

基于改进 YOLOv2 的山西传统民居门窗识别方法研究

Research on the Recognition Method of Shanxi Traditional Dwelling Door and Window based on Improved YOLOv2 Model

文 / 杨旭晖 余亮 唐铭婕 付蒙
Yang Xuhui Yu Liang Tang Mingjie
Fu Meng

作者简介

杨旭晖 苏州大学金螳螂建筑学院 硕士
余亮 苏州大学金螳螂建筑学院 教授 (通讯作者)
唐铭婕 苏州大学金螳螂建筑学院 硕士
付蒙 苏州大学金螳螂建筑学院 硕士

基金项目

国家自然科学基金项目 (项目编号: 41371173)

ABSTRACT

针对现有传统民居识别研究缺乏的现状,文章基于改进当下主流识别网络 YOLOv2 提出了一种山西传统民居门窗识别方法,通过增设卷积层、BN 层等手段设计出三种改进网络结构,通过对同一数据集进行训练并对比识别效果,进而选出适用于门窗识别的最优模型——YOLOv2-TDDWNet。该模型对目标的识别精度可达 91.2%,不同类型门窗的 mAP 达到了 75%,具有较高的识别准确度、较小的误检率和良好的鲁棒性。

In view of the lack of research on the identification of traditional dwellings, this paper proposes a method for the identification of traditional dwelling doors and windows in Shanxi Province based on the improvement of the current mainstream identification network YOLOv2. Three improved network structures are designed by adding convolution layer and BN layer. By training the same sample set and comparing the recognition effect, the optimal model——YOLOv2-TDDWNet, which is most suitable for door and window recognition. The target recognition accuracy of this model can reach 91.2%, and the mAP of different types of doors and windows can reach 75%, with high recognition accuracy, small error detection rate and good robustness.

KEY WORDS

传统民居门窗; 识别; YOLOv2; 山西
traditional dwelling door & window; recognition; YOLOv2; Shanxi Province

DOI

10.19875/j.cnki.jzywh.2020.10.016

引言

将深度学习技术应用于目标检测领域是近年来突飞猛进的目标检测方法^[1],在众多领域均得到了长足的发展。本文提出将图像识别技术用于传统民居门窗信息提取的数字遗产保护方法,以期全方位提升对历史建筑的保护力度。门窗属于传统民居各部分组成构件中图像特征相对明显、简单且数量最多的一部分单元构件,因此选取门窗这一传统民居构件进行传统民居识别初探。

1 卷积神经网络及其改进研究现状

Girshick 等人^[2]基于卷积神经网络提出了 R-CNN (Region-Convolutional Neural Network) 目标检测模型。这种模型基于传统机器学习与深度学习原理,借助卷积神经网络提取目标特征,继而再使用 SVM (Support Vector Machine) 进行分类,最后得到的 mAP (mean Average Precision) 值相对于传统识别方法有了很大的提升。后续人们又先后提出了新的目标检测算法: Fast R-CNN、Faster R-CNN、YOLO (You Only Look Once) 和 SSD (Single Shot MultiBox Detector) 等^[3]。在这些新兴检测算法中,YOLO 以网络结构简单、检测速度快等优势深受推崇。

关于对 YOLOv2 神经网络的改进,李珣等提出改变激活函数与初始学习率;白中浩等^[4]提出使用交叉熵函数替代 sigmoid 激活函数与采用自组织数据分析算法进行聚类;宋建国等^[5]提出将

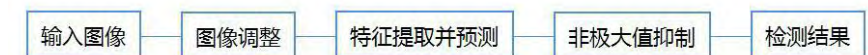


图 1 YOLO 目标检测原理 (图片来源:作者自绘)

K-means 算法改进为 DA-DBSCAN 算法等。以上改进措施大多是针对 YOLOv2 算法进行的优化改进,对网络结构本身的调整不大,且识别对象大多为车辆、墙体裂缝、植物茎叶^[6]等,未涉及到传统民居门窗范畴。

综上,针对现有传统民居识别研究缺乏的现状,本文提出了一种基于改进 YOLOv2 网络模型的山西传统民居典型门窗图像识别算法,通过使用不同改进模型对门窗样本集进行训练后识别效果的对比,获得了适用于传统民居门窗识别的改进模型 YOLOv2-TDDWNet (Traditional Dwelling Door & Window Net)。

2 基于改进 YOLOv2 民居门窗目标检测算法设计

2.1 基于 YOLOv2 的目标检测算法

YOLO (You Only Look Once) 是 Redmon J 等^[7]于 2016 年提出的一种端到端实时目标检测网络,其基本原理如图 1 所示。

当图像被输入 YOLOv2 之后,其分辨率会被调整为 448×448dpi,接下来网络将图像划分为 S×S 个网格,这 S×S 个网格就是 YOLOv2

来学习预测目标的基础。如图 2 所示,输入的目标猫的中心位于整幅图片中上方的一个红色网格中,那么这个网络就担负起预测猫位置的任务。而每个黑色网格的任务是预测边界框 (Bounding Box, BBox)。若目标检测任务需要检测多个类别,则每个网格负责多个条件概率 $Pr(Class_i | Object)$ 。每个 BBox 由四个坐标点信息 (X, Y, W, H) 和一个置信度 c (ConfidenceScore) 组成,其中, (X, Y) 是目标中心相对于当前网格位置的偏移量, (W, H) 为 BBox 相对于整个图像的宽和高,最后这四个值将被归一化到区间 $[0, 1]$ 之间。置信度 c 包括两层信息,一是目标 $Pr(Object)$ 是否出现在当前网格;二是该网格预测目标的准确性如何,置信度 c 可以通过公式 1 计算^[8]。

$$c = Pr(Object) * IOU_{pred}^{truth} \quad (1)$$

公式 1 中, IOU_{pred}^{truth} 表示预测物体位置的准确性,简称交并比。如图 3 所示, truth 表示实际框, pred 表示预测框,二者的交集与它们的并集之比也就是预测框与实际框两个矩形框的交集与并集之比即为交并比^[9]。直观理解为,若预

测框与实际框重合,即代表网格中包含目标,则 IOU_{pred}^{truth} 为 1, 否则 IOU_{pred}^{truth} 为 0。 IOU 公式如式 2 所示。

$$IOU_{pred}^{truth} = \frac{pred \cap truth}{pred \cup truth} \quad (2)$$

每个 BBox 的最终得分 G 由公式 3 得到。
 $G = Pr(Class_i | Object) * Pr(Object) * IOU_{pred}^{truth} \quad (3)$

YOLOv2 共预测 $S \times S \times B$ 个 BBox, 根据识别要求设置一个阈值 $thresh$, 网络会自动把低于阈值的 BBox 剔除, 对剩余 BBox 做非极大值抑制 (Non-maximum suppression, NMS) 处理, 这样我们就得到了最后想要的检测结果。在多方面领先于其他卷积神经网络, YOLOv2 是当前图像识别深度学习的首选工具, 其适用范围也可完美解决不同种类门窗检测与识别的功用。

2.2 基于 YOLOv2 目标检测识别模型的改进
针对现有传统民居识别研究空缺的现状,

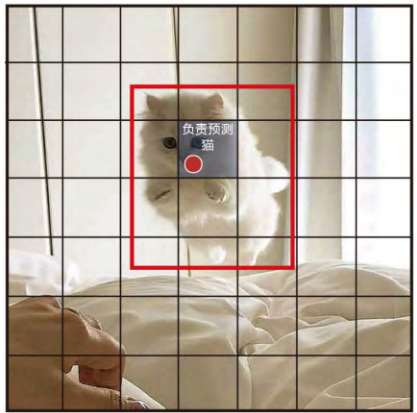


图 2 目标猫预测 (图片来源: 作者自绘)

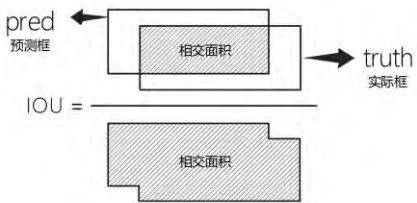


图 3 IOU 交并比示意图 (图片来源: 作者自绘)

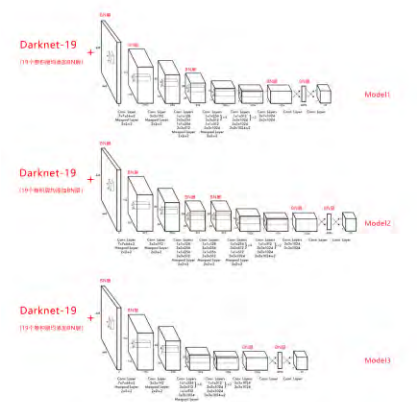


图 4 Model 1-3 神经网络示意图 (图片来源: 作者自绘)

模型编号	网络结构	
	卷积层数	BN 层数
Model 1	24	24
Model 2	25	25
Model 3	23	23

表 1 三种模型中的卷积层、BN 层数目 (表格来源: 作者自绘)



图 5 数据集截图 (图片来源: 作者自绘)

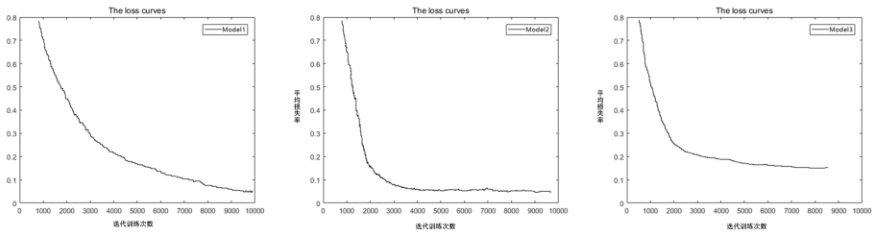


图 6 Model 1-3 训练过程平均损失变化曲线图 (图片来源: 作者自绘)

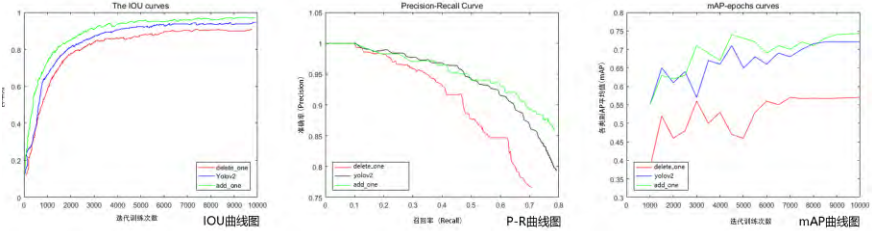


图 7 三种模型 IOU 曲线图 (图片来源: 作者自绘)

本论文提出了一种基于改进 YOLOv2 网络模型的山西传统民居典型门窗图像识别算法, 将其命名为 YOLOv2-TDDWNet (Traditional Dwelling Door & Window Net), 改进措施主要包括四个方面。

2.2.1 添加批量标准化层

YOLOv2 之所以是当前性能较优的一种识别网络, 主要归功于它的特征提取手段采取的是卷积神经网络, YOLOv2 的神经网络结构由 24 个卷积层加上 2 个全连接层所组成。通过前文对于卷积神经网络结构总结出, YOLOv2 中全新引入的批量标准化 BN 层能够提高识别精度, 而原有 Darknet-19 网络及 YOLOv2 网络中仅有 19 个卷积层中添加了 BN 层, 本文对其进行增加改进设计, 对 24 个卷积层均添加 BN 层。将这一改进网络结构模型命名为 Model 1。图 4 是 Model 1 的网络结构示意图。

2.2.2 添加卷积层

卷积层是卷积神经网络对样本进行特征提取的直接手段, 因此我们有理由推测, 卷积层数目的增加将会提高识别精度。

我们在 BN 层已经进行优化的 Model 1 的基础上, 在 YOLOv2 网络中添加一个卷积

层, 命名为 Model 2 (图 4)。识别层数的增加有导致训练速度降低的可能性, 因此我们在 Model 1 的基础之上, 在 YOLOv2 网络中删减一个卷积层作为 Model 2 的对比网络, 将其命名为 Model 3 (图 4)。三种模型网络具体特点如表 1 所示。

设计完成模型网络结构以后, 接下来就开始用三种模型对同一样本集进行训练, 记录训练过程中的各项参数, 对比三种网络识别精度的各项参数, 以求得最优网络模型。

3 实验结果与分析

3.1 建立数据集

本文利用作者实地调研拍摄的山西传统民居典型门窗图像数据, 共 5216 张 (图 5), 图像分辨率为 1600×1200 dpi, 其中 3830 张图像作为训练集, 1386 张图像作为验证集, 图片格式为 jpg。样本含有六种山西传统民居典型门窗样式, 即板门、隔扇门、直根窗、固定窗、槛窗、门联窗。样本目标包括了日常多种天气、光照及门窗本身完整度等。同时, 为了使样本数据分布均匀, 样本中的六种传统民居门窗类型图像各 600 多张。

模型	Recall	Accuracy	TP	FN
Model 1	0.89	0.74	2697	333
Model 2	0.91	0.76	2831	279
Model 3	0.72	0.64	2369	921

表2 三种模型召回率与准确率统计表 (表格来源: 作者自绘)

	模型	mAP	板门	隔扇门	直棂窗	固定窗	槛窗	门联窗
1	Model 1	0.73	0.75	0.74	0.7	0.72	0.69	0.78
2	Model 2	0.75	0.77	0.75	0.73	0.74	0.72	0.79
3	Model 3	0.65	0.66	0.64	0.65	0.64	0.6	0.71

表3 三模型六类传统民居典型门窗检测 mAP (Average Precision) 值 (表格来源: 作者自绘)



图8 检测效果图 (图片来源: 软件导出)

3.2 监控训练过程及绘制参数曲线

设置好训练参数后即开始训练, 在训练过程中会实时显示 IOU、Loss、Recall 等训练信息。统计各类信息如表 2 所示。

图 6 为三种模型以 0.001 的初始学习率在迭代近 10000 次时的 Loss 曲线图。从 Loss 曲线图上可以看出, 在收敛速度上, Model 1 在迭代 9000 多次之后平均损失收敛趋于 0, 相比 Model 2 在迭代 3000 多次之后便开始收敛趋于 0 较慢, 而 Model 3 收敛效果较差, 将导致模型识别准确性较低。

图 7 是三种模型以 0.001 的初始学习率在迭代近 10000 次时的 Loss 曲线图、PR 曲线图和 mAP 曲线图。参照 Rosebrock 提出的 IOU 评价参考标准^[10], IOU 超过 50.00% 即为良好、超过 73.30% 为好、超过 92.64% 为优秀。结合图 7 中 IOU 曲线图可以看出, 其中门窗的平均 IOU 都超过了 79%, 可知门窗检测效果属于“好”。在经过 10200 训练批次之后, Model 2 的平均损失 Loss 基本稳定在 0.05 上下, 且 IOU 基本稳定在 0.95 左右, 至此训练基本已经达到了一个较好的效果。综上可知, 对于总体的门窗识别模型, Model 2 最有优势, 本文将将其命名为 YOLOv2-TDDWNet (Traditional Dwelling Door & Window Net) 网络模型。

3.3 基于 YOLOv2-TDDWNet 的传统民居门窗检测

三种模型分别对训练集进行 10200 次训练之后得到各自训练好的网络模型。本节进一步对验证集进行了测试, 使用三种改进网络模型可以对传统民居门窗进行检测。

3.3.1 测试结果与分析

本文测试样本选用六种传统民居门窗类型, 包括单张门窗图像和混合门窗图像共 1386 张。

将三种网络各自训练过程中的精确率 Precision 和召回率 Recall 的值绘制成 P-R 曲线^[11], 如图 7 中 P-R 曲线图所示。曲线下纵轴 Precision 与横轴 Recall 所围合的面积即为 mAP 值, 可以看出 Model 2 的围合面积最大, mAP 值最高。

通过 mAP 曲线 (图 7) 我们可以看出, 在初始学习过程中, 曲线抖动剧烈, 这代表随着迭代次数逐渐增加, 网络学习情况还不稳定, 平均准确率甚至存在降低趋势, 这是由于网络发生了过拟合所导致的。但当迭代次数达到 9000 次左右时, 三种网络模型 mAP 均趋于稳定, 此时的模型已经有了各自稳定的学习能力。由此可以看出, Model 2 的平均准确率最高。

表 3 统计了三种模型对于六类典型门窗的检测准确率及网络平均准确率, 从数据角度展示了 Model 2 的准确率最高。

由图 7、表 3 可以看出, 三种传统民居门窗识别算法的 mAP 值存在一定差异, Model 2 即 YOLOv2-TDDWNet 的 mAP 值较高, 且单张图片检测耗时仅为 13ms, 实时性良好。

总体而言, YOLOv2-TDDWNet 对于山西传统民居典型门窗已具有较优的识别效果。

3.3.2 检测效果展示

Model 2 山西传统民居典型门窗的检测效果如图 8 所示。其中包括: (1) 单目标检测效果; (2) 存在遮挡时检测效果; (3) 不同光照下检测效果; (4) 多种门窗同时存在时检测效果; (5) 复杂背景下小尺度目标检测效果。

根据检测结果来看, 基于改进 YOLOv2 的传统民居门窗识别网络 YOLOv2-TDDWNet 检

测效果理想, 当目标存在遮挡^[12]、光照变化、模糊等干扰情况时仍具有较佳检测效果, 符合优质网络精准度标准。对于复杂背景下小尺度目标, YOLOv2-TDDWNet 显现出强大优势, Model 1 与 Model 3 均会出现漏检的现象, 而 YOLOv2-TDDWNet 可完美识别出图像中出现的所有的已知类别门窗, 识别效果较佳^[13]。

结语

本文研究了山西传统民居六种典型门窗图像识别的探索尝试。通过分析 YOLOv2 原模型的待优化之处, 设计出三种不同的改进模型, 使用同样本集分别对不同模型进行训练, 记录训练过程中的各项参数绘制成 mAP, IOU, Loss 等参数曲线, 挑选出适用于山西传统民居门窗识别的网络结构 YOLOv2-TDDWNet。

参考文献

- [1] 李珣, 时斌斌, 刘洋, 等. 基于改进 YOLOv2 模型的多目标识别方法 [J]. 激光与光电子学进展, 2020, 57(10): 113-122.
- [2] Girshick R, Donahue J, Darrell T, et al. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation [C]// CVPR. IEEE, 2014: 8-13.
- [3] 余玉琴, 魏国亮, 王永雄. 基于改进 YOLOv2 的无标定 3D 机械臂自主抓取方法 [J]. 计算机应用研究, 2020, 37(05): 1450-1455.
- [4] 白中浩, 李智强, 蒋彬辉, 等. 基于改进 YOLOv2 模型的驾驶辅助系统实时行人检测 [J]. 汽车工程, 2019, 41(12): 1416-1423.
- [5] 宋建国, 吴岳. 基于 YOLOv2 模型的道路目标检测改进算法 [J]. 软件导刊, 2019, 18(12): 126-129.
- [6] 谢欢芳. 地理实践力视域下的校内地理课程资源开发——以“形色”App 运用于校园植物调查为例 [J]. 地理教育, 2020(02): 51-52.
- [7] Redmon J, Divvala S, Girshick R, et al. You Only Look Once: Unified, Real-Time Object Detection [J]. 2016: 779-788.
- [8] 刘奇奇. 基于机器视觉的内窥镜器械检测和跟踪算法研究 [D]. 武汉: 华中科技大学, 2019.
- [9] 周志华. 机器学习 [M]. 北京: 清华大学出版社, 2016.
- [10] Rosebrock A. Intersection over Union (IoU) for object detection [J]. <http://www.pyimagesearch.com/2016/11/07/intersection-over-union-iou-for-object-detection/>, 2016.
- [11] 刘雪奇, 孙胜利. 基于 YOLO 网络模型的异常行为检测方法研究 [J]. 电子设计工程, 2018, 26(20): 154-158+164.
- [12] 芮红娟, 孙跃祥, 成洁. 基于 YOLO 与改进的 DLIB 多角度遮挡人脸判别方法 [J]. 计算技术与自动化, 2018, 37(04): 83-89.
- [13] 苏州大学. 基于改进 YOLOv2 的民居门窗图像识别提取软件 V2.0 [CP/CD]. 著作权登记号: 2020 SR0279183. 2020-03-20.