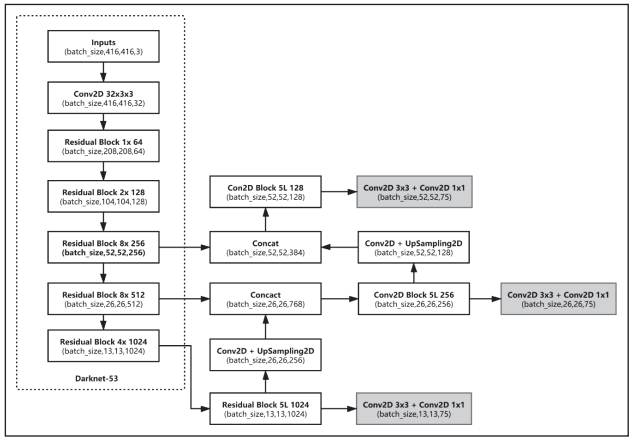

图 1　Darknet-53 网络结构

2.1.2 对象分类 softmax 改成 logistic

预测对象类别时不使用 softmax，改成使用 logistic 的输出进行预测。这样能够支持多标签对象（比如一个人有 Woman 和 Person 两个标签）。

输入 416×416×3 的图像，通过 darknet 网络得到三种不同尺度的预测结果，每个尺度都对应 N 个通道，包含着预测的信息；每个网格每个尺寸的 anchors 的预测结果。

对比 YOLOv1, 有 7×7×2 个预测；

对比 YOLOv2，有 13×13×5 个预测；

YOLOv3 共有 13×13×3+26×26×3+52×52×3 个预测。

每个预测对应 85 维，分别是 4（坐标值）、1（置信度分数）、80（coco 类别数）。

■ 2.2 MOT 模型 Deepsort[4]

在 deepsort 中，使用了更加可靠的度量来代替关联度量，并使用 CNN 网络在大规模行人数据进行训练，并提取特征，已增加网络对遗失和障碍的鲁棒性。在 deepsort 中使用 8 维空间去刻画轨迹在某时刻的状态，在轨迹处理中使用一个阀值 a 用于记录轨迹从上一次成功匹配到当前时刻，值大于 Amax 则认为改轨迹终止。在处理 Kalman 状态和新来的状态之间的关联度时它提出一种级联匹配的策略来提高匹配精度，主要由于当一个目标被遮挡很长时间，Kalman 滤波的不确定性就会大大增加，并会导致连续预测的概率弥散，假设本来协方差矩阵是一个正态分布，那么连续的预测不更新就会导致这个正态分布的方差越来越大，那么离均值欧氏距离远的点可能和之前分布中离得较近的点获得同样的马氏距离值。在最后阶段，使用 SORT 算法中的 IOU 关联去匹配 n=1 的 unconfirmed 和 unmatched 的轨迹。这可以缓解因为表观突变或者部分遮挡导致的较大变化。当然有好处就有坏处，这样做也有可能导致一些新产生的轨迹被连接到了一些旧的轨迹上。由此可得出人和物的运动轨迹和计数。

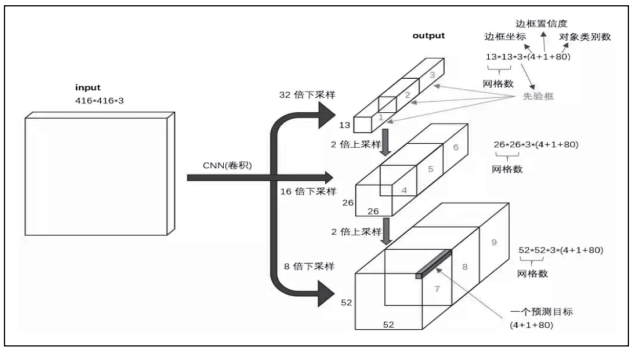

图 2　输入输出映射图

2.2.1 轨迹处理和估计

状态估计：使用一个 8 维空间去刻画轨迹在某时刻的状态 (u,v,r,h,x^*,y^*,r^*,h^*)。其分别表示为 bounding box 中心的位置、纵横比、高度、以及在图像坐标中对应的速度信息。然后使用 Kalman 滤波器预测更新轨迹，该 Kalman 滤波器采用匀速模型和线性观测模型。其观测量为 (u,v,r,h)。

轨迹处理：这个主要说轨迹什么时候终止、什么时候产生新的轨迹。首先对于每条轨迹都有一个阈值 a 用于记录轨迹从上一次成功匹配到当前时刻的时间。当该值大于提前设定的阈值 A_{max} 则认为改轨迹终止，直观上说就是长时间匹配不上的轨迹认为已经结束。然后在匹配时，对于没有匹配成功的检测都认为可能产生新的轨迹。但由于这些检测结果可能是一些错误警告，所以对这种情形新生成的轨迹标注状态“tentative”，然后观查在接下来的连续若干帧中是否连续匹配成功，是的话则认为是新轨迹产生，标注为“confirmed”，否则则认为是假性轨迹，状态标注为“deleted”。

2.2.2 分配问题

在 deepsort 中将目标运动和表面特征信息相结合，通过融合这两个相似的测量指标。

Motion Metric：使用马氏距离来测评预测的 Kalman

状态和新来的状态：

$$d^{(1)}(i,j)=\left(d_j-y_i\right)^T S_i^{-1}\left(d_i-d_j\right) \quad (1)$$

表示第 j 个 detection 和第 i 条轨迹之间的运动匹配度，其中 S_i 是轨迹有 Kalman 滤波器预测得到的在当前时刻观测空间的协方差矩阵，y_i 是轨迹在当前时刻的预测观测量，d_j 是第 j 个 detection 的状态（u,v,r,h）。

考虑到运动的连续性，可以通过该马氏距离对 detections 进行筛选，文中使用卡方分布的 0.95 分位点作为阀值 $t^{(1)}$ =0.4877，其中可以定义一个门限函数：

$$b_{i,j}^{(1)}=1\left[d^{(1)}(i,j)\le t^{(1)}\right] \quad (2)$$

Appearance Metric: 当目标运动不确定性较低时，马氏距离是一个很好的关联度量，但在实际中，如相机运动时会造成马氏距离大量不能匹配，也就会使这个度量失效，因此，我们整合第二个度量标准，对每一个 Bounding-Box 检测框 d_j 我们计算一个表面特征描述子 r_j，$\left|r_j\right|=1$，我们会创建一个 gallery 用来存放最新的 $L_k=100$ 个轨迹的描述子，即 $R_k=\left\{r_k^{(i)}\right\}_{k=1}^{L_k}$，然后我们使用第 i 个轨迹和第 j 个轨迹的最小余弦距离作为第二个衡量尺度。

$$d^{(1)}(i,j)=\min\left\{1-r_j^T r_k^{(i)}\middle|r_k^{(i)}\grave{o}R_i\right\} \quad (3)$$

当然，这也可以用一个门限函数来表示：

$$b_{i,j}^{(2)}=1\left[d^{(2)}(i,j)\le t^{(2)}\right] \quad (4)$$

接着，我们把这两个尺度相融合为：

$$C_{i,j}=\lambda d^{(1)}(i,j)+(1-\lambda)d^{(2)}(i,j) \quad (5)$$

$$b_{i,j}=\prod_{m=1}^{2}b_{i,j}^{(m)} \quad (6)$$

总之，距离度量对于短期的预测和匹配效果很好，而表现信息对于长时间丢失的轨迹而言，匹配度度量的比较有效。超参数的选择要看具体的数据集，比如文中说对于相对运动幅度较大的数据集，直接不考虑运动匹配程度。

■ 2.3 全景分割 Maskrcnn[5]

Mask R-CNN 是一个概念简单、灵活、通用的实例分割框架，能够有效地检测图像中的目标，同时为每个实例生成高质量的分割。而相对于 Faster R-CNN 模型，Mask R-CNN 是对 Faster R-CNN 模型的改进，使得模型训练的效率更高，并且 Mask R-CNN 推广性更强，比如能推广在识别同一个实例，不同的状态或者姿态，因此 Mask R-CNN 具有很好的实例追踪效果。

从图 3 可以看出，Mask R-CNN 是在 Faster R-CNN 的基础之上进行改进，在每一层的兴趣感知区域（Region of Interest）增加了一个 mask 分支。因此 Mask R-CNN 模型构建，更具灵活性，除此之外，这些 mask 分支占有较少的计算资源，所以更适用于快速进行实验。

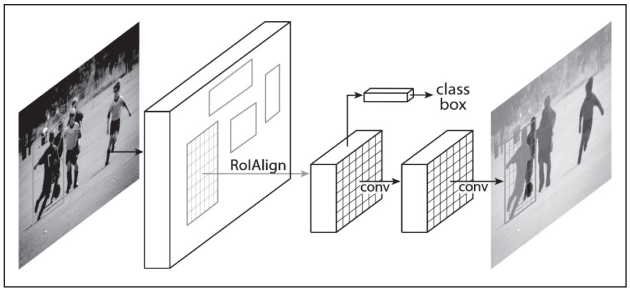

图 3　Mask R-CNN 模型框架

Mask R-CNN 基本结构，相对于 Faster R-CNN，采用了相同的两步：首先是找出 RPN，然后对 RPN 找到的每个 RoI 进行分类、定位、并找到 binary mask。这与其他网络是先找到 mask 然后在进行分类，所以与其他网络有明显的差别。

通过 detectron2 项目，使用全景分割模型 Mask-rcnn，Mask-rcnn 提出 Mask Prediction，我们利用这个 mask 信息得到人、物体的轮廓图，并且改良 ROI Pooling 提出 ROI Align，这样使得被检测到的目标更加精确。模型的 backbone 有多种选择，当选择 ResNet50+FPN 时，模型的推断时间会上去，但是模型会因为特征提取提取层数太浅而不能很好地识别训练数量少的物体和小物体，最后选择的 backbone 是 ResNeXt101+FPN，它能提取更深层次的特征。

3 总结

本次项目以街景图像作为观测城市物质空间的新型大数据源，我们采用动态追踪模型，实时跟踪物体轨迹。其结果如图 4 所示。

图 4　结果处理图

通过目标检测算法获取物体框信息，为进行实时计数，运用全景分割模型对国外城市图像数据进行训练和测试，为

（下转第 99 页）

继电器不能被励磁的不良情况，提高整体电力设备的运行能力，有效杜绝断路器的拒动隐患。

经该方案的实际落实情况可知，此改进方案可同时满足“两组跳闸回路直流之间不允许直流回路采用自动切换”和“若断路器操作机构箱内或保护操作箱内只有一组压力闭锁回路，电源消失时跳闸回路压力接点应处于闭合状态”的技术规范准则。由此便验证了改进思路方向正确，改进方案制定合理，能具备现场实践可行性。

■ 3.3 增加监视回路

继电器的状态是断路器中的必要巡检项目，因此在巡检人员不能及时监测到位的情况下，可在电路回路系统中增加监视回路，以此来严密监控继电器的实际运行状态，从基础运行层面中解决断电器一系列隐患。设计的监视回路系统，并联至两组继电器回路中，通过联锁回路方式，保障监视信号的稳定性、精准性，并将联锁回路的故障进行良好区分。在正常运行中，该监视回路并接至后台设备，电路装置运行时，监控仪表仪器前的值班人员可实时监控该联锁回路的运行情况，并观测打压行为的有效性。

当后台发出联锁回路的故障警报时，值班人员可立刻通过信息传输设备将故障险情通报有关部门，并派遣专业人员赶往故障现场抢修。该维修工作的重点部分，将会由监视回路的具体分析系统做出指示，明确故障位置后，则维修效率得到了充分保障。通常的重点检测位置位于继电器的线圈结构中，因此排查线圈运行状态，防止因线圈故障导致的两组继电器相继损坏情况，进而影响到电路设备的零压闭锁功能。

■ 3.4 加强巡检频率

最后，进行高频率、有效率的巡检视察工作，能为断路器设备的安全运行提供较好帮助，值班人员在巡视中，尤其要将重点项目做出高频检查，比如继电器、线圈等易损坏装置。继电器的巡视过程，应确保继电器设备该时段运行下的状态，通过与设备正常运行准则对比，判断继电器的舌片是否吸合；线圈的巡检项目包括线圈运行温度等状态，如果发现异常，务必及时上报隐患信息，并派遣维护单位赶赴现场解决隐患。

4 结论

综上，电网线路运行中，应实现断电继电保护装置的正常运作，保障电网线路及各级设备的安全运行。本文将断路器设备可能会发生的压力低闭锁分闸回路隐患做出了对策分析，通过以上方案的改进，使断路器在运行时能与分相操作箱紧密配合。满足该设备运行时的技术要求同时，又可将该电力系统的运行可靠性加以保障，有效解决运行隐患，提高电网稳定运行的技术能力。

参考文献

* [1] 王堃 , 黄岗 , 肖力 .220 kV 断路器操作箱内压力低闭锁分闸回路的隐患探究及改进方案 [J]. 农村电气化 ,2020(08):63-64.
* [2] 曾智桢 . 某 220kV 变电站 220kV 断路器 SF_6 闭锁回路隐患的研究 [J]. 电工技术 ,2018(24):15-16.
* [3] 黎慧 , 邵千 .ABB 技术 SF_6 高压断路器压力低闭锁回路的综合性改进方案 [J]. 通讯世界 ,2018(11):124-125.
* [4] 熊泽群 , 杨波 , 陈源 . 断路器液压机构压力低闭锁重合闸方式的实现 [J]. 安徽电力 ,2017,34(03):40-43.

（上接第 61 页）

降低模型之间的干扰，所作处理都是针对原图操作，并将全景分割得到的 mask 信息添加到追踪模型上，所训练的模型能识别 80 类物体，模型中框预测的 AP 达到 42.4，图片分割的 AP 约 38.5，具有识别数量多，视频分割流畅，动态追踪久等优势，但模型还存在改进之处，例如不是端到端模型，且当物体被遮掩时间过长会失去追踪效果。因此，本项目拟面向视频，动态图像，利用动态追踪、目标监测、全景分割模型，实时检测物体数量与分割。本项目研究成果具有可扩展性，可在三个方面进行应用，具有重要的现实意义和应用价值。在个人领域，通过街景图像绘制的街景地图，感知交通路况，帮助市民优化出行方案；在企业领域，通过市民行为轨迹、社会密集度等集中监控和分析，为公安部门指挥决策、情报研判提供有力支持；在政务服务领域，依托统一互联网地图实时服务平台，利用街道交通信息流，天空开放度，植被覆盖率等信息，实现城市规划信息化，一体化。

参考文献

* [1]邱锡鹏 . 神经网络与深度学习 [M]. 机械工业出版社 ,2020:105-126.
* [2] 阿斯顿・张 , 李沐等 . 动手学深度学 [M]. 人民邮电出版社 ,2019:106-110.
* [3]Joseph Redmon,Ali Farhadi.YOLOv3:An Incremental Improvement.arXiv:1804.02767, 2017.1-3
* [4]Alex Bewley,Zongyuan Ge,Lionel Ott,Fabio Ramos,Ben Upcroft.Simple Online and Realtime Tracking,arXiv:1602.00763, 2016.2-3.
* [5]Kaiming He,Georgia Gkioxari,Piotr Dollár,Ross Girshick. Mask R-CNN.arXiv:1703.06870, 2017.1-5.