Computer Science and Application 计算机科学与应用, 2021, 11(11), 2788-2794
Published Online November 2021 in Hans. http://www.hanspub.org/journal/csa
https://doi.org/10.12677/csa.2021.1111283

Hans汉斯

基于YOLOv3的建筑工地目标检测研究

王晓宇[1]，张长伦[1,2]，何　强[1]，王恒友[1]，刘屹伟[3]

[1]北京建筑大学理学院，北京
[2]北京建筑大学，北京未来城市设计高精尖创新中心，北京
[3]北京建筑大学理学院，北京

收稿日期：2021年10月24日；录用日期：2021年11月22日；发布日期：2021年11月29日

摘　要

随着智慧工地的产生和发展，建筑工地施工现场各类监测技术的要求日益提高，为了更好地监测施工现场各类行为是否符合规范需要提高目标检测算法的精确度。本文为了更准确地检测建筑工地场景下的真实图像，采用MOCS数据集验证目标检测效果。首先用无监督的深度学习去噪网络Noise2noise进行去噪，其次将去噪后的图像送入深度学习网络YOLOv3进行目标检测。经过去噪后的图像目标检测的效果有一定的提升。

关键词

目标检测，建筑工地场景，YOLOv3，Noise2noise

Research on Construction Site Target Detection Based on YOLOv3

Xiaoyu Wang[1], Changlun Zhang[1,2], Qiang He[1], Hengyou Wang[1], Yiwei Liu[3]

[1]Science School, Beijing University of Civil Engineering and Architecture, Beijing
[2]Beijing Advanced Innovation Center for Future Urban Design, Beijing University of Civil Engineering and Architecture, Beijing
[3]Beijing University of Civil Engineering and Architecture, Beijing

Received: Oct. 24th, 2021; accepted: Nov. 22nd, 2021; published: Nov. 29th, 2021

Abstract

With the emergence and development of smart construction sites, the requirements of various monitoring technologies on construction sites are increasing day by day. In order to better monitor the compliance of various behaviors on construction sites, the accuracy of target detection algorithms needs to be improved. In order to more accurately detect the real image in the construc-

文章引用: 王晓宇, 张长伦, 何强, 王恒友, 刘屹伟. 基于 YOLOv3 的建筑工地目标检测研究[J]. 计算机科学与应用, 2021, 11(11): 2788-2794. DOI: 10.12677/csa.2021.1111283

tion site scene, this paper uses MOCS data set to verify the target detection effect. Firstly, Noise2noise, an unsupervised deep learning denoising network, is used for denoising. Secondly, the denoised images are sent to YOLOv3, a deep learning network for target detection. After denoising, the effect of image target detection is improved to some extent.

Keywords

Object Detection, Construction Site Scene, YOLOv3, Noise2noise

1. 引言

近年来人工智能的发展推动了智慧工地[1]的产生，智慧工地采用先进的计算机网络通信技术、数字视频处理技术、视频监控技术和计算机视觉技术加强对建筑工地的实时监测管理。其中计算机视觉技术在建筑工地目标检测方面取得了巨大的进展，例如工人安全监测、安全生产监控等。而计算机视觉技术的底层算法是基于深度学习的目标检测算法，通过获得建筑工地的数字视频和图像，目标检测算法可以实现建筑工地的目标识别。

近几年来，目标检测算法取得了很大的突破。其中一种主流的算法叫做两阶段目标检测算法，如 Faster R-CNN [2]等。Faster R-CNN 网络中提出的区域提案网络属于第一阶段用于产生提案框，第二阶段用于对提取的特征进行分类与回归。但两阶段方法由两部分组成，在速度上很难达到实时的效果，在应用上也产生了很大的局限性，故多采用一阶段算法。一阶段算法是一个端到端的算法，如 YOLOv1 [3]，它大大提升了目标检测的速度，但其直接回归检测框的宽和高导致检测效果不佳。所以 YOLOv2 [4]网络回归基于先验框的偏移量，达到了精度与两阶段算法相当。YOLOv3 [5]延续了 YOLOv2 的优势并采用多尺度网络进一步完善小目标检测的需求，大大提升了目标检测的速度并使目标检测达到了实时的效果，常用于工业落地使用。

建筑工地施工现场获得的图片在照明条件差、相机抖动、物体运动、空间像素未对准、颜色亮度不匹配、拍摄角度不合适、恶劣天气等情况下存在噪声。真实图片中存在的噪声未知，噪声类型多种多样且噪声分布复杂。这些未知噪声的存在进而影响目标检测的准确性。图像去噪研究是图像处理的重要手段，图像去噪的本质就是从观测值中分离噪声，保留干净图像。

基于上述分析，本文提出了一种适用于建筑工地施工现场的目标检测方法。本文的主要贡献分为两个部分：

1) 将建筑工地场景下的真实图像用于无监督的深度学习去噪方法 Noise2noise [6]进行去噪，解决噪声存在影响目标检测的准确性。

2) 将去噪后的图像送入深度学习目标检测网络 YOLOv3 进行目标检测。

本文结构如下：第二章介绍 Noise2noise 与 YOLOv3 网络，第三章展示实验，第四章总结本文。

2. 相关知识

2.1. Noise2noise 去噪网络

图像去噪是去除被噪声污染过的图像，并且把图像中的原有信息尽可能多地保留下来。针对不同噪

声的类型，产生了各种算法大致分为如下几类：滤波类、稀疏表达、外部先验、聚类低秩、深度学习[7]等。相对于传统的去噪算法，基于深度学习的图像去噪算法可以拟合复杂的噪声分布并节约运算时间。当前去噪方法被广泛地应用于合成噪声，而对日常生活中的真实噪声泛化性较差。真实噪声往往难以被参数化，并且一幅真实图像中不同位置的噪声类型和噪声分布可能是不一样的。所以对真实图片处理真实噪声是一个有重要意义的问题。

Noise2noise 在 2018 年被提出，是一种基于深度学习无监督的去噪算法，如图 1。可以看出该网络结构主要分为三个部分：下采样、上采样和横向连接。下采样过程通过卷积和池化操作实现特征提取是网络的编码部分，解码部分是通过卷积和上采样实现。横向连接采用通道拼接的方式将编码、解码阶段的特征图拼接在一起，以结合网络的浅层特征与深层特征。

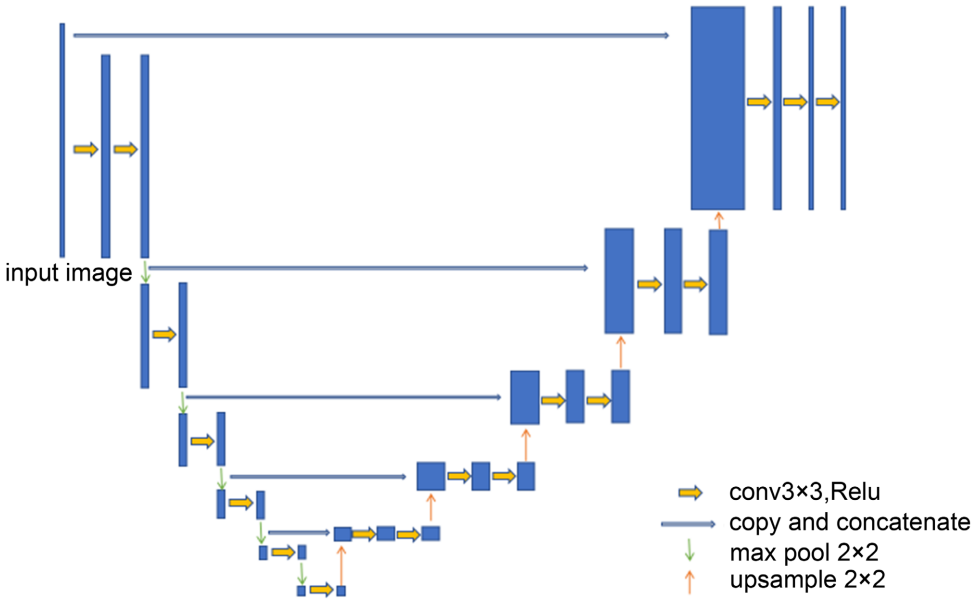

Figure 1. Network structure
图 1. 网络结构

这种方法不需要无噪声图像即可训练。具体来说，Noise2noise 使用成对的含噪图像进行训练，成对的含噪图像是指对同一张图像分别施加两次噪声得到两个不同的含噪图像，施加噪声的操作是相互独立的。即两张含噪图像的噪声部分是来自于同一个噪声分布的独立采样。Noise2noise 模型的训练数据首先要满足噪声在不同像素之间是不相关的，即独立的。例如，真实图像上的噪声可以看成很多不同的概率分布的随机变量的加和，并且每一个随机变量都是独立的。其次 Noise2noise 模型要求噪声均值为零，一般的高斯白噪声即可满足，如在拍摄时不够明亮、亮度不够均匀导致的噪声。所以对于 Noise2noise 在训练过程中优化损失函数如等式(1)所示。

$$L(f)=E\left\|f(x_1)-x_2\right\|_2^2 \tag{1}$$

其中 $f(x):R^m\to R^m$ 表示去噪网络，$x_1=y+n_1$，$x_2=y+n_2$，n_1 和 n_2 为相同的噪声模型在两张图像上的不同的实现，他们是独立同分布的。当样本数量较少时，卷积神经网络会学习到两种噪声的映射关系。当样本数量足够多时，由于噪声是随机的，网络在最小化损失函数的时候，就可以学习到干净的图像。因为噪声是随机的，有限的参数是无法将一个随机量映射到另一个随机量，那么唯一的策略就是将噪声去除。所以 Noise2noise 模型训练后的网络，可以处理真实图片的去噪问题。

2.2. YOLOv3 目标检测网络

YOLOv3 将整张图片输入训练好的卷积神经网络，则会直接输出目标的类别、置信度以及边界框，这在保证检测精度的同时大大加快了检测速度。模型相对于 YOLOv2 框架改进不大，在骨干网络、先验框设计与损失函数方面进行了改进。

骨干网络：为了达到更好的分类效果，YOLOv3 设计并训练了 darknet-53，相比于 resnet-101 和 resnet-152 有效减少了网络层数并提高了训练速度，相比于 YOLOv2 的 darknet-19 提高了网络精度。为了降低池化带来的梯度负面效果，darknet-53 采用全卷积网络利用步长进行降采样并且通过调节步长控制输出特征图的尺寸，因而未对输入图片尺寸进行限制。为了加强算法对目标尺度检测的多样性，借鉴特征金字塔思想，相比于 YOLOv2 采用 passthrough 结构来检测细粒度特征，YOLOv3 进行了 3 个尺度特征图的融合并进行 3 条支路的预测。

先验框设计：为了匹配图像目标尺度和输出特征图尺度的多样性，YOLOv3 延续了 YOLOv2 的技巧并采用 K-means 聚类得到先验框的尺寸，并为每种采样尺度设计了三种比例的先验框，共聚类 9 种先验框尺寸。

损失函数：为了支持多标签分类，YOLOv3 放弃 YOLOv2 预测对象时采用的 softmax，改用 logistic 回归判断目标位置的置信度。

YOLOv3 的框架大体分为主干网络和预测头，主干网络又名 darknet53，其包含 53 个卷积层。输入主干网络的过程可以理解为对特征进行编码的过程。主干网络最后三层采用特征金字塔网络模型，由于分别参与不同尺度目标的预测，而分别进行卷积和上采样，结果即为各个特征层的预测结果，这可以理解为解码的过程。解码过程的输出结果经过置信度的排序和非极大值抑制的筛选可以得到与目标物体预测最契合的边界框。其具体步骤在于首先对图片分为方块区域，对每个方块区域预设锚框，其次采用卷积神经网，对整张图片的每个分块进行预测框并分类，从而得到检测结果。

3. 实验与分析

3.1. 实验平台及参数设置

本文实验测试环境在 Intel(R)Core(TM) i7-10875H CPU @ 2.30GHZ，16 GB 内存，Ubuntu 18.04 系统下搭建的 tensorflow.keras 平台环境完成对图像去噪方法的测试任务。实验环境的配置参数如表 1 所示。

Table 1. Configuration parameters of the test environment
表 1. 测试环境的配置参数

参数	版本或数值
操作系统	Ubuntu18.04
CPU	Intel(R)Core(TM) i7-10875H CPU @ 2.30GHZ
GPU	GeForce RTX 2060 with Max-Q Design
CUDA	CUDA10.2
Keras	Keras2.4.1

在训练过程中为无噪图像分两次分别加入均值为零，方差为(0, 50)的高斯白噪声，输入图片的大小 128*128，批处理数设为 8，迭代次数设置为 60，初始学习率设置为 0.001，且随着迭代次数的增加，学习率越来越低。测试过程中不对真实图片加入噪声，直接送入网络进行测试。

3.2. 数据集

本次实验采用 MOCS 建筑工地数据集[8]。MOCS 数据集包含 174 个建筑工地覆盖了公路、桥梁、隧道、水坝、人行道和室内装饰等各种项目。建筑工地中的移动物体数据集包含 41,668 张图像，并通过边界框和掩模的样式对其中 13 个类别进行标注，数据集中带有 222,861 个附注释的实例。针对不同的自然条件采用智能手机、数码相机、无人飞机以及监控等设备采集了不同天气条件和照明条件下的数据，例如晴天、下雨、多雾和下雪情况以及夜间和隧道中的现场照明条件。并未从特定角度收集物体特征而是尝试记录现场环境。故为建筑工地场景下含有真实噪声的真实图片。

3.3. 实验效果与分析

本研究是基于目标检测算法 YOLOv3 提高建筑工地场景下目标检测的效果。为了测试模型的有效性，我们通过将含有噪声的建筑工地数据集输入 Noise2noise 网络进行去噪，去噪后的输出图输入 YOLOv3 网络进行目标检测，将使用 Noise2noise 网络去噪后的检测结果与未使用 Noise2noise 网络去噪的检测结果进行对比如图 2、图 3 所示。建筑工地场景下具有明显特点的目标例如：人和施工现场的建筑设备例如：塔吊、泵车和卡车等检测方面，由于拍摄角度导致目标遮挡聚集且背景复杂。并且由于光照、雾气等自然噪声和建筑工地场景下的脚手架、防尘布等特定噪声的存在，原始的目标检测结果易产生冗余，从而降低了网络的精确度，而我们的 Noise2noise 网络大大抑制了这种情况的发生，经过 Noise2noise 网络去噪后的检测效果有一定的提升。这充分说明本文提出的模型适用于建筑工地场景下的目标检测。

Figure 2. The target detection result of the original image
图 2. 为原始图片目标检测结果

Figure 3. The target detection result of the denoised image in this paper
图 3. 为本文经过去噪图片目标检测结果

4. 结语

本文对基于深度学习的无监督去噪算法 Noise2noise 与目标检测算法 YOLOv3 做了表述，并将这两种方法应用到建筑工地场景下解决含有真实噪声的真实图片目标检测效果不佳的实际问题。本文采用建筑工地场景下的数据集MOCS，首先将图片送入Noise2noise网络进行去噪，其次将输出结果送入YOLOv3目标检测网络进行目标检测。与直接送入目标检测网络的检测结果进行对比，目标检测效果有一定的提升。但是对小目标的检测效果有所下降并且对由于二维图片丢失深度空间信息导致的错检、漏检等问题没有涉及到，后续继续对类似问题进行研究，以尽可能提高目标检测精度。

基金项目

国家自然科学基金(No. 62072024)；北京建筑大学北京未来城市设计高精尖创新中心资助项目(UDC2017033322，UDC2019033324)；北京建筑大学市属高校基本科研业务费专项资金资助(NO. X20084，ZF17061)。

参考文献

[1] 李瑞平, 杜瑞. 智慧工地管理平台在建筑工程中的应用探究[J]. 智能建筑与智慧城市, 2021(10): 60-61.

[2] Ren, S., He, K., Girshick, R., *et al.* (2015) Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **39**, 1137-1149. https://doi.org/10.1109/TPAMI.2016.2577031

[3] Redmon, J., Divvala, S., Girshick, R., *et al.* (2016) You Only Look Once: Unified, Real-Time Object Detection. 2016 *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Las Vegas, 27-30 June 2016, 779-788.

https://doi.org/10.1109/CVPR.2016.91

[4] Redmon, J. and Farhadi, A. (2016) YOLO9000: Better, Faster, Stronger. 2016 *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Las Vegas, 27-30 June 2016, 6517-6525. https://doi.org/10.1109/CVPR.2017.690

[5] Redmon, J. and Farhadi, A. (2018) YOLOv3: An Incremental Improvement.

[6] Lehtinen, J., Munkberg, J., Hasselgren, J., *et al.* (2018) Noise2Noise: Learning Image Restoration without Clean Data. *Proceedings of the* 35*th International Conference on Machine Learning*, Stockholm, PMLR 80 2018.

[7] 刘迪, 贾金露, 赵玉卿, 钱育蓉. 基于深度学习的图像去噪方法研究综述[J]. 计算机工程与应用, 2021, 57(7): 1-13.

[8] An, X.H., Zhou, L., Liu, Z.G., Wang, C.Z., Li, P.F. and Li, Z.W. (2021) Dataset and Benchmark for Detecting Moving Objects in Construction Sites. *Automation in Construction*, **122**, Article ID: 103482. https://doi.org/10.1016/j.autcon.2020.103482