文章编号：2095-2163(2022)02-0115-05 中图分类号：U463.6 文献标志码：A

改进 YOLOv4 模型的车辆检测算法

韩 帅[1]，罗素云[1]，陈杨钟[2]
（1 上海工程技术大学 机械与汽车工程学院，上海 201620；2 大工科技(上海)有限公司，上海 200000）

摘 要： 随着深度学习的不断发展和应用，目标检测效果有了明显的改善。但由于目标检测任务需要检测多尺度信息，而目前的检测器在不同尺度物体检测方面仍有不足，尤其是对小目标物体，易出现漏检以及误检的情况。本文针对场景中出现的小目标漏检及误检问题，对小目标检测进行研究，对 YOLOv4 网络进行改进，在 YOLOv4 网络上增加专门针对小目标物体的特征层，实现语义信息和定位信息更好的融合。同时，增加数据集中小目标物体的占比，来提高小目标物体的检测精度。实验结果表明，所进行的网络改进达到了提高目标检测效果的目的。

关键词： 一阶段算法；车辆检测；小目标检测

Vehicle detection algorithm based on improved YOLOv4 model

HAN Shuai[1], LUO Suyun[1], CHEN Yangzhong[2]
(1 School of Mechanical and Automotive Engineering, Shanghai University of Engineering Science, Shanghai 201620, China;
2 Dagong Technology (Shanghai) Co., Ltd., Shanghai 200000, China)

【Abstract】 With the continuous development and application of deep learning, the effect of target detection has been significantly improved. However, since the target detection task needs to detect multi-scale information, the current detectors are still insufficient in the detection of objects of different scales, especially for small target objects that are prone to detection loss and false detection. Aiming at the small target detection loss problems in the scene, this paper conducts research on small target detection, improves the YOLOv4 network, and adds a feature layer specifically for small target objects on the YOLOv4 network to achieve better semantic information and positioning information fusion. At the same time, the proportion of small target objects is increased in the dataset to improve the detection accuracy of small target objects. The experimental results show that the proposed network improves the target detection effect.

【Key words】 one-stage algorithm; vehicle detection; small target detection

0 引 言

随着科技水平的不断发展，自动驾驶技术也不断成熟。通过车辆自身的传感器，对车辆、行人、信号灯以及可行驶区域的检测，是自动驾驶的重中之重。由于传统目标检测算法的实时性无法满足自动驾驶的需求，因此基于深度学习的目标检测算法，逐渐进入了人们的视野。

以 RCNN 为代表的二阶段检测算法，具有更高的检测精度。但需要经过候选区域筛选以及目标分类两大步骤，因此网络的实时性较差。以 YOLO 和 CenterNet 为代表的一阶段检测算法，将感兴趣区域的筛选和目标分类集成到一个网络，在保证精度的同时，提升了网络的实时性。与此同时，随着 MobileNet 和 ShuffleNet 等轻量化网络的提出，进一步降低了网络的参数量和计算量，使得网络可以在移动端进行部署。

1 YOLOv4 简介

1.1 YOLO 系列算法

YOLOv1[1] 和 YOLOv2[2] 算法是 YOLO 系列算法的开山之作，其检测思路不同于 RCNN 系列的二阶段算法[3]。二阶段算法先利用 RPN(区域生成网络)[4] 等方法选取目标位置所在的候选区域，然后在感兴趣区域中利用卷积神经网络的方法提取图像特征。以 YOLO 系列为代表的一阶段目标检测算法，其检测流程是一个单一的网络，创造性的把目标检测问题看成一个简单的回归问题。在一个网络中，直接回归出检测框的位置，并得到框内物体的种类以及置信度，可以实现端对端的实时监测。

YOLOv3[5] 和 YOLOv4[6] 算法的输出都是具有 3 种不同尺度、不同感受野的特征层，其主干网络均采用残差结构。一般情况下，卷积层数越多，网络就越深，得到的特征信息也就越丰富。但是，如 VGGNet

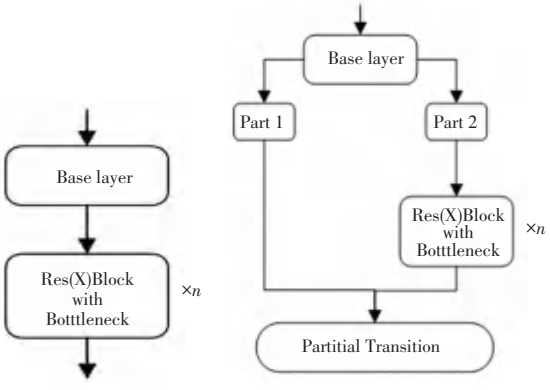
作者简介：韩 帅(1996-)，男，硕士研究生，主要研究方向：机器视觉、图像处理；罗素云(1975-)，女，硕士，副教授，主要研究方向：无人驾驶汽车环境感知及控制；陈杨钟(1983-)，男，硕士，工程师，主要研究方向：控制理论与控制工程。

收稿日期：2021-08-18

等网络[7],其加深到一定程度便无法继续加深。因为随着网络深度的增加,其检测效果不但不会得到优化,反而可能会变的更差。而残差网络[8]可以在网络不断加深,得到更强语义信息的同时,避免出现梯度爆炸和梯度消失等情况。

1.2 主干特征提取网络

YOLOv3 在 Darknet53 中,利用残差网络共进行了 5 次特征提取。当输入为 416 * 416 * 3 的特征图时,分别得到 208 * 208、104 * 104、52 * 52、26 * 26、13 * 13 5 层输出。随着图片不断被压缩,特征层深度不断增加,得到的语义信息更加丰富。同样,YOLOv4 的主干网络在其基础上改用了跨阶段局部网络[9](Cross Stage Partitial Network)。CSPNet 是一个逐层的特征融合机制,通过截断方式可以有效避免同一梯度信息被反复学习,从而得到最大化梯度组合的差异。残差网络基本结构如图 1 所示。



(a) ResNet 基本结构 (b) CSPResNet 基本结构

图 1 残差网络示意图

Fig. 1 Schematic diagram of residual network

1.3 特征融合框架

主干特征提取网络获取的特征层需要进一步特征融合,才能得到语义信息和位置信息都强大的特征聚合。YOLOv4 不仅包含与 YOLOv3 相同的自顶向下的 Feature Pyramid Networks 网络,还在此基础上对 13 * 13、26 * 26、52 * 52 的特征层利用 Path Aggregation Network 进行下采样,提高了浅层特征图的信息利用率。图 2 为 YOLOv4 网络中各部分代表的含义;YOLOv4 的整体框架如图 3 所示。

其中:CBL 模块由卷积层、归一化和 Leaky Relu 激活函数组成;CBM 模块由卷积层、归一化和 Mish 激活函数组成;Res_unit 由两个 CBM 模块经过残差连接而成;CSPx 中 x 代表包含几个 Res_unit;UP 代表上采样的操作;DOWN 代表下采样的操作;SPP 结构将 4 次不同尺度的最大池化进行通道的堆叠,然后再输入特征融合网络[10]。

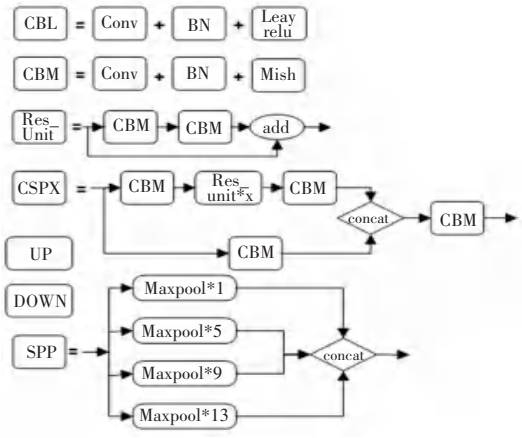


图 2 YOLOv4 框架组成模块

Fig. 2 YOLOv4 framework components

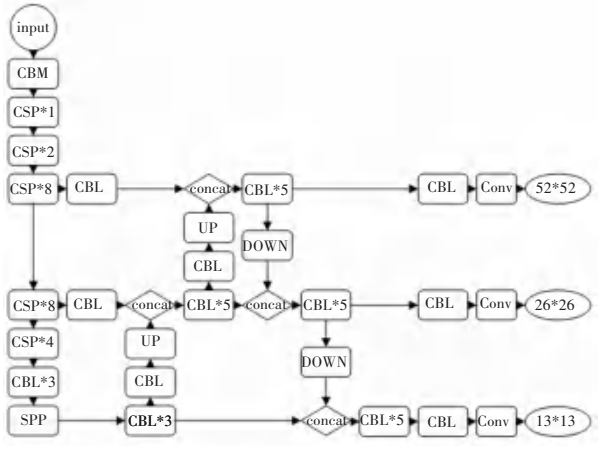


图 3 YOLOv4 整体框架

Fig. 3 YOLOv4 overall framework

1.4 损失函数及预测框

目前,多目标检测函数通常由两部分组成,即分类损失函数和回归损失函数。近年来,随着回归损失函数的发展,目标检测的精度和速度也有了一定提升。如:IOU_Loss 主要考虑检测框和目标框重叠面积;在 IOU_Loss 的基础上,GIOU_Loss[11]解决了边界框不重合时的问题;DIOU_Loss[12]还将考虑边界框中心点距离的信息;而 CIOU_Loss[13]则将重叠面积、边界框不重合、边界框中心距离和边界框宽高比的尺度信息进行了融合,得到对于目标检测最优损失函数。

$$CIOU = IOU - \frac{d^2}{c^2} - \alpha v \tag{1}$$

$$L_{CIOU} = 1 - IOU + \frac{d^2}{c^2} + \alpha v \tag{2}$$

其中:v 代表纵横比一致性;

$$v = \frac{4}{\pi^2}\left(\arctan\frac{w^{gt}}{h^{gt}} - \arctan\frac{w}{h}\right)^2 \quad (3)$$

α 代表折中系数；

$$\alpha = \frac{v}{1 - IOU + v} \quad (4)$$

d 代表预测框和真实框的中心点的欧氏距离；

$$d = \rho(b, bgt) \quad (5)$$

c 同时包含预测框和真实框的最小闭包区域对角线距离。

Yolov4 中采用 Kmeans 聚类的方式，分别得到 3 个特征层的 9 个不同尺度的先验框，并采用 CIOU_Loss 的回归方式，使得预测框回归的速度和精度得到提高。

2　改进的 YOLOv4 模块

2.1　小目标检测难点

小目标检测在许多任务中至关重要。如，从汽车高分辨率场景照片中，检测小的或远处的物体对于安全部署自动驾驶汽车是必要的。

在目标检测实际应用场景中，需要检测的目标的尺寸往往大小不一，即多尺度检测。多尺度检测要求训练的模型应具有较强的鲁棒性，可以检测出不同尺度的各类物体。然而在检测过程中，大尺度物体占据图像的面积大，经过多次卷积得到的特征也比较丰富，因此较容易被检测出来。所以，对于目标检测来说，其难点在于如何精确定位和识别出占据图像比例较小的目标。

小目标物体难以检测的原因可以分成两类：

(1)训练所用的数据集中，小目标物体出现的次数较少。主要原因有：

①a)数据集包含小目标物体的图片数量较少；

②即使图片中包含小目标物体，但其在一张图片中出现的次数较少，很容易被忽略；

③许多图片本身的分辨率较低、图像模糊，导致其携带的信息较少。

(2)特征提取网络无法很好的提取到小目标的特征。当输入进来的图片经过不断的卷积和采样，使得图片不断地被压缩，使小目标物体所占据的比例更小，且即使检测框可以检测到小物体所在的位置，也会在检测框内部包含大量不属于小目标的特征。此外，特征融合网络没有充分利用语义信息和位置信息。

不同阶段的特征图对应的感受野不同，表达的信息抽象程度也不一样。浅层特征图中含有更多的位置信息，深层特征图中含有更多的语义信息，如何将语义信息和位置信息进行更好的特征融合，得到特征更加丰富的特征层，也是一个亟待解决的难题。

2.2　数据集增强

数据增强就是在现有数据的情况下，让有限的数据通过变换，得到更多有价值的数据。传统的数据增强方式包括翻转、旋转、裁剪、变形、缩放等。本文在 Mosaic 数据增强方式的基础上，增加了对小目标物体的复制与粘贴，使得场景远处的小目标物体在数据集中占据更大的比例。

首先，在图片中选取一个小目标物体，在图片的任意位置进行多次粘贴。在复制粘贴过程中，要保证粘贴的位置不能与图像中现有的目标有遮挡，如图 4 所示。

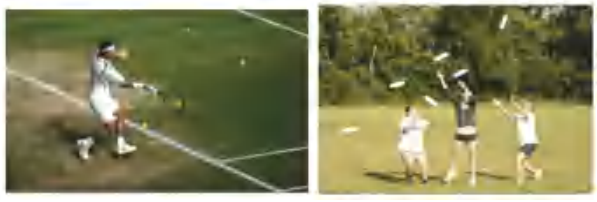


图 4　小目标的随机增强

Fig. 4　Random enhancement of small targets

粘贴完成后，利用 Mosaic 数据增强方法随机选取 4 张图片进行缩放，再随机分布进行拼接，大大丰富了检测数据集，特别是随机缩放增加了很多小目标，让网络的鲁棒性更好。Mosaic 数据增强实例如图 5 所示。

图 5　Mosaic 数据增强

Fig. 5　Mosaic data enhancement

2.3　优化的多尺度特征融合

传统的卷积神经网络，都是自上而下进行的，随着网络层数的加深，图像包含的语义信息也更加丰富。但与此同时，小目标的特征可能会随着网络层数的加深而逐渐被忽略。在 YOLOv3 当中，通过 FPN 中的上采样结构，将深层特征图的语义特征传递给浅层特征图，实现特征融合[14]。YOLOv4 在此基础上，增加了 PANet 结构[15]，将浅层特征图的强定位特征传入深层网络中，进行进一步的特征聚合，得到网络的 3 个用于检测的输出层。

本文主要针对解决场景中出现的小目标漏检及误检的问题。YOLOv4 利用 FPN 和 PANet 网络将主干特征提取网络中 13 * 13,26 * 26,52 * 52 的 3 层输出进行融合,得到语义信息和定位信息更加丰富的 3 个特征层,然后进行先验框的预测和回归。

浅层的特征图感受野小,比较适合检测小目标。因此,本文算法在传统的 YOLOv4 的基础上,将主干特征提取网络的第二层 104 * 104 的输出经过上采样和下采样,再与前 3 层输出融合在一起,得到 4 个特征层,来提高网络对于小目标物体的检测能力(如图 6 中红色虚线部分)。并且,利用 Kmeans 聚类方法[16],得到适合每个数据集的先验框的宽和高。根据聚类得到的 4 组先验框的大小,将其划分到对应的特征层上,大的先验框分配到深层上,用于检测大物体;小的分配到浅层上,用于检测小物体。

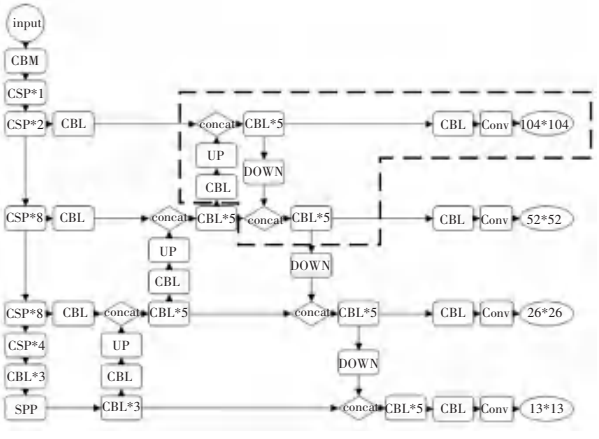


图 6 改进的 YOLOv4 结构

Fig. 6 Improved YOLOv4 structure

3 实验效果

3.1 数据集

本文采用了在图像分类、目标检测和图像分割比赛中运用最为广泛的 VOC 数据集对网络进行训练。VOC 数据集共包含人、车辆、动物、室内家具、背景等 5 大类,总计 21 个小类。

首先,选取 vehicle 中的 4 个公路交通工具进行检测,分别是 bicycle、bus、car、motorbike。VOC_2007 数据集和 VOC_2012 数据集分别包含 9 963 张和 17 125张尺度丰富的图像,再从中随机挑选出 16 551张图片组成 VOC_2007+2012 数据集。这 3 个数据集中,训练集、验证集、测试集的比例分别为 8∶1∶1。

首先,利用 VOC_2007、VOC_2012、VOC_2007+2012 数据集分别对 vehicle 中的 4 类进行训练。然后,再对经过数据增强后的 3 个数据集进行同样的训练,分别得到结果,进行对比。最后,将改进的 YOLOv4 与其它检测网络进行对比。

3.2 实验效果

为了验证本文所改进的网络在目标检测当中的有效性,在服务器上进行了实验。

图 7 中,图 7(a)是利 YOLOv4 网络在 VOC_07+12 数据集上经过 100 轮训练后得到的 val_loss 最低的权重进行预测后,得到的效果图。图 7(b)为同等条件下,改进的 YOLOv4 网络的预测结果。从两图对比可以看出,改进的 yolov4 网络比原始的网络可以检测出更多场景远处的小目标车辆,提升了小目标的检测精度,从而使得整体检测精度有了提升。

(a) YOLOv4 网络的预测结果

(b) 改进 YOLOv4 网络的预测结果

图 7 检测效果对比图

Fig. 7 Comparison of detection results

表 1、表 2 分别为各类检测器在 VOC_2007+2012 数据集和增强 VOC_2007+2012 数据集下的检测效果。其中包括 YOLOv3 和 YOLOv4,以及将 YOLOv4 主干特征提取网络 CSPDarkNet53 替换为 MobileNetv1、MobileNetv2、MobileNetv3 的轻量化网络 YOLOv4_M1、YOLOv4_M2、YOLOv4_M3。将以上 5 种改进的目标检测网络与本文提出的改进 YOLOv4 网络作比较,得到的实验结果。

表 1 VOC_2007+2012 数据集结果

Tab. 1 Results on VOC_2007+2012 dataset

VOC_07+12	AP				mAP
	bicycle	bus	car	motorbike	
YOLOv3	78.07	77.65	80.22	81.64	79.39
YOLOv4	87.26	88.82	85.69	84.55	86.58
YOLOv4_M1	79.02	81.28	83.24	81.39	81.23
YOLOv4_M2	78.56	82.20	82.39	80.34	80.87
YOLOv4_M3	82.33	86.07	85.18	82.36	83.99
改进 YOLOv4	**88.35**	**90.01**	**87.23**	**85.07**	**87.67**

表 2　增强 VOC_2007+2012 数据集下实验结果

Tab. 2　Results under the enhanced VOC_2007+2012 dataset

VOC_07+12	AP				mAP
	bicycle	bus	car	motorbike	
YOLOv3	79.32	80.36	82.59	80.03	80.58
YOLOv4	88.21	89.89	88.36	85.94	88.10
YOLOv4_M1	81.37	88.53	86.56	84.02	85.12
YOLOv4_M2	84.12	86.17	88.40	83.30	84.50
YOLOv4_M3	84.96	91.41	85.33	**86.11**	86.95
改进 YOLOv4	**88.65**	**92.06**	**90.27**	85.66	**89.16**

4　结束语

本文在 VOC 数据集基础上，筛选出属于 Vehicle 的 4 类图片，并利用 Mosaic 和增加小目标的方式丰富了数据集中的 Vehicle。同时，在 YOLOv4 算法的基础上，增加了特征输出，使得网络可以在同等条件下检测更多场景远处的小目标物体，提升了整体的检测精度。最后，利用改进的 YOLOv4 在增强的 VOC 数据集下进行训练，对 4 类交通工具的检测精度均有不同程度的提升。

参考文献

[1] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 779-788.

[2] REDMON J, FARHADI A. YOLO9000: better, faster, stronger [C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 7263-7271.

[3] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2014: 580-587.

[4] MA J, SHAO W, YE H, et al. Arbitrary-oriented scene text detection via rotation proposals [J]. IEEE Transactions on Multimedia, 2018, 20(11): 3111-3122.

[5] REDMON J, FARHADI A. Yolov3: An incremental improvement [J]. arXiv preprint arXiv:1804.02767, 2018.

[6] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. Yolov4: Optimal speed and accuracy of object detection[J]. arXiv preprint arXiv:2004.10934, 2020.

[7] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [J]. arXiv preprint arXiv:1409.1556, 2014.

[8] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.

[9] WANG C Y, LIAO H Y M, WU Y H, et al. CSPNet: A new backbone that can enhance learning capability of CNN [C]// Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops. 2020: 390-391.

[10] HE K, ZHANG X, REN S, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE transactions on pattern analysis and machine intelligence, 2015, 37(9): 1904-1916.

[11] REZATOFIGHI H, TSOI N, GWAK J Y, et al. Generalized intersection over union: A metric and a loss for bounding box regression [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 658-666.

[12] ZHENG Z, WANG P, LIU W, et al. Distance-IoU loss: Faster and better learning for bounding box regression[C]//Proceedings of the IEEE Conference on Artificial Intelligence. 2020: 12993-13000.

[13] ZHENG Z, WANG P, REN D, et al. Enhancing geometric factors in model learning and inference for object detection and instance segmentation[J]. arXiv preprint arXiv:2005.03572, 2020.

[14] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 2117-2125.

[15] LIU S, QI L, QIN H, et al. Path aggregation network for instance segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 8759-8768.

[16] COATES A, NG A Y. Learning feature representations with k-means [M]//Neural networks: Tricks of the trade. Springer, Berlin, Heidelberg, 2012: 561-580.

(上接第 114 页)

参考文献

[1] 张明心. 基于认知诊断的贝叶斯知识追踪模型改进与应用[D]. 上海：华东师范大学硕士学位论文，2019.

[2] 吴虑. 大数据支持下学习评价的价值逻辑[J]. 清华大学教育研究，2019 (1)：15-18.

[3] SANTORO A, BARTUNOV S, BOTVINICK M, et al. Meta-learning with memory - augmented neural networks [C]// Proceedings of The 33rd International Conference on Machine Learning, 2016:1842-1850.

[4] 李梦蕾，李爽，沈欣忆. 2007—2017 年我国学习分析研究进展与现状分析——基于国内核心学术期刊文献的分析[J]. 中国远程教育，2018 (10)：78-79.

[5] 李景奇，卞艺杰，方征. 基于 BKT 模型的网络教学跟踪评价研究[J]. 现代远程教育研究，2018 (5)：104-112.

[6] 王怀波，李冀红，杨现民. 目标导向的学习分析模型构建[J]. 中国电化教育，2018 (5)：96-98.

[7] 朱静宜. “互联网”背景下高职物联网专业实践教学云平台的建设探索[J]. 物联网技术，2017，7(12)：106-109.

[8] 刘邦奇，李鑫. 智慧课堂数据挖掘分析与应用实证研究[J]. 电化教育研究，2018 (6)：41-47.

[9] 郭佳盛. 基于深度学习的自适应学习的学生模型研究[D]. 中国优秀硕士学位论文全文数据库，2017.

[10] ZHANG J, SHI X , KING I, et al. Dynamic Key-Value Memory Networks for Knowledge Tracing[C]// International Conference on World Wide Web. International World Wide Web Conferences Steering Committee, 2017:765-774.

[11] VINYALS O, BLUNDELL C, LILLICRAP T, et al. Matching networks for one shot learning[J]. Advances in neural information processing systems, 2016, 29: 3630-3638.

[12] 张荣花. 基于 ASSURE 模式的课程教学设计——以移动学习为背景[J]. 文教资料，2019 (4)：34-37.

[13] 杨诚. 基于学习风格和概率推理的智能教学系统研究[D]. 中国优秀硕士学位论文全文数据库，2018.