

# 基于分段线性激活的多任务行人目标检测识别算法研究

朱亚旋 张达明 尹荣彬 吴继超

(一汽(南京)科技开发有限公司,南京 211100)

【欢迎引用】朱亚旋,张达明,尹荣彬,等.基于分段线性激活的多任务行人目标检测识别算法研究[J].汽车文摘,2024(4):38-43.

【Cite this paper】ZHU Y X, ZHANG D M, YIN R B, et al. Research on Multi-Task Pedestrian Target Detection and Recognition Algorithm Based on Piecewise Linear Activation Function[J]. Automotive Digest (Chinese), 2024(4): 38-43.

【摘要】通过在已有检测网络结构的基础上增加解耦预测支路,实现了车载摄像头下行人位置、人头位置、行人年龄与遮挡属性识别任务,经过试验论证了分类支路在不同网络位置对检测性能的影响。针对网络部署中一些非线性激活函数算子不受工程环境支持这一问题,提出了一种使用分段线性函数来拟合非线性激活函数的方法。经过验证表明,该方法既提升了网络识别的性能又为工程部署提供了便捷性。

关键词:多任务识别;解耦预测;拟合非线性;激活函数

中图分类号:TP277 文献标志码:A DOI: 10.19822/j.cnki.1671-6329.20220316

## Research on Multi-Task Pedestrian Target Detection and Recognition Algorithm Based on Piecewise Linear Activation Function

Zhu Yaxuan, Zhang Daming, Yin Rongbin, Wu Jichao

(FAW(Nanjing) Technology Development Co., Ltd., Nanjing 211100)

【Abstract】On the basis of the existing detection network structure, a new decoupling prediction branch is added to realize the multi-task recognition of pedestrian position, head position, pedestrian age and occlusion attributes under the vehicle-mounted camera in this paper. The influence of classification branch on detection performance at different network locations is demonstrated by experiments. Aiming at the problem that some nonlinear activation function operators in engineering applications are not supported by the deployment environment, a method using piecewise linear function to fit the nonlinear activation function is proposed, which not only improves the performance of network identification but also provides convenience for engineering deployment.

**Key words: Multi-task recognition, Decoupling prediction, Nonlinear fitting, Activation function**

### 缩略语

PAN Path Aggregation Network

FPN Feature Pyramid Networks

YOLOX You Only Look Once X

### 0 引言

随着神经网络技术与信号采集技术的发展,基于

深度学习的视觉感知识别技术越来越多地应用于生产生活当中。而受限于硬件算力支持与工程应用环境的局限性,基于多任务集成、算子优化的神经网络开发需求与日俱增。

在当前学术领域中,基于“编码-解码”(Encoder-decoder)结构的目标检测网络近年来不断地刷新着各大公开数据集的指标记录,其中YOLOX<sup>[1]</sup>(You Only Look Once X)作为一个“无需设定锚框”(Anchor-free)

方法,备受业界学者与技术人员青睐。然而,在实际智能网联汽车研发需求中,除了对目标检测与分类有识别需求外,图像中目标的其他属性,甚至图像中除目标外的其他背景信息均有着智能化识别的需求,良好的多任务模型<sup>[2]</sup>是解决该类问题的首选方案。

通过设计网络结构,实现目标位置与类别预测的同时,实现其他维度的属性信息识别,再通过将网络模型工程化部署至相应的硬件平台,最后集成到产品中,是当前工业界的主流智能化开发流程。其中,如何设计一个好的多任务预测结构,如何快速而又稳定地将网络模型部署到相应的硬件平台是当前技术开发的主要挑战。

在使用神经网络进行训练时,目前学术界常用的 Mish<sup>[3]</sup>与 Swish<sup>[4]</sup>激活函数均能很好地满足梯度回传需求,两者函数形状分布较为相似,且均满足无上界有下届、平滑、非单调的特性,性能在总体上优于 Relu 函数效果。然而,在实际部署中,因两者计算方式复杂,不利于网络特征量化后的数值计算,甚至 Mish 算子在一些部署环境中存在算子不支持的情况。

本文针对这一问题,提出了一种针对复杂激活函数映射进行拟合的方案,在保障网络性能的基础上,去除了激活函数中的非线性运算部分,方便了后续的工程化部署。同时,为了达到目标检测与图像内容分类任务的最优性能,本文通过试验,探究了不同的识别任务之间的相互影响,基于不同的损失约束对网络进行训练。

## 1 多任务识别网络

基于实际的生产应用需求,在 YOLOX 网络结构的基础上,增加了一个分类支路来实现目标检测与图片相关属性分类的多任务识别,探究了分类支路在网络不同位置处对检测性能的影响。

### 1.1 解耦多任务支路

图 1 为所采用的神经网络方案处理流程,在原有的 YOLOX 网络方案的基础上,提出了一种解耦的多任务识别结构,输入为待识别图像,输出则既有目标的检测框与类别信息又有目标相关的其他属性(如年龄、遮挡率、头部遮挡率)信息。

从图 1 可以看出,新增的分类支路的作用是判断图片中目标的年龄、目标的遮挡程度。相较于额外使用一个分类网络模型针对相同的输入图像进行识别,将其集成至原有的目标检测网络,只需要在网络结构上新增一个支路即可,在时效性上,后者具有巨大优势。

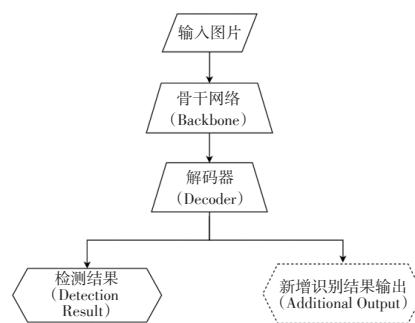


图 1 多任务网络识别流程示意

图 2 为目标新增属性支路在网络模型中的可拼接位置示意,可以看出,对于新增的网络属性分类,其既可以基于检测支路的特征图进行分类<sup>[5]</sup>,也可以在检测支路处理前接入解耦支路进行分类识别<sup>[6]</sup>,本文基于这 2 种选择分别进行了实验。

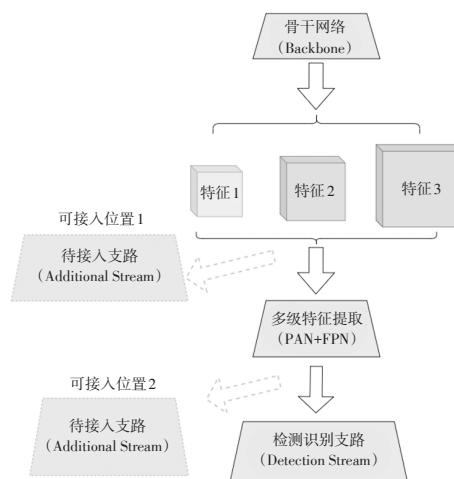


图 2 新增属性分类支路在网络中的可选位

### 1.2 多维度属性分类

图 3 为新增多属性分类通道示意。本文所设计的网络要实现目标包括年龄阶段(成人、小孩)、人体遮挡程度、头部是否遮挡 3 个维度属性,且属性信息彼此之间相互独立,针对这一情况,本文采用了“一头多通道”的方式<sup>[7]</sup>,将输出的特征图通道进行拆分,分别计算各自属性损失。

根据不同分类任务实际分类类别数的情况,具体的损失函数的设计如式(1)所示

$$L_{cls} = L_{bce1} + L_{bce2} + L_{ce3} \quad (1)$$

式中:  $L_{ce}$  表示分类常用的多分类交叉熵损失函数,  $L_{bce1}$  为目标年龄的分类损失,  $L_{bce2}$  为目标头部是否遮挡的分类损失,  $L_{ce3}$  为整体遮挡程度的分类。

## 2 多任务损失配比方法

为了更好地了解不同任务支路同时训练时彼此

之间的相互性能影响,本研究通过对分类支路的损失添加比例系数 $k^{[8]}$ ,如式(2)所示:

$$Loss_{total} = loss_{det} + kloss_{cls} \quad (2)$$

式中: $k \in (0.01, 2)$ ,  $Loss_{total}$  为所有任务的损失总和,  $loss_{det}$  为检测支路损失,  $kloss_{cls}$  为分类支路的损失。

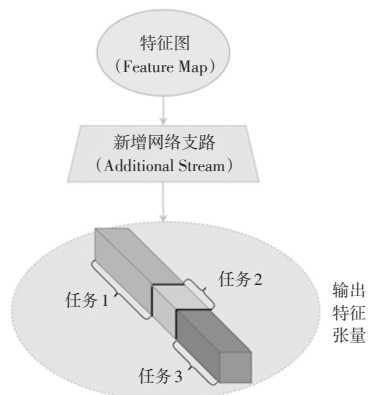


图3 新增的多属性分类输出示意

对于检测支路,分别基于当前目标检测领域常用的 ciou、diou<sup>[9-10]</sup>损失函数进行训练,而多属性分类支路,考虑到属性之间的独立性,本文网络训练时采用的是1:1:1的比例即对不同属性分类损失进行同等比重计算损失。

### 3 分段线性拟合方法

图4和图5分别为 Mish 和 Swish<sup>[3-4]</sup>激活函数的分布形状,其函数表达分别为式(3)和式(4)。

$$f(x) = x \cdot \tanh(\ln(1 + e^x)) \quad (3)$$

$$f(x) = x \cdot \text{sigmoid}(\beta x) \quad (4)$$

式中: $x$  为输入特征张量元素值; $\beta$  为该函数引入的超参,可根据实际应用场景调节。

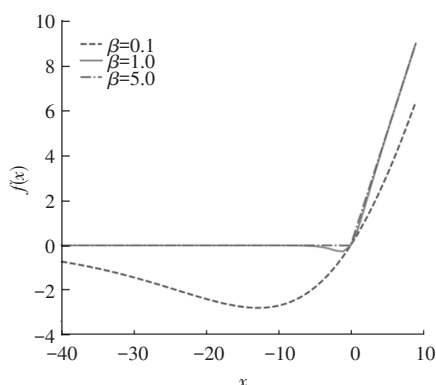


图4 Swish激活函数分布

可以看出 Mish 与 Swish 在 $(-\infty, 0)$ 区间内的分布形状较为相似,核心思想是当输入的 $x$ 值非常小时,其梯度值变化较小,而在趋近于0的一段区间内,存在非线性的同时,梯度分布有正有负<sup>[11]</sup>。

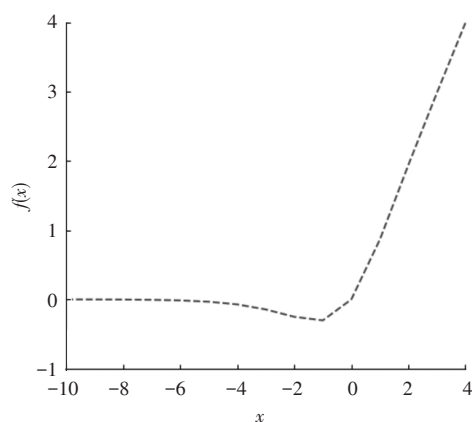


图5 Mish激活函数分布

针对这一现象,基于使用正多边形来拟合圆形的思想,提出了一种分段拟合的激活映射方式。该方法共分为2个阶段,首先,使用过原点的线性函数与分段的边界值来确定每个激活点位置,然后通过两个相邻的激活点来实现该分段区间的线性激活函数的映射。

#### 3.1 激活点位置确定

图6所示为本文所示方案的激活点位置选取的示意,其中每一个分界区间点横坐标为 $x$ ,其对应的激活点纵坐标 $f(x)$ 的计算方式如式(5)所示:

$$f(x) = \begin{cases} x \cdot e^{(x-\alpha)/\beta} & x < \alpha \\ x \cdot e^{(\alpha-x)/\beta} & \alpha < x < 0 \end{cases} \quad (5)$$

式中: $\alpha$ 、 $\beta$  为该拟合方案的参数,实际应用时,可根据试验效果来选取对应的值。

本研究经过试验发现,在自采的车载行人数据集中,取 $\alpha=0$ 和 $\beta=2$ 时模型识别效果最好,不同的取值对应的激活点分布见图6。

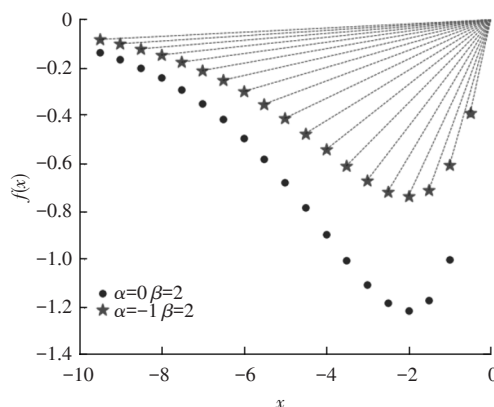


图6 不同分段点处对应的激活点值

#### 3.2 分段线性拟合

在得到一系列指定区间的激活点之后,对相邻的2个端点进行连接,得到分段的线性映射函数<sup>[12]</sup>,这一系列分段的线性映射函数最终能够拟合原有非线性

函数映射的效果。

如图7所示,当取 $\alpha=0$ 、 $\beta=1$ 时,本方案与Swish及Mish函数的分布较为相近。相较于Swish所实现的非线性映射形态,本文提出的方法通过2个参数能够实现现在 $x < 0$ 范围<sup>[13]</sup>内引入非线性映射效果的同时,拥有更宽的值域映射区间,梯度分布较为均匀,有利于网络的训练和推理。

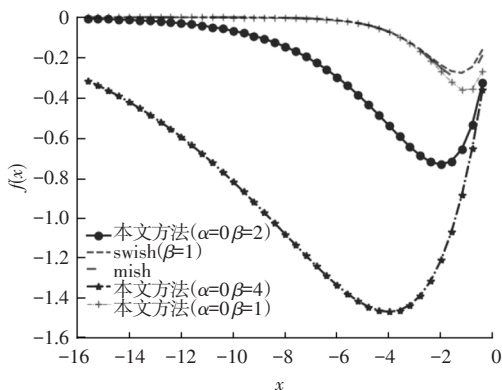


图7 分段拟合效果与原有激活函数在(-16,0)范围内的映射分布对比

#### 4 试验结果与分析

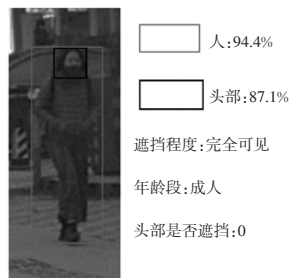
为了验证本文所提出的网络结构的识别效果,现使用车载摄像头进行图片的采集,对包含目标的区域进行裁剪,得到11 197张图片包含26 110个目标检测框用于训练,使用2 143张图片包含5 149个目标框进行测试,由于车载相机成像分辨率较高,裁剪后的远处行人较于原图属于较小的目标,且区域清晰度较低。

图8为网络识别的效果图展示,其中矩形框部分为目标检测的类别与区域信息,文字部分为对图中目标的分类识别的结果。

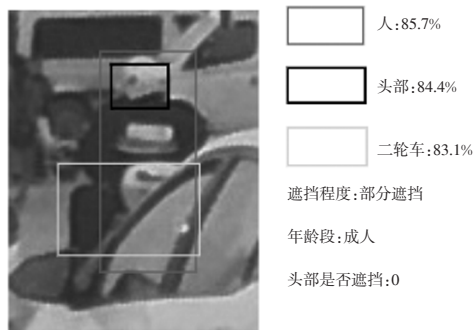
针对前文所提出的一些问题,本文分别从分类支路位置、分类损失权重以及替换拟合函数后的模型性能表现3个方面进行了对比试验并计算了量化指标。



(a)模型识别可视化效果示例1



(b)模型识别可视化效果示例2



(c)模型识别可视化效果示例3

图8 部分测试图片识别效果示例

#### 4.1 分类支路对检测支路的影响

表1为新增的分类支路与检测支路在结构上是否解耦的性能表现,其中检测支路主要预测人体(Person)、人头(Head)和非机动车(Non-motor vehicle)3个类别目标框位置,而属性分类支路既可以使用检测支路的特征,也可以另起一个支路。

表1 预测支路在不同位置的效果 %

评测维度 分类支路位置	行人框(ap)	人头框(ap)	非机动车框(ap)	总体遮挡程度(acc)	人头是否遮挡(acc)	目标年龄(acc)
使用解耦支路特征图	81.60	57.20	88.00	71.58	98.30	92.95
使用检测支路特征图	76.80	50.10	80.10	61.70	98.18	89.20

注: ap为平均精度(average precision),为目标检测常用指标; acc为精度(accuracy),为分类常用检测准确率指标。

由表1可知,对于新增分类属性的需求,将新增分类的支路与原有的检测支路解耦,确实有利于神经网络对两个任务的学习。

在额外的分类支路已经解耦的条件下,由于分类支路与检测支路虽然识别任务不同,但都是对图像特征语义的理解与表达,两者存在部分共同的特征编码阶段,因而两个支路的损失函数梯度回传而存在一定的相互影响。

本文在实际训练时发现,分类支路的损失值(loss)收敛速度较快,故为了达到较好的综合性能表现,避免因分类支路的网络权重迭代影响目标检测性

能的表现,对分类支路的 loss 梯度回传时的权重设置较小的比例,故设置  $k$  分别为 0.01、0.10、0.50<sup>[14]</sup> 时进行对比试验,结果如表 2 所示。

表 2 分类支路损失在不同权重下的影响 %

评测维度 分类支路损失权重 $k$	行人框 (ap)	人头框 (ap)	非机动车框 (ap)	总体遮挡程度 (acc)	人头是否遮挡 (acc)	目标年龄 (acc)
$k=0.01$	78.00	54.67	82.30	57.40	98.20	91.10
$k=0.10$	79.20	58.20	86.80	70.20	97.50	92.30
$k=0.50$	81.60	57.20	88.00	71.58	98.30	92.95

为了提升网络对目标框的检测性能,分别采用 giou、ciou<sup>[15-16]</sup> 损失函数与 YOLOX 网络原生的损失函数进行对比试验,结果如表 3 所示。

表 3 不同 iou 损失对检测框性能的影响 %

评测维度 检测损失权重	行人框(ap)	人头框 (ap)	非机动车框(ap)
原损失函数	81.60	57.20	88.00
giou loss	79.00	54.80	86.00
ciou loss	78.00	54.67	82.30

由表 3 可以看出,本文所涉及的网络在使用 YOLOX 原生损失约束的条件下,检测性能最好,初步分析可能是人头的目标区域过小,而 giou 和 ciou 损失函数所计算的信息在目标过小时对检测性能的促进意义不明显。

#### 4.2 拟合激活函数训练

表 4 所示为使用本文的线性拟合方案与原生非线性激活函数在相同网络结构下的训练效果对比。

表 4 激活函数拟合方法对模型性能的影响 %

评测维度 激活函数	行人框 (ap)	人头框 (ap)	非机动车框 (ap)	总体遮挡程度 (acc)	人头是否遮挡 (acc)	目标年龄 (acc)
Swish ( $\beta=1$ )	81.60	57.20	88.00	71.58	98.30	92.95
线性拟合方案 (本文方法)	81.20	58.34	88.12	71.88	98.18	92.98

如表 4 所示,在同等网络结构、训练与验证数据集条件下,使用本文自定义的分段激活函数进行训练,个别类别的检测效果甚至优于使用原激活函数时的效果,总体性能表现与原方案能够保持对齐。

在试验过程中发现,使用该方案并没有产生额外的时间成本,因此,所设计的分段线性拟合激活方案能够实现常见非线性激活函数的完美替换。

## 5 结束语

本文通过试验验证了网络模型训练时,不同任务支路解耦对总体性能的影响,同时发现在人头等小目标检测方面,即使一些损失约束在理论上存在改进,但是实际网络训练时效果提升不大。

本文结合理论与实践,提出了以分段线性函数拟合非线性激活函数的方法,试验表明,该方法能够完美地替换非线性激活函数在网络中的作用。相较于传统非线性激活函数,本文设计的拟合方法能够映射的特征值域范围更广,在工程部署上更加便捷,同时没有额外的时间成本,有着重要的理论研究意义与工程应用价值。

最后,受限于采集与标注条件,试验所使用的图片质量较为驳杂,选取的目标为车载摄像头中的远端行人与骑行行人区域的裁剪,后续可以在数量更多、分辨率更高的数据集下进行试验研究。

### 参 考 文 献

- [1] ZHENG G, SONG T, LIU, FENG W, et al. YOLOX: Exceeding YOLO Series in 2021[C]. CVPR(Computer Vision and Pattern Recognition Conference), 2021.
- [2] JIAN X. exYOLO: A Small Object Detector Based on YOLOv3 Object Detector[J]. Procedia Computer Science, 2021, 188(10): 18-25.
- [3] PRAJIT R, BARRET Z, QUOC V L. Swish: a Self-Gated Activation Function[J]. Neural and Evolutionary Computing, arXiv:1710.05941, 2017.
- [4] ALEXEY B, CHIEN-YAO W, HONGYUAN M L. Yolov4: Optimal Speed and Accuracy of Object Detection[C]. CVPR, (Computer Vision and Pattern Recognition Conference), 2020.
- [5] 王彩云. 目标检测的研究进展[C]. 中国计算机用户协会网络应用分会 2019 年第二十三届网络新技术与应用年会论文集, 2019.
- [6] MENG TIAN L, YUX W, DEVA R. Towards Streaming Perception[C]. ECCV(European Conference on Computer Vision), 2020.
- [7] 张有健, 陈晨, 王再见. 深度学习算法的激活函数研究[J]. 无线电通信技术, 2021, 47(1): 115-120.
- [8] 徐增敏, 陈凯, 郭威伟, 等. 面向轻量级卷积网络的激活函数与压缩模型[J]. 计算机工程, 2022, 48(5): 242-250.
- [9] 朱亚旋, 张小国, 陈刚. 基于图像纹理与矩特征的转子绕线检测研究[J]. 测控技术, 2018, 37(2): 16-19.
- [10] 王攀杰, 郭绍忠, 侯明, 等. 激活函数的对比测试与分析[J]. 信息工程大学学报, 2021, 22(5): 551-557.

- [11] 付志航. 基于空间与时间上下文的深度学习目标检测算法研究[D]. 杭州: 浙江大学, 2019.
- [12] 尤轩昂, 赵鹏, 慕晓冬, 等. 基于多任务学习的可见光与近红外虹膜融合研究[J]. 计算机工程与应用, 2022, 58(21): 197-204.
- [13] 王晓峰, 王昆, 刘轩, 等. 自适应重加权池化深度多任务学习的表情识别[J]. 计算机工程与设计, 2022, 43(4): 1111-1120.
- [14] ROBERTO C, YARING, ALEX K. Multi-task Learning Using Uncertainty to Weigh Losses for Scene Geometry and Semantics[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018.
- [15] HAMID R, NATHAN T, JUN Y G, et al. Generalized Intersection over Union: A Metric and A Loss for Bounding Box Regression[C]. CVPR, (Computer Vision and Pattern Recognition Conference), 2019.
- [16] ZHAO H Z, PING W, DONG W R, et al. Enhancing Geometric Factors in Model Learning and Inference for Object Detection and Instance Segmentation[J]. IEEE Transac-

tions on Cybernetics, 2022, 52(8):8574-8586.

(责任编辑 梵铃)

#### 【作者简介】

朱亚旋(1994—), 男, 一汽(南京)科技开发有限公司, 硕士研究生, 研究方向为车载视觉障碍物、交通元素感知算法。

E-mail:zhuyaxuan@faw.com.cn

张达明(1994—), 男, 一汽(南京)科技开发有限公司, 硕士研究生, 研究方向为车载视觉车道线感知算法开发。

E-mail:zhangdaming6@faw.com.cn

尹荣彬(1989—), 男, 一汽(南京)科技开发有限公司, 学士, 研究方向为车载智能化高低阶辅助驾驶技术合成。

E-mail:yinrongbin@faw.com.cn

吴继超(1992—), 男, 一汽(南京)科技开发有限公司, 硕士研究生, 研究方向为基于车载信号的地图定位算法。

E-mail:wujichao1@faw.com.cn