

文章编号: 2095-2163(2021)08-0087-05

中图分类号: TP391.41

文献标志码: A

# 基于改进 YOLOv3 网络的非机动车检测

杨紫辉, 江 磊, 任洪娟

(上海工程技术大学 机械与汽车工程学院, 上海 201620)

**摘要:** 中国目前的交通主体主要由机动车、非机动车和行人构成。非机动车在日常的交通参与中占据了相当大的比例,同时随着无人车自动驾驶技术和机动车驾驶辅助技术的不断突破与升级,非机动车检测成为了路况信息获取中的重要一环。由于非机动车数量大,在视频中的所占面积小,极易互相遮挡等原因,容易出现检测不准确和漏检等问题。针对该问题,基于 YOLOv3 算法,本文提出了一种改进模型,通过采用双重特征提取网络,同时在辅助网络和骨干网络的特征信息融合采用注意机制提高检测精度。实验结果表明,所提出的改进模型在公开数据集中取得了优于 YOLOv3 的检测结果,将平均检测准确率由 79.33% 提高至 83.26%。

**关键词:** YOLOv3; 非机动车检测; 特征提取; 注意机制

## Non-motor vehicle detection based on improved YOLOv3 network

YANG Zihui, JIANG Lei, REN Hongjuan

(School of mechanical and automobile engineering, Shanghai university of engineering science, Shanghai 201620, China)

**[Abstract]** At present, the main body of traffic in china is mainly composed of motor vehicles, non-motor vehicles and pedestrians, and non-motor vehicles account for a considerable proportion in daily traffic participation. With the continuous breakthrough and upgrading of unmanned vehicle automatic driving technology and motor vehicle driving assistance technology, non-motor vehicle detection has become an important part of road information acquisition. Due to the large number of non-motor vehicles, small area in the video and easy to block each other, it is easy to detect inaccurate and missed detection. to solve this problem. Based on the yolov3 algorithm, an improved model is proposed in this paper. By using the dual feature extraction network, the feature information fusion of the auxiliary network and the backbone network uses the attention mechanism to improve the detection accuracy. The experimental results show that the improved model achieves better detection results than YOLOv3 in public datasets, and the average detection accuracy is improved from 79.33% to 83.26%.

**[Key words]** YOLOv3; non-motor vehicle detection; feature extraction; attention mechanism

## 0 引言

近年来,自动驾驶技术和机动车驾驶辅助技术不断突破与升级,智能驾驶系统发展势头迅猛。在此基础上,路况信息的获取、识别便显得尤为重要,逐渐成为了智能驾驶系统研究应用的基础与关键,而非机动车的精准检测识别便是智能驾驶系统的重要组成部分<sup>[1-2]</sup>。

传统的目标检测方法是使用人工设计的图像特征来进行检测,如 HOG 算子、SIFT 算子等,但由于真实场景复杂多样性,难以提取非机动车等目标特征,传统方法难以满足实际使用的需求。同时随着深度学习技术的不断发展,基于深度卷积神经网络的目标检测和跟踪算法被不断提出,如 R-CNN (Region Convolutional Neural Network)、Fast-RCNN 等,目标检测率得到了大大提升,基于深度卷积网络

的目标检测已经成为了主流检测方法。在 2015 年,Joseph Redmon 提出了 YOLO (You Only Look Once) 系列算法,极大地提升了算法检测的速度。

在驾驶场景中执行目标检测背景复杂,要测量的目标密集分布或重叠,摄像机的观看距离不固定,导致目标大小不同。特别是较小的目标,如非机动车、行人和交通标志,具有像素少、分辨率低、特征不明显的特点,YOLO 算法对该类目标的检测性能并不理想<sup>[3-4]</sup>。由此,本文基于改进优化的 YOLOv3 网络对非机动车进行检测,通过复制增加骨干网络得到特征提取辅助网络,由此提高整个特征提取网络的性能,并在辅助网络与骨干网络的特征信息融合时采用注意机制,重点对有效特征通道进行处理,抑制无效信息通道,提高网络的处理效率,经测试整个优化网络的性能和功能满足实际应用。

**作者简介:** 杨紫辉(1996-),男,硕士研究生,主要研究方向:机器视觉技术。

**收稿日期:** 2021-01-12

## 1 YOLOv3 网络结构分析

YOLO 系列是一种典型的网络结构为端到端的算法<sup>[5]</sup>,相较于 R-CNN 系列的两阶段网络算法, YOLO 算法的网络结构更加简洁。YOLO 算法的网络是首先生成候选推荐区域,然后执行检测与判断。其检测速度更加迅捷,原因在于其将候选区域机制和检测集成到了同一网络中。

### 1.1 训练过程

YOLOv3 的网络结构通过使用预定义的候选区域来取代 R-CNN 网络中的 RPN。其将特征映射分成  $s \times s$  网格,为了预测目标会在每个网格生成  $b$  个边界框,最后在特征地图上生成可以覆盖整个特征地图区域的预测边界框,其数量为  $s \times s \times b$ ,同时直接对生成的预测边界框进行边界回归。为了防止预测边界框是冗余的,需要对每个预测边界框进行置信度计算,然后为置信度设置阈值,阈值以上的预测边界框保留用于回归,低于阈值的边界框直接删除。其中,每个边界框的置信度由两部分组成:预测目标类别概率和预测边界框与实际帧的重合度。

置信度计算公式(1)为:

$$C = \Pr(class_i | object) * \Pr(object) * IOU_{pred}^{truth} \quad (1)$$

式中,  $\Pr(class_i | object)$  表示对象在预测边界框中分类的概率;  $\Pr(object)$  是一个参数,用于确定划分的网络中是否存在要检测对象的中心点,存在为 1,否则为 0;  $IOU_{pred}^{truth}$  表示预测框与实际框之间的重合面积与联合面积的比值。

为了减少工作量,可以通过对预测边界框设置阈值的方式消除多数无用边界框,但某些单体对象可能同时持有多个边界框来预测对象,从而在特征映射上产生冗余的预测边界框。因此,YOLOv3 使用非极性非最大抑制算法去除冗余检测框,从而得到一个对目标精确检测的目标框。

### 1.2 网络结构

由于网络的不断深化,梯度消失和梯度爆炸等问题会在训练过程中凸显出来,这类问题可以通过引入残差网络来解决。通常为了提取更深层次的特征信息,会使用将进入残差模块前的特征与残差模块输出的特征相结合的方法。YOLOv3 采用了新的网络结构 darknet-53。darknet-53 主要由 53 个卷积层组成,包含大量的  $3 \times 3$ ,  $1 \times 1$  卷积核。YOLOv3 与 v1 和 v2 的网络结构相比,其利用剩余网络设计了快捷连接模块,如图 1 所示。

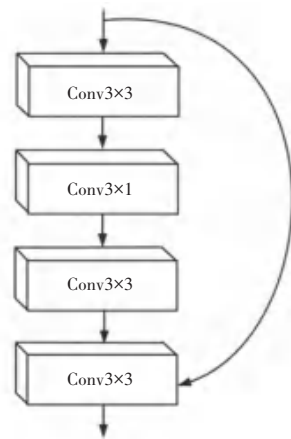


图 1 剩余结构网络图

Fig. 1 Residual structure network diagram

快捷连接模块的使用一方面有利于解决网络层过多造成的梯度消失问题,另一方面使整个网络的总层数达到 106 层,更适合于特征提取。同时, YOLOv3 采用多尺度检测机制,分别检测  $13 \times 13$ ,  $26 \times 26$  和  $52 \times 52$  的特征映射,增强了提取小目标的能力,其网络结构如图 2 所示。

YOLOv3 使用 3 个不同比列的特征映射来预测检测结果。当输入图像的分辨率确定时,基本尺度特征图的大小为原始分辨率的  $1/32$ ,其余 2 个尺度分别为  $1/16$  和  $1/8$ 。例如,当输入图像的分辨率为  $416 \times 416$  时,其基本尺度特征映射的大小为  $13 \times 13 \times n$ ,通过上采样获得  $26 \times 26 \times n$  的特征映射;将其与上一个卷积层的输出融合,从而获得  $26 \times 26 \times m$  的第二尺度特征图;基于第二尺度特征图,用同样的方法得到  $52 \times 52 \times w$  的第三尺度特征图;通过对每个尺度特征图的类别预测、边界框和目标分数来预测三维张量编码。在此过程中,检测帧共 4 个参数,目标评价 1 个参数,类别数为 80 个。每个尺度特征映射单元预测 3 组上述信息,即  $3 \times (4 + 1 + 80) = 255$  维信息。最后 3 个尺度的输出张量维分别为  $y_1 = 13 \times 13 \times 255$ ,  $y_2 = 26 \times 26 \times 255$ ,  $y_3 = 52 \times 52 \times 255$ 。

## 2 研究改进 YOLOv3

本文的设计旨在研究非机动车的检测问题,其检测内容为道路中行驶的非机动车,包括自行车、电动助力车等,需要测试的目标具有多样性特征。针对该情况,本文将 YOLOv3 的特征提取网络分为两部分进行优化。第一部分是通过增添骨干网络来获得特征提取辅助网络,提高整个特征提取网络的性能;第二部分是对骨干网络和辅助网络的特征信息

融合采用注意机制,提高对有效特征通道的关注度,抑制无效信息通道,提高网络处理效率。

### 2.1 特征提取辅助网络结构

YOLOv3 的特征提取网络采用 darknet-53,以残差结构形式加深采样深度。其中剩余模块内部结构

相对简单,使得整个网络简单易用,但特征提取能力无法优化。本文优化骨干网络的主要方式是通过增加复制剩余模块来拓宽整个网络,并对增加复制所得到的剩余模块结构进行微调,优化后的网络如图 3 所示。

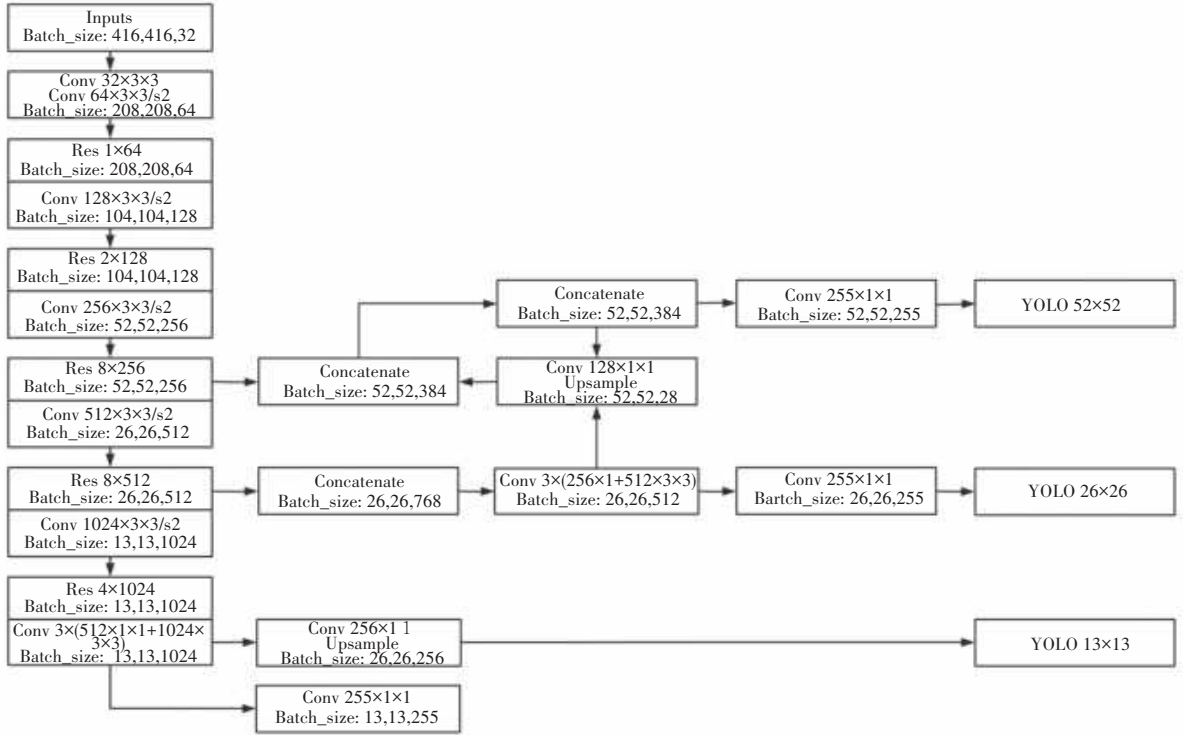


图 2 YOLOv3 网络结构图

Fig. 2 YOLO v3 network structure

特征合并到骨干网络中。

### 2.2 注意机制

二级网络与骨干网络的连接有两种不同的方式。第一种方式是辅助模块的输出首先由 1x1 卷积核集成,然后传输到骨干网络;第二种方式是将注意机制增添到深度辅助网络的两个网络之间。当网络达到一定深度时,其中的语义信息也变得更高,辅助模块将发挥对有效特征进行处理和传输,对无效特征进行信道抑制的作用。

由于 SE 模块具有结构简单,易于操作的优点,本文采用 SE 模块实现两个网络之间的注意机制。通过 SE 模块的增加,达到对辅助模块输出特性的重新校准。工作流程大致可分为挤压和激励。挤压是为了更好地现实每层中各信道的特征值分布,将二维特征映射通过平均池转化为一维特征映射,此时特征映射大小转换为 1x1xc;激励是使用一个全连接神经网络,对压缩之后的结果做一个非线性变换,该过程由两个全连接层组成,其中会加入一个缩放参数 SERadio 来减少通道个数,从而降低计算量。

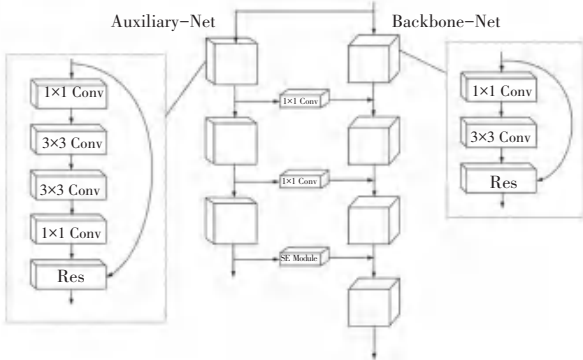


图 3 YOLOv3 优化网络结构图

Fig. 3 Optimized YOLO v3 network structure

与单一结构的原始网络相比,本文增加了一个比骨干网络规模更小的特征提取辅助网络,骨干网络的旁路由多个剩余模块组成。与 YOLO 残差模块相比,本文对辅助网络残差模块进行了改进,其具体实现方式为将原来的残差模块使用 3x3 卷积核进行特征提取,而辅助网络中的残差模块使用两个连续的 3x3 卷积核获得 5x5 的接受域,最后将提取的

### 3 实验结果与分析

#### 3.1 数据集训练

本文使用的数据集来自 BDD100K。2018 年 5 月加州大学伯克利分校 AI 实验室对外发布了 BDD100K, 该数据集是目前规模最大、内容最具多样性的公开驾驶数据集, 主要内容由 10 万段高清视频组成, 其中每段视频内容约 40 s, 分辨率为 720 p, 帧数为 30 fps。BDD100K 发布时, 该实验室同时设

计了一个图片标注系统, 该系统通过对数据集中每个视频的第 10 s 关键帧进行采样, 从而得到了 10 万张图片, 并对其进行标注。

本文在通过多次调整参数后, 对模型进行了 12 000 次的迭代训练。通过损失曲线的实时绘制, 可以直观地观察到训练过程中的实时动态, 如图 4 所示。图 4 中蓝色折线为本文模型训练工程中对应的平均损失曲线。其中, 横坐标表示训练迭代次数, 纵坐标表示训练过程中的损失值。

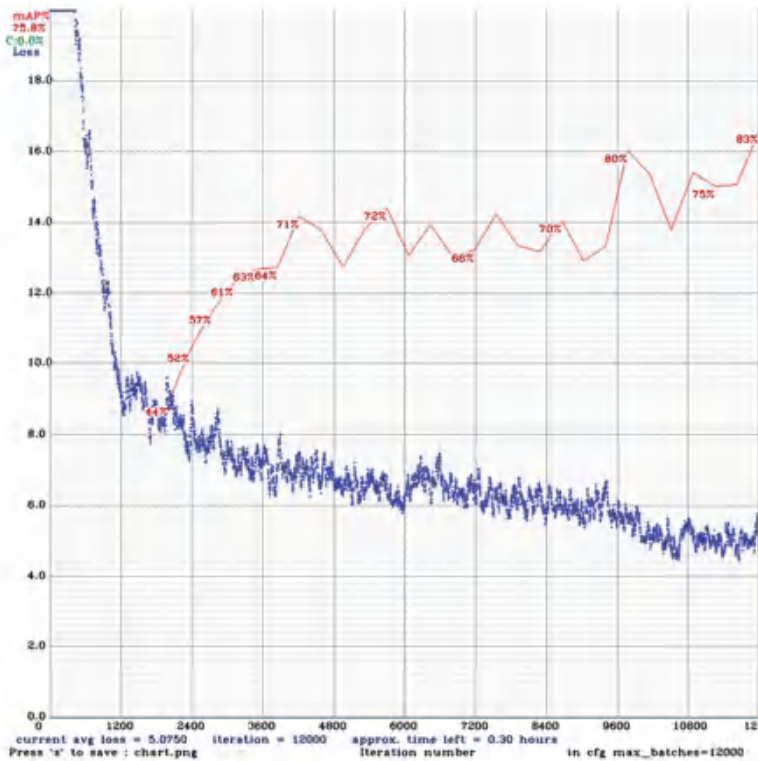


图 4 训练函数损失图

Fig. 4 Training loss function graph

#### 3.2 模型测试与比较

使用测试集对训练好的模型进行测试, 测试指标主要为对目标的召回率和检测的准确率。其中, 目标召回率  $R$  (Recall) 和检测准确率  $P$  (Precision) 计算公式如下:

$$R = \frac{X_{TP}}{X_{TP} + X_{FN}}, \quad (2)$$

$$P = \frac{X_{TP}}{X_{TP} + X_{FP}}. \quad (3)$$

其中,  $X_{TP}$  表示算法中正确分类的正样本, 即被正确检测出的目标数;  $X_{FN}$  表示错误分类的负样本, 即没有被检测出的目标数;  $X_{FP}$  表示错误分类的正样本, 即被错误检出的目标数。

$AP$  代表某一分类的精度, 通过找到在不同查全

率下最高的查准率得到。  $mAP$  代表多分类检测模型中所有类别的  $AP$  均值,  $mAP$  的值越大表示该模型的定位与识别的准确率越高。

测试所使用的 200 幅图像中共有 482 个目标, 使用 YOLOv3 算法和改进后的新 YOLOv3 算法在数据集上进行测试, 分别计算  $R$  和  $P$ , 结果见表 1。

表 1 优化模型与原网络模型试验结果对比表

Tab. 1 Comparison between optimized model and original network model test result

Detection Algorithm	$P$ /%	$R$ /%	$mAP$ /%
YOLOv3	79	82	79.33
Improved YOLOv3	83	85	83.26

本文模型与 YOLOv3 模型检测对比结果如图 5 所示。



(a) YOLOv3 检测效果  
(a) Detection effect of yolov3



(b) 本文模型检测效果  
(b) Detection effect of this paper

图 5 检测结果对比图

Fig. 5 Comparison of test results

## 4 结束语

本文主要介绍了一种基于 YOLOv3 网络优化的可用于非机动车检测的新 YOLO 网络模型。本文的主要工作:

(1) 新 YOLO 网络模型在原有 YOLOv3 模型的基础上, 采用了双重特征提取网络结构。采用规模为  $13 \times 13$ ,  $26 \times 26$ ,  $52 \times 52$  的骨干网络, 配置不同特征

提取的接收域辅助网络;

(2) 辅助网络和骨干网络的特征信息融合采用注意机制, 重点对有效特征通道进行处理, 抑制无效信息通道, 提高网络处理效率。

与 YOLOv3 相比, 优化后的网络检测能力得到了提升, 但由于辅助网络的增加, 计算量也会增加, 影响了检测时间, 但仍可满足实际应用的条件。为了进一步提升网络的检测应用能力, 可以使用边缘计算, 将模型用于嵌入式设备, 推动无人车智能驾驶技术发展。

## 参考文献

- [1] 叶佳林, 苏子毅, 马浩炎, 等. 改进 YOLOv3 的非机动车检测与识别方法[J]. 计算机工程与应用, 2021, 57(1): 194-199.
- [2] 鞠默然, 罗海波, 王仲博, 等. 改进的 YOLO V3 算法及其在小目标检测中的应用[J]. 光学学报, 2019, 39(7): 253-260.
- [3] 张祥越, 丁庆海, 罗海波, 等. 基于改进 LCM 的红外小目标检测算法[J]. 红外与激光工程, 2017, 46(7): 0726002.
- [4] TAIGMAN Y, YANG M, RANZATO M, et al. Deepface: closing the gap to human-level performance in face verification [C]// IEEE Conference on Computer Vision and Pattern Recognition, 2014: 1701-1708.
- [5] REDMON J, FARHADI A. YOLO9000: Better, Faster, Stronger [C]// IEEE Conference on Computer Vision and Pattern Recognition, 2017: 6517-6525.
- [6] 王紫, 李强, 张明, 等. 基于改进 YOLOv3 的行人检测算法[J]. 计算机工程与应用, 2020, 56(12): 2345-2352.
- [7] 李强, 王紫, 张明, 等. 基于改进 YOLOv3 的物体检测算法[J]. 计算机工程与应用, 2020, 56(12): 2353-2360.
- [8] 张明, 李强, 王紫, 等. 基于改进 YOLOv3 的车辆检测算法[J]. 计算机工程与应用, 2020, 56(12): 2361-2368.
- [9] 王紫, 李强, 张明, 等. 基于改进 YOLOv3 的行人检测算法[J]. 计算机工程与应用, 2020, 56(12): 2369-2376.
- [10] 李强, 王紫, 张明, 等. 基于改进 YOLOv3 的物体检测算法[J]. 计算机工程与应用, 2020, 56(12): 2377-2384.
- [11] 张明, 李强, 王紫, 等. 基于改进 YOLOv3 的车辆检测算法[J]. 计算机工程与应用, 2020, 56(12): 2385-2392.
- [12] 王紫, 李强, 张明, 等. 基于改进 YOLOv3 的行人检测算法[J]. 计算机工程与应用, 2020, 56(12): 2393-2400.
- [13] 李强, 王紫, 张明, 等. 基于改进 YOLOv3 的物体检测算法[J]. 计算机工程与应用, 2020, 56(12): 2401-2408.
- [14] 张明, 李强, 王紫, 等. 基于改进 YOLOv3 的车辆检测算法[J]. 计算机工程与应用, 2020, 56(12): 2409-2416.
- [15] 王紫, 李强, 张明, 等. 基于改进 YOLOv3 的行人检测算法[J]. 计算机工程与应用, 2020, 56(12): 2417-2424.
- [16] 李强, 王紫, 张明, 等. 基于改进 YOLOv3 的物体检测算法[J]. 计算机工程与应用, 2020, 56(12): 2425-2432.
- [17] 张明, 李强, 王紫, 等. 基于改进 YOLOv3 的车辆检测算法[J]. 计算机工程与应用, 2020, 56(12): 2433-2440.
- [18] WANG P, FAN Y, NIU S, et al. Hierarchical matching network for crime classification [C]// In Proceedings of SIGIR, 2019: 325-334.
- [19] XIAO C, ZHONG H, GUO Z, et al. CAIL2018: A large-scale legal dataset for judgment prediction [J/OL]. [2018-7-4]. <https://arxiv.org/abs/1807.02478v1>
- [20] ZHONG H, GUO Z, TU C, et al. Legal judgment prediction via topological learning [C]// In Proceedings of EMNLP, 2018: 3540-3549.
- [21] LI S, ZHANG H, YE L, et al. Prison Term Prediction on Criminal Case Description with Deep Learning [J]. CMC - Computers Materials & Continua, 2020, 62(3): 1217-1231.
- [22] CHEN H, CAI D, DAI W, et al. Charge-based prison term prediction with deep gating network [C]// In Proceedings of EMNLP-IJCNLP, 2019: 6363-6368.
- [23] SARAVANAN M, RAVINDRAN B, RAMAN S. Improving legal information retrieval using an ontological framework [J]. Artificial Intelligence & Law, 2009, 17(2): 101-124.
- [24] CASANOVAS P, CASANOVAS P, PALMIRANI M, et al. Semantic web for the legal domain: the next step [J]. Semantic Web, 2016, 7(3): 213-227.
- [25] WAGH R S, ANAND D. Application of citation network analysis for improved similarity index estimation of legal case documents: a study [C]// In Proceedings of ICCTAC, 2017: 1-5.
- [26] MINOCHA A, SINGH N, SRIVASTAVA A, et al. Finding Relevant Indian Judgments using Dispersion of Citation Network [C]// In Proceedings of the Web Conference, 2015: 1085-1088.

(上接第 86 页)