

文章编号: 2095-2163(2021)12-0068-05

中图分类号: TP391

文献标志码: A

基于多任务学习的传感器融合目标检测

程书慧¹, 陈凌珊¹, 杨军典²

(1 上海工程技术大学 机械与汽车工程学院, 上海 201620; 2 德心智能科技(常州)有限公司, 江苏 常州 213002)

摘要: 本文针对多传感器融合的目标检测, 提出一种基于多任务学习的毫米波与相机目标级数据融合算法。利用任务间的强相关性, 采用硬共享机制的多任务学习模型, 完成目标位置及尺寸预测与目标类别的分类。网络模型使用批归一化与残差结构, 并对两个任务损失函数设置权重。模型训练集中加入某一传感器失效时的数据, 使模型在某一传感器失效时也能准确得到目标位置、尺寸和类别, 同时改善了相机对于距离检测的不确定性, 以及雷达对目标类别及横向位置检测的不准确性, 使融合后目标属性值更精确, 提高各任务的性能与泛化能力及智能车辆感知的稳定性与准确性。经实验验证, 本文算法得到了较高的分类准确率和召回率, 且回归任务的均方根误差较小。

关键词: 毫米波雷达; 相机; 多传感器融合; 多任务学习

Sensor fusion target detection based on multi-task learning

CHENG Shuhui¹, CHEN Lingshan¹, YANG Jundian²

(1 School of Mechanical and Automotive Engineering, Shanghai University of Engineering Science, Shanghai 201620, China;

2 Dexin Intelligent Technology (Changzhou) Co., Ltd., Changzhou 213002, China)

[Abstract] In this paper, aiming at multi-sensor fusion target detection, a millimeter wave and camera target-level data fusion algorithm based on neural network multi-task learning is proposed. Using the strong correlation between tasks, the multi-task learning model with hard sharing mechanism is used to complete the prediction and classification tasks, which improves the accuracy of the camera's range detection and the radar's accuracy of the target category and lateral position detection, so that the attribute value is more precise. The network model uses batch normalization and residual network and weights the loss functions of the two tasks at the same time. The data when a certain sensor fails is added to the training set to make the model accurately obtain the position, size and category of the target so as to improve the performance and generalization ability of each task and the stability and accuracy of intelligent vehicle perception. After experimental verification, the classification accuracy of the algorithm proposed in this paper is beyond 93%.

[Key words] millimeter wave radar; camera; multi-sensor fusion; multi-task learning

0 引言

随着无人驾驶领域的发展, 单个传感器的感知能力已不能满足复杂的交通场景需要, 多传感器融合逐渐成为汽车环境感知主要方式。目前, 毫米波雷达与相机是常用的传感器, 也是传感器融合的重要研究对象^[1]。相机可获得目标的纹理颜色等信息, 对目标分类有较高的识别能力, 且对横向位置的检测有较高的精度。但相机性能易受工作环境与光照的影响, 对深度信息的获取能力较差。毫米波雷达对目标纵向位置的检测有较高的精度, 且由于电磁波的特性具有穿透性好、抗干扰能力强等全天候的工作能力。但毫米波雷达检测得到稀疏点云, 分类能力较差。因此, 在工作特性上毫米波雷达与相

机完成优势互补, 进行多传感器融合可获取较为可靠的目标属性值^[2-3]。

基于单一的毫米波雷达目标检测, 张暄等^[3]基于雷达回波信号的距离多普勒图像作为输入, 进行图像处理完成雷达的目标检测。与传统的恒虚警目标检测算法作为对比, 较大的减少了虚警的数量。魏涛等^[4]提出一种基于机器学习的毫米波雷达车辆检测算法, 对毫米波雷达目标级数据预处理, 针对雷达检测点的各个属性构建决策树模型进行车辆有效目标的识别。基于毫米波雷达与视觉融合的目标检测, 目前多数采用由毫米波雷达确定图像检测的兴趣区域 (ROI, Region of interest) 的方案, 从而加快图像检测速度与准确度。赵望宇等^[5]基于毫米波雷达与视觉融合, 提出一种核相关的扩展卡尔

作者简介: 程书慧(1997-), 男, 硕士研究生, 主要研究方向: 多传感器融合、目标跟踪; 陈凌珊(1966-), 女, 博士, 教授, 硕士生导师, 主要研究方向: 新能源汽车技术、智能驾驶控制技术; 杨军典(1990-), 男, 硕士, 高级工程师, 主要研究方向: 毫米波雷达信号处理、多传感器融合算法。

收稿日期: 2021-05-26

曼滤波进行目标的跟踪。利用毫米波雷达的输出确定 ROI 后, 基于支持向量机 (SVM, Support Vector Machine) 进行车底阴影图像检测, 完成目标的检测任务。文献[6]研究了一种基于毫米波雷达与视觉融合的车辆检测系统。通过判断雷达检测点位置的预测值, 与计算值匹配生成航迹信息, 将毫米波雷达生成航迹信息与视觉检测融合, 完成目标的准确检测。

单一的雷达检测算法通过对雷达信号或数据处理完成目标的检测, 但不能确定目标的类别属性。融合系统利用毫米波雷达辅助相机完成目标检测, 但未考虑某一传感器失效时的情况, 且未较好的融合雷达与相机输出的属性值。因而, 若雷达失效则系统失去目标的精确位置, 若相机失效则检测系统将失去对目标的分类能力, 且两种检测方法均未输出目标的尺寸信息。针对上述问题, 本文提出了一种基于多任务学习的传感器融合目标检测的模型。模型可输出目标的位置、尺寸信息及预测物体的类别, 提高了传感器融合后目标属性值的精确度, 且融合结果不受一方传感器失效时的影响。

1 基于多任务学习的网络模型

相比于大多数机器学习模型, 多任务学习在性能上更有优势。若要得到性能优异的模型, 需要数以百万计的数据, 而很多领域获取数据集依赖高成

本的人力资源^[7]。多任务学习是在同一个模型中, 将具有相关性的任务同时进行训练, 利用任务之间相关性提高模型性能与泛化能力, 通过一个模型解决多个任务的机制, 可以解决数据样本集较少的问题^[8]。多任务学习模型常用基于参数的硬共享机制和软共享机制。硬共享机制是将所有隐藏层共享, 对于不同的任务指定特定的输出层。而软共享机制中每个任务都具有单独的模型, 不同任务模型间同一层的信息可进行交换^[9]。硬共享机制由于在每个任务中不存在独自的模型参数, 适用于强相关任务。软共享机制由于每个任务都有独立的模型, 则对任务间的相关性不做要求, 但模型个数与任务数相关会存在大量的参数^[10]。本文模型通过回归任务解决目标的位置与尺寸问题, 分类任务获得目标的类别。由于物体的位置为目标中心点所在位置, 不同类别的物体大小将影响目标位置, 两个任务具有强相关性。因此, 模型使用多任务学习的硬共享机制。

本文基于多任务学习进行视觉与雷达的融合, 其处理流程如图 1 所示。利用图像检测所得目标锚点信息与毫米波雷达所得点云作为输入, 经多任务学习网络模型训练得到目标的尺寸、位置与目标类别。为了解决某个传感器失效带来的影响, 在数据集中加入单个传感器检测失效时的数据进行训练。

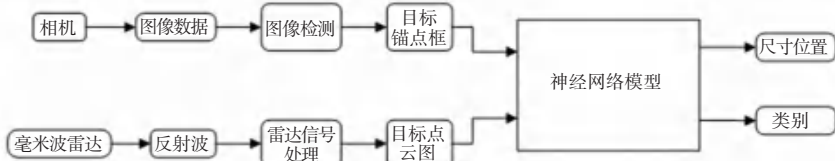


图 1 本文传感器融合流程图

Fig. 1 Flow chart of sensor fusion in this article

1.1 数据集分析

本文使用上海外高桥张杨北路实车采集的数据, 对不同场景下的检测数据进行筛选, 得到 15 万条数据样本。将数据集按 8 : 1 : 1 的比例划分为训练集、验证集与测试集。模型输入 10 个特征值数据, 预测任务输出 4 维数据, 分类任务对 6 种类别进行分类。输入特征为视觉检测框坐标, 4 个参数分别为: 检测框的左顶点横纵坐标、框的高度和宽度; 雷达检测点属性的 6 个特征值分别为: 检测点距离、横向距离 x 、横向速度 v_x 、纵向速度 v_y 、雷达散射截面积及雷达检测点的个数。其中, 雷达散射截面积表示目标在接收方向上反射雷达信号能力的度量,

其与距离有关且物体的面积越大 RCS 值越大, 则雷达检测点个数越多, 对该目标的识别能力越强。两种传感器检测结果如图 2 所示。图 2(a) 表示相机的检测结果, 黄色框为雷达检测目标的映射, 其余锚点框为图像检测所得物体; 图 2(b) 为雷达检测点部分属性信息; 图 2(c) 中横坐标为 x 轴, 纵坐标为 y 轴, 单位米, 白色方框为静态雷达检测点, 彩色框为动态监测点, 其中实体框表示雷达处理后得到目标。由于在复杂的道路场景中, 主车分析场景中目标的类别及其运动状态、尺寸和距离信息来判断是否对自身造成危险, 则本文对目标位置尺寸信息进行预测且对目标进行分类。回归任务包括目标的位

置及目标的长度与宽度,分类任务输出包括对轿车、火车、摩托车、自行车、行人及未分类的汽车6个类别进行分类。

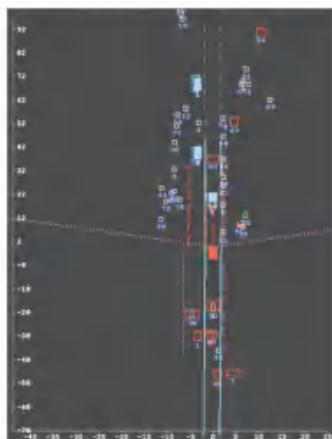
输入的各特征间量纲不同,且数值差异较大,会影响模型的收敛速度,且对数值较大的特征较为依赖,而忽略数值较小的特征,造成模型的精度下降。对输入特征数据进行线性归一化,将数值在0和1之间,减少特征间的相互影响,使模型均衡的学习各个特征。



(a) 相机检测图



(b) 检测数据属性值显示



(c) 毫米波雷达检测图

图2 传感器检测得到数据

Fig. 2 The sensor detection

1.2 网络结构

模型分为共享层、分类任务与回归任务3部分,如图3所示。3层隐藏层作为共享层,每层的神经元节点分别为10、256、128、256、64;分类任务3层全连接层,神经元节点数分别为64、256、128、6;预测任务3层全连接层,每层神经元节点数为64、256、64、4。归一化指数(*Softmax*)函数常用于解决分类问题,如公式(1)所示。带泄露随机线性整流

(*LeakyReLU*)函数可以避免梯度饱和与消失和神经元死亡的问题,如公式(2)所示,由经验值得知公式(2)中 α 取0.01。本文模型中,分类任务最后一层结点激活函数使用*Softmax*函数,其余激活函数均使用*LeakyReLU*函数。

$$\text{Softmax}(x_i) = \frac{e^{x_i}}{\sum_j e^{x_j}} \quad (1)$$

式中, x_i 为输出值中第*i*个元素, e 为自然对数。

$$\text{LeakyReLU}(x) = \max(\alpha x, x) \quad (2)$$

式中, x 为激活函数输入值, α 一般取0.01。

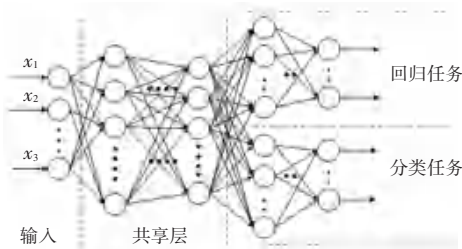


图3 多任务学习模型结构图

Fig. 3 Structure diagram of multi-task learning model

为了避免全连接层过多将造成梯度饱和或消失问题,对分类和预测任务每一层都加入批量归一化(*Batch Normalization*, *BN*)处理和残差结构。利用*BN*将某一层的输入分布变为均值为0方差为1的正态分布,使各输入分布保持一致,避免梯度饱和或消失问题,经过缩放平移还原数据自身的表达能力,使输入保持同一分布的同时,不丧失网络的表达能力,避免输入间的分布差异造成梯度问题。残差结构(*residual*)将不同层学习到的信息组合,让模型学习到更多的特征信息,避免全连接层过多带来的梯度问题同时加快了模型的收敛速度与稳定性。残差结构如图4所示。

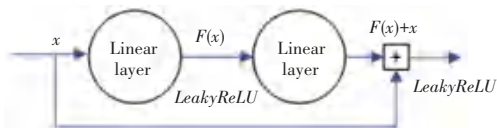


图4 残差结构图

Fig. 4 Residual structure diagram

分类任务使用交叉熵损失函数(*Cross Entropy Loss*, *CEL*)如公式(3)所示。预测任务使用均方误差损失函数(*MSE*, *Mean Squared Error*)如公式(4)所示。

$$\text{CEL} = \frac{1}{N} \sum_i L_i = -\frac{1}{N} \sum_i \sum_{c=1}^M y_{ic} \log(p_{ic}) \quad (3)$$

式中, c 为类别; i 为样本数量; y_{ic} 表示预测是否与真值相同(相同为1不同为0); p_{ic} 表示类别*c*的预测概

率。

$$MSE = \frac{1}{N} \sum_{i=1}^N (x_i - y_i)^2 \quad (4)$$

式中, x_i 表示预测值; y_i 表示真值; N 表示样本数量。

对任务模型需考虑任务间收敛速度, 使每个任务都能收敛至最优。若对分类与回归任务采用相同的学习率, 会使训练过程中两个任务模型后向传播的更新速度不同, 造成其中一个任务收敛很快, 另外一个任务收敛很慢的情况。对模型损失函数增加超参数 α 与 β , 模型后向传播时, 对学习率增加权重, 使两个任务具有不同的学习率, 从而达到较为接近的收敛速度。

2 模型训练及结果分析

本文算法在 Windows10 64 位操作系统上实现, 使用设备为 Intel(R) Core(TM) i7-10700 CPU、16GB 内存、显卡为 NVIDIA GeForce RTX2060、PyCharm2020 社区版、PyTorch 深度学习网络框架。

模型的参数设置为: 初始学习率为 0.001、批大小为 256、循环次数为 500。通过多次实验, 对分类任务的损失设定权重 0.8, 对回归任务的损失设置权重 10。使用自适应矩估计 (Adam) 优化器, 该优化器具有自动调节学习率, 加快模型训练过程, 减少模型参数调整等优点。

若对损失函数不做处理, 即任务间使用相同的学习率时, 模型中分类任务较回归任务收敛快。对各个任务损失函数加权重, 使任务具有不同的学习率。回归任务中相同与不同学习率的损失收敛对比如图 5(a) 所示。由图可见, 模型分类任务使用不同学习率的收敛速度更快。

模型分类任务中使用 BN 与残差结构与不做处理训练得到的准确率对比结果如图 5(b) 所示, 使用残差结构与 BN 的模型得到的分类任务的准确度与模型稳定性较好。

回归任务使用均方根误差作为评价指标, 如公式(5)所示; 分类任务使用准确率 (Accuracy) 与召回率 (Recall) 作为各个类别的评价指标, 如公式(6)和公式(7)所示。对测试集数据样本测试, 得到回归任务的各属性均方根误差值, 见表 1。传感器融合后对于目标的长度预测效果更好。分类任务的整体准确率为 94.62%, 召回率为 92.27%, 各类别的准确率 (Accuracy) 与召回率 (Recall) 见表 2。对于汽车、火车、行人预测较好, 对于机动性较强的小目标物体分类能力有进步空间。

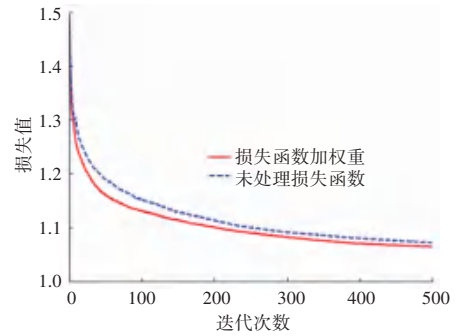
$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - y_i)^2} \quad (5)$$

式中, N 为样本数量; x_i 表示预测值; y_i 表示真值。

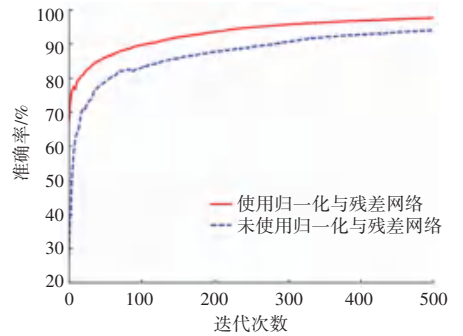
$$Accuracy = \frac{TP_c}{N_{true}^c} \quad (6)$$

$$Recall = \frac{TP_c}{N_{prediction}^c} \quad (7)$$

式中, TP_c 表示 c 类的预测值与真实值匹配数量; N_{true}^c 表示样本中实际为 c 类的样本数量; $N_{prediction}^c$ 表示预测为 c 的数量。



(a) 损失函数加权后回归任务损失值对比



(b) 使用归一化与残差网络分类任务准确度对比

图 5 处理前后模型对比

Fig. 5 Comparison of models before and after processing

表 1 回归任务的评价指标

Tab. 1 Evaluation metrics of regression task

目标属性	横向位置	纵向位置	宽度	长度
均方误差	0.595 4	0.473 99	0.523 380	0.168 2

表 2 分类任务的评价指标

Tab. 2 Evaluation metrics for classification tasks

类别	准确率/%	召回率/%
汽车	98.87	96.67
货车	99.21	99.68
摩托车	89.65	93.28
自行车	85.62	77.65
行人	95.37	88.76
未分类的车	99.05	97.55

(下转第 77 页)