基于深度卷积神经网络的人物检测改进算法

周 杨 杨文柱* 申 远
(河北大学网络空间安全与计算机学院 河北 保定 071000)

摘 要 基于深度卷积神经网络的人物检测方法是目前检测效果最好的方法。在同等环境下,YOLOv3 运行速度最快,但其采用的非极大值抑制算法(NMS)导致很多正确的检测框被错误移除。通过加入取回算法来恢复被 NMS 错误移除掉的人物检测框,而且将 NMS 替换为 Soft-NMS 进一步提高了准确率。在 PASCAL VOC 数据集上的实验表明,使用 Soft-NMS 和取回算法改进的 YOLOv3 相比于原算法提升了大约 3.1 百分点的准确率,同时运行速度没有发生太多的变化。

关键词 人物检测 非极大值抑制 取回算法 深度卷积神经网络

中图分类号 TP391 **文献标志码** A DOI:10.3969/j.issn.1000-386x.2022.07.033

IMPROVED HUMAN DETECTION ALGORITHM BASED ON DEEP CONVOLUTIONAL NEURAL NETWORK

Zhou Yang Yang Wenzhu* Shen Yuan
(*School of Cyber Security and Computer, Hebei University, Baoding 071000, Hebei, China*)

Abstract Human detection algorithm based on deep convolutional neural network is the most effective method at present. Under the same circumstances, YOLOv3 ran the fastest, but its None Max Suppression(NMS) algorithm caused many correct detection frames to be removed by mistake. The incorrectly removed human bounding boxes could be recovered by using the get back algorithm. NMS was replaced by Soft-NMS to further improve the accuracy. Experiments on the PASCAL VOC dataset show that compared with the former model, the accuracy of the improved YOLOv3 increases by about 3.1 percentage points and its speed does not change too much.

Keywords Human detection NMS Get back algorithm Deep convolutional neural network

0 引 言

人物检测通常用于检测图像中是否存在人目标,之后再获取图像中人目标的坐标。传统的方式中,HOG 和 SVM 通常被用于人物检测,但这非常消耗时间,通常准确度也不太高。深度卷积神经网络在模式识别中效果表现非常良好,基于深度卷积神经网络的人物检测模型现在变成了主流模型。最近几年提出了很多新的基于深度卷积神经网络的模型。Ren 等[1]构建了用于物体检测的深度卷积神经网络 Faster R-CNN,其包含两个平行的子网络,分别用于生成目标框的类置信度和提取这些目标框的位置信息。Redmon 等[2]提出了一种深度卷积神经网络——YOLO。YOLO 是一个没有区域建议部分的网络。Liu 等[3]提出了 SSD 网络,这像是 YOLO 网络和区域建议网络的混合体。Lin 等[4]提出了 RetinaNet,它使用 focal loss 来计算损失值。He 等[5]提出了 SPP-net,该网络把输入图像转换成了一个固定大小的图像。Dai 等[6]提出了 R-FCN,该网络修改并提高了 Faster R-CNN 的 RoI 池化部分。在这些网络中,YOLOv3 是速度最快的一个,而且准确率很高。YOLOv3 使用 Darknet-53 作为深度卷积神经网络部分来提取输入图像的特征,提取特征并预测检测窗口。

YOLOv3 使用了传统的 NMS(非极大值抑制)算法[7],该算法可以替换为最新的 Soft-NMS[8]来提高准确率。在 YOLOv3 的物体检测过程中,传统的深度卷积神经网络会计算得到很多重复的目标检测框,NMS

收稿日期:2019-12-04。**周杨**,硕士生,主研领域:目标检测。**杨文柱**,教授。**申远**,硕士生。

是用来消除这些重复的窗口的。但是传统的 NMS 移除了所有的重叠率超过阈值的检测框,这导致了很多正确预测的检测框被移除。尽管 Soft-NMS 提高了准确率,但还是有一些正确预测的检测框被移除。所以我们加入了取回算法,该算法可以恢复丢失的目标框,从而提高了准确率。

1 相关工作

1.1 传统人物检测方法

对于传统的人物检测,目标物体的特征首先作为模板提取了出来。之后使用不同尺度的滑动窗口用于裁剪图像。裁剪出的小块图像的特征会被提取出来,基于相似度和模板比较来确定它们是否属于目标物体。HOG[9] 和 SIFT[10] 通常用于获得特征。最后,会使用数据集来训练一个支持向量机来预测输入物体是否是一个人。传统的方式非常消耗时间,所以需要一种更快的方式。

1.2 用于人物检测的 YOLO 系列

YOLO 系列模型是目前最佳人物检测模型。YOLO 有三个版本,是 YOLOv1[11]、YOLOv2[12] 和 YOLOv3。YOLOv1 只是简单地把输入图像分成了几个格,并使用 GoogleNet[13] 对每个格进行检测框预测。YOLOv2 基于 YOLOv1 的模型基础上微调了分类网络。其使用了高分辨率的分类器,并像 Faster R-CNN 一样使用锚点。YOLOv2 还使用了 Darknet-19 作为深度卷积神经网络部分。

YOLOv3 使用了 Darknet-53 作为深度卷积神经网络部分用于从输入图像中提取特征。在使用像 MS-COCO[14] 这样的数据集来训练之后,就可以提取特征并预测物体的检测框了。它会在 3 个不同的尺度上预测检测框,这比较像是一个特征金字塔网络[15],这样 YOLOv3 就可以识别非常大的或者非常小的物体。在网络的结尾,YOLOv3 会得出输出数据,输出数据包括 4 个检测框的偏移、1 个目标置信度和 80 个类置信度。YOLOv3 把输入图像分成了很多单元,每个单元会预测几个检测框,对于相同的物体会存在很多重复错误预测的检测框。NMS 算法通常被用于移除这些重复的检测框。

由于简单的结构和 Darknet-53 的原因,YOLOv3 相比于其他基于深度卷积神经网络的人物检测模型,运行得非常快。使用 Nvidia 的 GTX-1080 或者 Titan-X 就可以实时地检测人目标。所以,它非常适合应用在工程中用于满足日常需要。

1.3 NMS 算法

NMS 是一种贪心算法,应用在计算机视觉检测上已经很多年了。NMS 可以被用于边缘检测[16]、特征检测、人脸检测[17] 和物体检测。

对于人物检测,NMS 算法会处理 YOLOv3 深度卷积神经网络生成的检测框数据。首先,它会通过检测框的置信度来排序,找出置信度最高的检测框 M。然后再计算其他更低置信度的检测框和 M 的重叠率,并且设置了一个重叠率阈值来确定检测框是否应该被移除。如果一个检测框的重叠率大于等于该阈值,它就会被移除出检测框列表。在这之后,检测框 M 会被加入到最终的结果列表中,第二高置信度的检测框会成为检测框 M,之后再计算剩下的更低置信度的检测框和新的检测框 M 的重叠率。这个过程会一直持续到列表中没有检测框为止。NMS 算法如算法 1 所示[8]。

算法 1 NMS 算法

输出:$B=\{b_1,b_2,\cdots,b_N\}$,$S=\{s_1,s_2,\cdots,s_N\}$,N_t。

B 是初始检测框列表。

S 包含了相应的检测数值。

N_t 是 NMS 阈值。

```
Begin
    D←{}
    While B≠empty do
        m←argmax S
        M←m
        D←D∪M; B←B − M
        for b_i in B do
            if iou(M, b_i) ≥ N_t then
                B←B − b_i;
                S←S − s_i
            end
        end
    end
    return D, S
end
```

但是,当存在遮挡的时候,就是一个高置信度得分的人被另一个高置信度得分的人遮住,只是通过重叠率来判断会导致错误移除预测检测框。如果阈值设置得太高,会导致重复检测框从识别框列表中被移除得太少。如果阈值设置得太低,一些高置信度的正确检测框也会被从列表中移除。所以,应该改进 NMS 算法来让其更加高效。

2 用于人物识别的 YOLOv3 改进方法

通过使用 Soft-NMS 和取回算法来改进 YOLOv3,

这样会恢复一些错误移除掉的检测框。改进的YOLOv3 算法如图 1 所示。

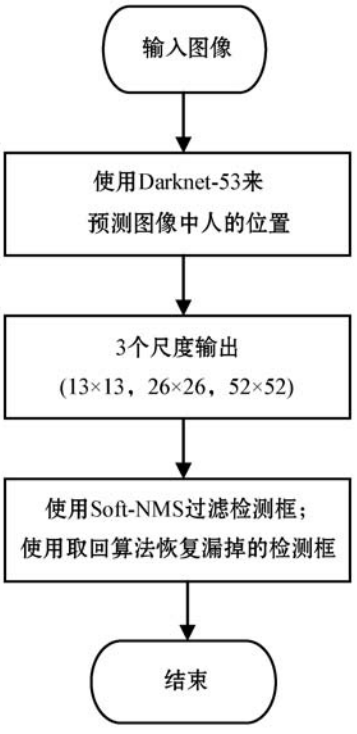

图 1　改进的 YOLOv3 算法

传统的 NMS 只是移除所有的重叠率高于阈值的检测框。这可能会导致很多正确的检测框被移除。Soft-NMS 对判断阈值进行了修改,尽管 Soft-NMS 提高了准确率,依然有一些正确的检测框被移除。我们添加了取回算法来发现未检测的人实例,并恢复人实例对应的丢失检测框,从而进一步提升了准确率。

2.1　使用 Soft-NMS 算法过滤错误预测的检测框

Boldla 等提出了 Soft-NMS 算法,该算法很像传统的 NMS,但是 Soft-NMS 没有立刻移除高重叠率的检测框。算法降低了重叠率大于等于阈值的检测框的置信度。

传统的 NMS 移除步骤可以描述如下:

$$\begin{cases} S_i = S_i & iou(M,b_i) < N_t \\ S_i = 0 & iou(M,b_i) \geqslant N_t \end{cases} \tag{1}$$

式中:$iou(M,b_i)$是最大置信度的检测框和剩余检测框之间的重叠率。式(1)通过比较 iou 和 N_t阈值来给检测框 i 的置信度 S_i重新赋值。

当检测框的重叠率大于等于阈值时,Soft-NMS 降低了检测框的置信度。因为依照 YOLOv3 的深度卷积神经网络的原理,检测框的重叠率越高,检测框越有可能是一个重复的错误的检测框。当一些检测框的重叠率高于阈值时,就需要被移除。但是当置信度很高时,这意味着它们更加可能是正确的检测框,应该被保留。所以 Soft-NMS 保留了重叠率高于阈值的检测框,但是不至于高到几乎完全和检测框 M 重叠的程度,而且这些保留的检测框的初始置信度也很高,这样在被算法降低置信度后依然可以保留。检测框几乎完全和检测框 M 重叠的会被移除,因为它们更可能是重复的错误的检测框。Soft-NMS 的移除标准定义如下:

$$\begin{cases} S_i = S_i & iou(M,b_i) < N_t \\ S_i = S_i(1 - iou(M,b_i)) & iou(M,b_i) \geqslant N_t \end{cases} \tag{2}$$

式(2)是线性函数,用来降低检测框的置信度。远离检测框 M 的检测框就会较少被影响,或者不会被影响。如果某检测框离检测框 M 非常近或者大部分被检测框 M 覆盖时,其置信度就会被降低非常多。最后,在所有的检测框置信度被降低后,还使用了另一个阈值来移除错误预测的检测框。降低这些检测框的置信度不会移除重复检测框,所以在降低置信度后,还需要设置一个用来过滤低置信度的检测框的阈值。

相比于传统的 NMS,Soft-NMS 对于 YOLOv3 没有增加更多的计算。Soft-NMS 的计算复杂度是 $O(N^2)$,与传统的 NMS 一样。N 是检测框的数量。每个检测框需要计算它和最大置信度检测框的重复率,所以 Soft-NMS 的计算复杂度是 $O(N^2)$。

Soft-NMS 对于 YOLOv3 来说是一个很小的部分。它不会需要对 YOLOv3 进行重新训练,所以集成到 YOLOv3 时不会花费太多的时间。

2.2　使用取回算法恢复被遗漏的检测框

因为 Soft-NMS 也是通过重叠率来判断是否移除检测框,所以肯定还存在错误移除的被算法漏掉的检测框。对此,我们可以通过取回算法来取回这些错误移除的检测框。

在取回算法中,我们提取了数据集中的人脸图像的 HOG 特征,并使用这些特征来训练了一个 SVM。使用 NMS 和一个滑动窗口来从图像中截取数据并提取对比特征,我们就可以检测人脸了。在改进的 YOLOv3 中,在检测出所有的人实例后,还需要检测出图像中的所有人脸。

因为人物检测框都围绕在人形状轮廓之外,所以对应人物检测框的检出人脸一定是完全在人物检测框里面。如果有人脸检测框在所有人物检测框外面,或者与人物检测框重叠时,肯定存在一个漏掉的人物检测框被 Soft-NMS 算法错误地移除。所以所有的被 Soft-NMS 移除的人物检测框都会被再次检查一遍,查找出那个完全覆盖了该人脸检测框的人识别框,因为

有时候会发现好几个人识别框符合要求,这时候就会找到置信度最高的那个来恢复。

在图 2 中,细线的检测框是人物检测框,粗线检测框是被 Soft-NMS 漏掉的检测框,已经被恢复了回来,虚线的检测框是检测到的人脸中没有被完全包含在人识别框中的人脸检测框。在图 2(a)中,人脸检测框明显和人物检测框边界重叠了,所以一定存在被 Soft-NMS 遗漏的人物检测框。

(a)

(b)

图 2 检测框结果

我们在这里会讨论如何判断一个人脸检测框是否完全在一个人物检测框里。定义(X_1,Y_1)是人脸检测框的右上角的坐标;(X_2,Y_2)是人脸检测框的左下角坐标;(M_1,N_1)是人物检测框的右上角坐标;(M_2,N_2)是人物检测框左下角坐标。如果这些点的坐标符合如下的条件,人脸检测框就属于完全在人物检测框里面的情况。

$$M_1 - X_1 > 0, Y_1 - N_1 > 0 \tag{3}$$

$$X_2 - M_2 > 0, N_2 - Y_2 > 0 \tag{4}$$

如果有一个人脸检测框在人物检测框外边或者与人物检测框重叠,我们就会搜索查找所有的原始的没有被 Soft-NMS 算法删减的检测框。计算它们哪一个完全覆盖了人脸检测框,最后找到最高的置信度的那个人物检测框。之后该人物检测框就会被取回恢复,所以人物检测准确度就会提高。

图 3 展示了取回算法的流程。

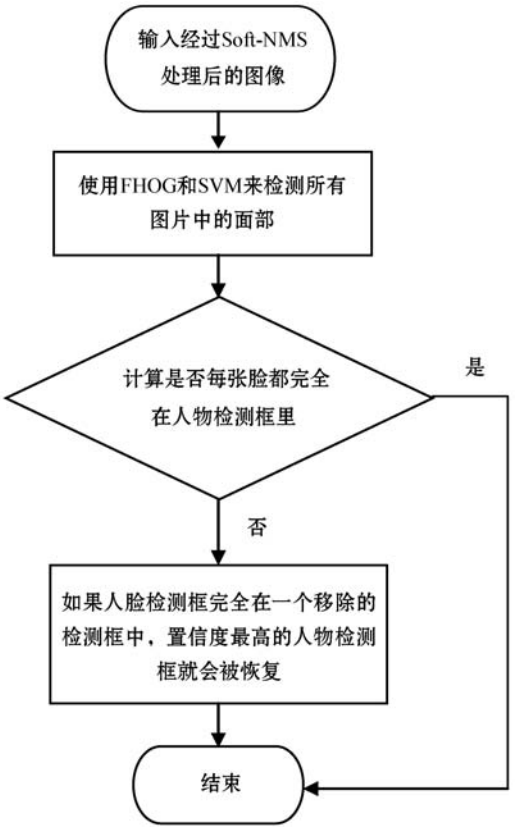

图 3 取回算法的流程

3 实 验

实验使用的数据集是 PASCAL VOC 2007[18]。YOLOv3 使用的权重是官网的作者训练好的权重,该权重训练使用的数据集是 MS-COCO。PASCAL VOC 数据集的测试部分被用于测试改进的 YOLOv3 的平均精确度。PASCAL VOC 测试部分包含了大约 5 000 幅图片。

实验中,我们设置了 NMS 重叠阈值为默认值 0.3,该默认值是作者发现的可以获得最高准确率的值。对于 Soft-NMS,除了重叠阈值 N_t 设置为 0.3,还有一个 Soft-NMS 作者设置的阈值 σ,通过对物体置信度进行与该阈值的比对,最终移除错误预测的检测框,该 σ 值设置为 0.4。该阈值设置得太高会移除掉所有的检测框,设置得太低也会降低检测准确度,因为检测框具有非常高的重叠率时,它就更可能是一个重复的检测框。设置一个低阈值意味着检测框很少会被移除。在对该阈值尝试了很多数值之后,数据结果如图 4 所示,当设置为 0.4 时,得到了最高的准确率。网络的输入分辨率设置为了 416。在对 PASCAL VOC 数据集进行检测结束后,我们计算了检测的准确率,使用了传统的 NMS 的 YOLOv3 和使用 Soft-NMS 和取回算法改进的 YOLOv3 的准确率如表 1 所示。

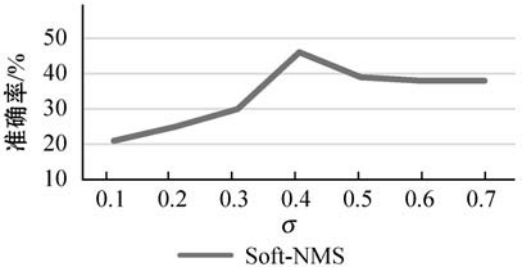

图 4　不同 σ 值下进行人物检测的准确率

表 1　使用 Soft-NMS 和取回算法改进的 YOLOv3 的准确率结果

方法	准确率/%
YOLOv3	43.00
使用 Soft-NMS 和取回算法的 YOLOv3	46.10

可以看出,在使用了 Soft-NMS 和取回算法后,准确率提升了 3.1 百分点。图 5 所示的部分实验结果证明了 Soft-NMS 带来的改进。图 5 中(a)、(c)、(e)、(g)是使用 YOLOv3 和 Soft-NMS 的检测结果,(b)、(d)、(f)、(h)是使用 YOLOv3 和 NMS 的检测结果。图 5(a)和图 5(b)中,上衣上印有 PM 的女孩被 NMS 给遗漏了。在图 5(c)和图 5(d)中,红色上衣的人被遗漏了。在图 5(e)和图 5(f)中,站在后面的女孩被遗漏了。在图 5(g)和图 5(h)中,左半部分中间的人被遗漏了。

由此我们可以看出,使用 NMS 时,一些人的检测框由于超过了阈值而被移除。通过使用 Soft-NMS,由于其使用了置信度判断更加合理,NMS 中错误移除的检测框可以被保留,提高了最终的检测精度。

(a)

(b)

(c)

(d)

(e)

(f)

(g)

(h)

图 5 使用 Soft-NMS 和 NMS 的 YOLOv3 的人物检测结果

图 6 中,虚线的检测框是被 Soft-NMS 漏掉的人物检测框对应的人脸检测框。他们被取回算法恢复了,检测框被标为了粗线。

(a)

(b)

(c)

图 6 使用 Soft-NMS 和取回算法的 YOLOv3 人物检测

我们比较不同模型消耗的时间,结果如表 2 所示。模型和它们的速度通过每秒处理帧数(Frames Per Second, FPS)来测量。从表 2 中看到,改进的 YOLOv3 相比其他模型有着最高的速度,同时精确度也被提升了。

表 2 人物检测模型和它们的检测速度

模型	训练	测试	FPS
SSD321	COCO trainval	Test-dev	16
DSSD321[19]	COCO trainval	Test-dev	12
R-FCN	COCO trainval	Test-dev	12
SSD513	COCO trainval	Test-dev	8
DSSD513	COCO trainval	Test-dev	6
FPN FRCN	COCO trainval	Test-dev	6
Retinanet-50-500	COCO trainval	Test-dev	14
improvedYOLOv3	COCO trainval	Test-dev	27

4 结 语

通过替换 NMS 为 Soft-NMS,加入了取回算法,本文对 YOLOv3 进行了改进。通过使用 Soft-NMS 算法,高置信度的检测框的置信度被降低了,而不是彻底从最终结果中移除掉,从而提升了准确度。Soft-NMS 的算法复杂度与传统的一样。取回算法恢复了 Soft-NMS 遗漏的检测框,进一步提升了准确度。

Soft-NMS 和取回算法可以被集成到其他未来的人物检测模型中去来提高性能。它们也可以集成到其他模型中用于人追踪。关于取回算法,还有更多可以改进的东西。据我们所知,检测的人脸检测框必须完全在人物检测框中。但是如果有两个人脸检测框完全在一个人物检测框中,一个人脸会被遗漏掉,其所属的人物检测框也不会恢复。为了解决这个问题,YOLOv3

的神经网络可以通过修改输出结果为包含人识别框信息和人脸识别框信息,并通过计算取回算法中的人脸检测框和修改的 YOLOv3 中输出的人脸识别框信息,并设置一个像 NMS 的阈值,就可以移除相应的人物检测框。然后遗漏的人脸就不会完全在人物检测框中,其所属的人物检测框就可以恢复,准确度就可以进一步提升。

参 考 文 献

[1] Ren S, He K, Girshick R B, et al. Faster R-CNN:Towards real-time object detection with region proposal networks[C]// Neural Information Processing Systems,2015:91 - 99.

[2] Redmon J, Farhadi A. YOLOv3: An incremental improvement[EB]. arXiv:1804.02767,2018.

[3] Liu W, Anguelov D, Erhan D, et al. SSD: Single shot multibounding box detector[M]//Computer Vision—ECCV 2016. Springer International Publishing,2016.

[4] Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence,2020,42(2):318 - 327.

[5] He K, Zhang X, Ren S, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[C]//European Conference on Computer Vision,2014:346 - 361.

[6] Dai J, Li Y, He K, et al. R-FCN: Object detection via region-based fully convolutional networks[C]//30th International Conference on Neural Information Processing Systems, 2016:379 - 387.

[7] Neubeck A, Gool L V. Efficient non-maximum suppression [C]//18th International Conference on Pattern Recognition, 2006:850 - 855.

[8] Bodla N, Singh B, Chellappa R, et al. Soft-NMS—Improving object detection with one line of code[C]//International Conference on Computer Vision,2017:5562 - 5570.

[9] Dalal N, Triggs B. Histograms of oriented gradients for human detection[C]//Computer Vision and Pattern Recognition,2005:886 - 893.

[10] Lowe D G. Distinctive image features from scale-invariant keypoints[J]. International Journal of Computer Vision, 2004, 60(2):91 - 110.

[11] Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition(CVPR), 2016:779 - 788.

[12] Redmon J, Farhadi A. YOLO9000: Better, faster, stronger [C]//Computer Vision and Pattern Recognition,2017:6517 - 6525.

[13] Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition(CVPR),2015:1 - 9.

[14] Lin T, Maire M, Belongie S J, et al. Microsoft COCO: Common objects in context[C]//European Conference on Computer Vision, 2014:740 - 755.

[15] Lin T Y, Dollar P, Girshick R, et al. Feature pyramid networks for object detection[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition(CVPR),2017.

[16] Rosenfeld A, Thurston M. Edge and curve detection for visual scene analysis[J]. IEEE Transactions on Computers, 1971,20(5):562 - 569.

[17] Viola P, Jones M. Rapid object detection using a boosted cascade of simple features[C]//2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2001.

[18] Everingham M, Gool L V, Williams C K, et al. The pascal visual object classes(voc) challenge[J]. International Journal of Computer Vision,2010,88(2):303 - 338.

[19] Fu C Y, Liu W, Ranga A, et al. DSSD: Deconvolutional single shot detector[EB]. arXiv:1701.06659,2017.

(上接第 108 页)

[4] 陈冠军,王英健,徐大远. 基于无源 RFID 的卸货列车行程定位方法研究[J]. 自动化技术与应用,2015,34(10):6 - 10,73.

[5] 林颖,王长林. 基于 CBTC 的车载 ATP 安全制动曲线计算模型研究[J]. 铁道学报,2011,33(8):69 - 72.

[6] 董海鹰,刘洋,李欣,等. 基于模糊神经网络预测控制的高速列车 ATP 研究[J]. 铁道学报,2013,35(8):58 - 62.

[7] 柏卓彤,柏赟,李佳杰,等. 基于制动距离表的高速铁路 ATP 常用制动曲线研究[J]. 铁道标准设计,2018,62(11): 139 - 143,149.

[8] 谭莉,王长林. CTCS3 级列控系统 ATP 防护曲线算法研究[J]. 铁路计算机应用,2014,23(7):48 - 52,57.

[9] 姜俊彤,李鸿,苏醒. 模糊神经网络在列车防冒进系统中的应用[J]. 自动化与仪表,2019,34(12):92 - 97.

[10] Yang H, Fu Y T, Zhang K P, et al. S-peed tracking control using an ANFIS model forhigh-speed electric multiple unit [J]. Control Engineering Practice,2014,23(1):57 - 65.

[11] Fu Y, Yang H, Wang D. Real-time optimal control of tracking running for high-speed electric multipleunit[J]. Information Sciences,2017,376(10):202 - 215.

[12] Chou M, Xia X. Optimal cruise control of heavy-haul trains equipped with electronically controlled pneumatic brake system[J]. Control Engineering Practice,2007,15(5):511 - 519.