

一种提高目标图像识别准确率的数据增强技术

吴天雨¹,许英朝^{2,3},晁鹏飞¹

(1. 厦门理工学院 电气工程及其自动化学院,厦门 361024;
2. 厦门理工学院 光电与通信工程学院,厦门 361024;
3. 厦门理工学院,福建省光电技术与器件重点实验室,厦门 361024)

摘要:针对当下使用神经网络算法检测目标时,普遍存在的真实数据量缺乏,人为采集难度大,图像标注工作繁琐等问题,提出一种基于卷积神经网络的图像数据增强方法。通过对原始图像数据进行几何变换、水平翻转,改变图像像素位置并保证其对应特征不变,实现数据量呈指数型增长。通过模拟不同的光照条件和拍摄角度,丰富数据库种类的多样性,提高实际检测的鲁棒性。实验结果表明,在对图像数据进行增强处理后,训练出的神经网络模型对于目标的检测性能更高。

关键词:神经网络;目标检测;图像标注;数据增强;几何变换;光照强度

中图分类号: TN249 **文献标识码:** A **DOI 编码:** 10. 14016/j. cnki. jgzz. 2020. 05. 096

Dataaugmentation technology for improving target image recognition

WU Tianyu¹, XU Yingchao^{2,3}, CHAO Pengfei¹

(1. School of Electrical Engineering and Automation, Xiamen University of Technology, Xiamen 361024, China;
2. School of Optoelectronics and Communication Engineering, Xiamen University of Technology, Xiamen 361024, China;
3. Fujian Key Laboratory of Optoelectronic Technology and Devices, Xiamen University of Technology, Xiamen 361024, China)

Abstract: Aiming at the current use of neural network algorithm to detect targets, the lack of real data, the difficulty of human collection and the cumbersome image labeling, a method of image data augmentation based on convolutional neural network is proposed. By geometrically transforming and horizontally flipping the original image data, the image pixel position is changed and the corresponding features are kept unchanged, and the data volume is exponentially increased. By simulating different lighting conditions and shooting angles, the variety of database types is enriched, and the robustness of actual detection is improved. The experimental results show that the trained neural network model has higher detection performance for the target after the image data is enhanced.

Key words: neural network; target detection; image annotation; data augmentation; geometric transformation; illumination intensity

1 引言

对于机器学习任务,例如图像分类、机器翻译和文本到语音互译,可用于训练的样本数量对于实现高

性能至关重要。在原始训练数据中应用数据增强方法并综合创建成新数据集被广泛用于神经网络的训练。当使用小的训练集训练模型时,训练后的模型往往会过度拟合训练集中的样本,从而导致测试效果不佳。数据增强通过扩大训练集的大小来避免这种过度拟合的问题,例如 Krizhevsky^[1],从 256×256 像素的原始图像中随机裁剪 224×224 像素的色块,水平翻转提取的图像补丁,来更改 RGB 通道的强度,通过多种增强方法从原始训练数据中合成新的样本数据^[2-3]。常见的数据增强技术包括翻转或者扭曲输入图像,添加少量噪声或从随机位置裁剪。数据增强的使用通

收稿日期:2019-11-12

基金项目:福建省教育厅(No. JK2017036);福建省自然科学基金面上项目(No. 2019J01876);厦门市科技计划项目(No. 3502Z20183060);厦门市科技计划重大项目(Micro-LED 技术研发及产业化)

作者简介:吴天雨(1995-),男,在读硕士生,主要研究方向为智能照明电器与光电子器件。E-mail: 756738690@qq.com

通信作者:许英朝(1980-),男,教授,博士,硕士生导师,主要研究方向为从事光电器件与工艺研究、半导体照明技术与新能源等方面的研究。E-mail: yexu@xmut.edu.cn

常是赢得图像分类竞赛的标准。

在图像识别中,为了避免出现过拟合,通常需要大量的数据训练模型。但很多情况下,经常会遇到数据不足的问题。为了丰富图像训练集,更好的提取图像特征,泛化模型,通常需要通过数据增强来扩充数据。但已存在的扩充方式多是采用物理扩充的方式,这样的方式往往会占据大量磁盘空间。

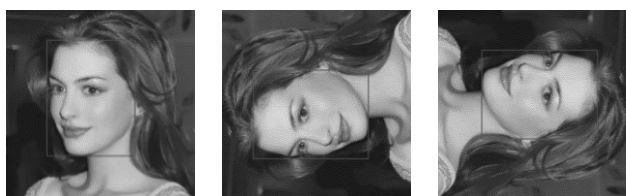
针对这些问题,本文在 faster-rcnn 内部实现了多种数据增强方式^[4]。Faster-rcnn 模型是当前目标检测任务中主流的深度学习框架,凭借较快的速度和较高的精度,该模型得到了广泛的应用。由于只在训练时占用内存,同时实现了检测时的数据增强,并通过实验来对比验证各种增强方法的有效性,最终的识别准确率平均保持在 90% 以上。

2 数据增强方式介绍

在 faster-rcnn 框架下,主要实现了四种数据增强方式:左右各旋转 90°、随机裁剪、颜色扰动以及随机尺度缩放。在 faster-rcnn 框架内部实现数据增强时,需将图片本身以及对应的目标框分别增强。

2.1 左右旋转 90°(Rotate±90)

读取一张图片,同时实现图片和目标框的左、右各翻转 90°。尤其对于形状不规则的目标物体,通过旋转,可以增加多种角度下的数据,如图 1 所示。



(a) 原图及目标框 (b) 左转 90° (c) 右转 90°

图 1 图像及目标框的旋转

每张图片对应一个标注文件(XML),标注文件的作用是记录了图片里目标框的坐标位置,在进行图像旋转时,图像本身会旋转,而标注里的目标框也要相应地旋转,XML 标注文件里对坐标的标注是以图像的左上角为坐标原点,每个目标框的起始坐标为(x_{\min} , y_{\min} , x_{\max} , y_{\max})如图 2 所示。

假设旋转后的坐标为(X_{\min} , Y_{\min} , X_{\max} , Y_{\max}),正 90°变换的坐标计算如下:

$$X_{\min} = H - y_{\max} \quad (1)$$

$$Y_{\min} = x_{\min} \quad (2)$$

$$X_{\max} = -H - y_{\min} \quad (3)$$

$$Y_{\max} = x_{\max} \quad (4)$$

负 90°变换坐标如下:

$$X_{\min} = y_{\min} \quad (5)$$

$$Y_{\min} = W - x_{\max} \quad (6)$$

$$X_{\max} = y_{\max} \quad (7)$$

$$Y_{\max} = -W - x_{\min} \quad (8)$$

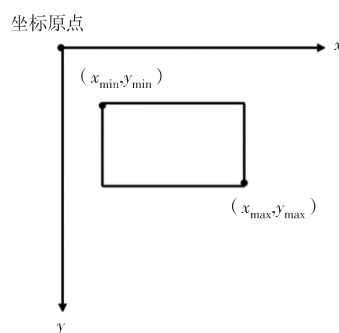
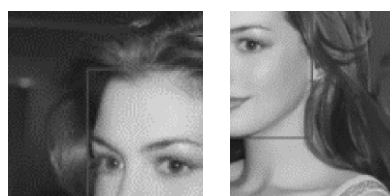


图 2 XML 文件中目标框的坐标标注形式

2.2 随机裁剪(Random crop)

图像中的目标有可能会被遮挡,因此有的目标并不是以完整的形态出现。对于这类目标,原有的训练数据中缺少这类数据,检测时难以检测出这类不完整目标。因此,在训练过程中生成图像及目标框同时裁剪的训练数据就很有必要,采用随机裁剪的方式可以实现,裁剪后的示例图如图 3 所示。



(a) 随机裁剪 1 (b) 随机裁剪 2

图 3 随机裁剪

此外,随机裁剪时容易将目标剪裁的太小,如图 3 (b),导致背景过多,不利于训练,同时在检测时很可能造成局部检测重复的问题,因此在裁剪时限制目标剪裁掉部分不能太大,最好不要超过目标本身大小的 1/3。

2.3 颜色扰动(PCA Jettering)

图像目标的颜色有时可能会随着环境的变化而变化,因此除了几何变换外,还可以通过颜色变换来进行数据增强,如图 4 所示。首先按照 RGB 三个颜色通道计算均值和标准差,随后计算协方差矩阵进行

特征分解,得到特征值和特征向量用来进行 PCA Jettering^[5-7]。



(a) 原图 (b) PCA Jettering

图 4 原图与 PCA Jettering 后的对比图

2.4 尺度缩放(Scale augmentation)

图像中的目标可能会因拍摄角度而造成目标本身比例大小不同,在 faster-rcnn 中,图像短边会被固定到某个值,长边按相应的比例进行缩放,在实验时可以通过改变图像的长宽比来进行随机尺度缩放(0.5~2 倍),如图 5 所示。



(a) 宽/高比为 0.5 (b) 宽/高比为 1.2

图 5 随机缩放后的图像对比

2.5 高斯噪声

当神经网络试图学习可能无用的高频特征时,通常会发生过度拟合,从而扭曲高频特征,这也意味着较低频率的数据也会逐渐失真。具有零均值的高斯噪声基本上在所有频率中具有数据点,添加适量的噪声可以增强学习能力,它表现为随机的黑白像素在图像中传播。通过向图像添加高斯噪声获得较低的信息失真水平,并提高算法的鲁棒性。如图 6 所示,左边为原始图片,右边为引入噪声后的效果图。



(a) 原图 (b) 引入少量高斯噪声

图 6 引入高斯噪声后的图像对比

3 实验结果与分析

3.1 实验参数与配置

实验时采用 faster-rcnn 框架下默认参数,在 Ubuntu14.04 Intel Xeon CPU TITAN X 下,分别进行物品识别实验和物品数量检测实验。

3.2 实验数据

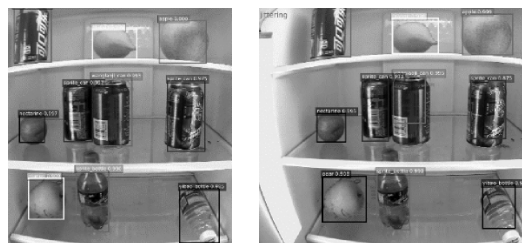
训练数据来源为两部分,一部分为网络爬虫获取的网络图片(每类约 500 张),一部分为摄像头拍摄的饮料水果等图像数据(每类约 500 张)。

3.3 实验结果

图 7 为部分检测结果。其中图(a)(b)为物品数目检测实验的结果。图(a)中原模型无法检测出竖直角度的盒装物品,而经过左右旋转 90°的增强训练后,就可以准确地将其检测出。在图(c)中,原检测器误将左下角的梨误检为芒果,且概率值较高(0.9 以上),图(d)是在 PCA Jettering 增强下训练的模型检测结果,由结果可知,该检测器能够准确地检测出梨。图(f)为尺度缩放增强的检测结果,与(e)对比可知,这种增强方式可在一定程度上减少较小目标框的生成。



(a) faster-rcnn 检测结果 (b) Rotate $\pm 90^\circ$ 增强检测结果



(c) faster-rcnn 检测结果 (d) PCA Jettering 增强检测结果



(e) faster-rcnn 检测结果 (f) 尺度缩放增强检测结果

图 7 部分检测结果对比

目标检测中衡量识别精度的指标是 mAP(mean average precision) [8-10], 由表 1 可知, 使用不同的数据增强方式后, 除随机裁剪外, 结果均有所改善。在检测模型中, 随机裁剪会将目标切分成多块, 检测时可能会引入重复检测的问题, 因此造成检测精度下降。

表 1 不同增强方式下的实验结果

增强方法	mAP
Faster-rcnn	0.930 2
Rotate ± 90	0.935 1
Random crop	0.921 8
PCA Jettering	0.948 0
Scale augmentation	0.947 0

除了在训练过程中进行数据增强, 还可以在检测时进行数据增强。检测时的数据增强可以直接使用训练时的数据增强方式, 图 8 为部分检测结果。其中图 8(a) (b) 为检测时对顺时针 90° 和逆时针 90° 变换的检测结果。通过对比可以看出, 虽然原 faster-rcnn 中未检测的目标在经过旋转处理后可以检测出来, 但原 faster-rcnn 中可以检测出来的部分目标, 在旋转处理后会被漏检或错检。图 8(c) 为原 faster-rcnn 检测结果, 图 8(d) 为检测时对图像进行 PCA Jettering 增强后的识别结果。由图可以看出, 原本误检的梨可以被正确检测, 但原 faster-rcnn 中可以正确检测的芒果, 在经过增强后出现漏检现象。图 8(e) 和 (f) 为改变图像宽高比时的检测结果, 由图可知, 对宽高比的改变会导致物体本身发生形变, 降低检测效果。

通过实验可以发现, 训练时的所用的数据增强方式并不适用于检测时的增强。此外, 通过图像增强, 改变图像光照分布, 会改善检测结果, 如图 8(g) (h) 所示。



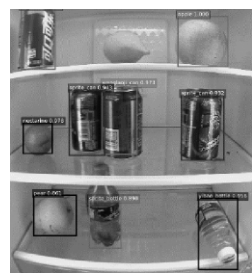
(a) 顺时针 90°



(b) 逆时针 90°



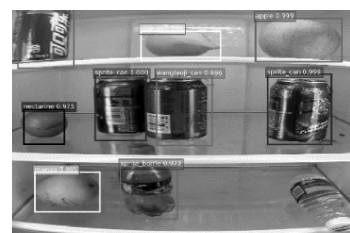
(c) faster-rcnn 检测结果



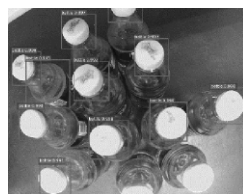
(d) PCA Jettering 增强检测结果



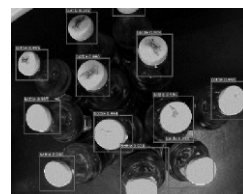
(e) 宽高比为 0.5



(f) 宽高比为 1.8



(g) faster-rcnn 检测结果



(h) gamma 增强检测结果

图 8 检测时增强的部分检测结果

4 其他数据增强方法

4.1 增加 anchors 个数

在 RPN 提取的特征图中, 一个特征点映射为原图, 对应 9 个 anchors(3 种尺度 \times 3 种宽高比例)。这种方式生成的候选框较少, 可能会遗漏掉较小或者较大的目标 [11-13]。为了生成尺度更多的 anchors, 本文在实验中增加了 3 种宽高比例(即生成 18 个 anchors), 由表 2 可知, 通过增加 anchors 的比例, 可以提高检测精度。

表 2 改变 anchors 的对比实验

方法	mAP
Faster rcnn	0.930 2
Anchors ratio	0.941 6

4.2 畸变校正

在用摄像头采集图像进行检测时, 采集的图片在边缘会产生畸变, 称为桶形失真 [14-15]。本次实验在检测前将图像进行桶形失真校正。由实验结果可以发现, 经过校正后的图像, 检测结果并没有改善, 有些甚至会出现错检的情况, 如图 9(a) (b) 所示。在物品数量检测实验中, 校正后的图像反而会出现误检。此外, 校正后的图像都会损失视野, 在边缘处的目标可能会被丢失。



(a) 校正前检测结果



(b) 校正后检测结果

图 9 校正前后检测对比实验

5 结语

针对神经网络训练过程中普遍存在的数据量不足、数据种类不够丰富等问题,提出了一种应用在训练数据和实际检测两大过程中的数据增强技术。为了验证本文方法的有效性,通过对水果、饮料等生活用品进行了检测实验。实验结果表明,在训练时进行数据增强可以提高模型的检测效果,而在检测时采用类似的数据增强方式和桶形失真校正可能会使检测结果恶化,同时增加了检测时间。另外光照条件对检测结果的影响较大,因此在检测时进行图像增强、改变图像的光照可能会改善最后的检测结果。如何选择有效的数据增强和图像增强方式来改善检测结果,是下一步需要研究的内容。

参考文献

- [1] KRIZHEVSKY A, SUTSKEVER I, HINTON G. ImageNet Classification with Deep Convolutional Neural Networks [C]// NIPS. Curran Associates Inc. 2012: 815-823.
- [2] 卢宏涛,张秦川. 深度卷积神经网络在计算机视觉中的应用研究综述 [J]. 数据采集与处理, 2016, 31(01): 1-17.
- [3] 黄斌,卢金金,王建华,吴星明,陈伟海. 基于深度卷积神经网络的物体识别算法 [J]. 计算机应用, 2016, 36(12): 3333-3340+3346.
- [4] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2015, 39(6): 1137-1149.
- [5] 史鹤欢,许悦雷,马时平,李岳云,李帅. PCA 预训练的卷积神经网络目标识别算法 [J]. 西安电子科技大学学报, 2016, 43(03): 161-166.
- [6] 栾孟杰. 一种多分辨多尺度的红外图像增强算法 [J]. 激光杂志, 2019, 40(08): 81-84.
- [7] 南瑞亨,蒋晨杰. 面向暗室照度计检定低对比度示值图像智能识别技术研究 [J]. 激光杂志, 2019, 40(08): 56-59.
- [8] 王茜,张海仙. 深度学习框架 Caffe 在图像分类中的应用 [J]. 现代计算机, 2016(5): 72-75.
- [9] 韩袁琛,程勇. 基于朗伯反射模型的光照估计及鲁棒人脸识别 [J]. 电视技术, 2017, 41(01): 79-83.
- [10] 刘超,秦川,陶忠,等. 基于混合蛙跳优化算法的图像畸变校正研究 [J]. 应用光学, 2017(2).
- [11] 杜兰,刘彬,王燕,等. 基于卷积神经网络的 SAR 图像目标检测算法 [J]. 电子与信息学报, 2016, 38(12): 3018-3025.
- [12] 刘凯品,应自炉,翟懿奎. 基于无监督 K 均值特征和数据增强的 SAR 图像目标识别方法 [J]. 信号处理, 2017, 33(3): 452-458.
- [13] 蒋杰,熊昌镇. 一种数据增强和多模型集成的细粒度分类算法 [J]. 图学学报, 2018, 39(2): 244-250.
- [14] 刘万军,梁雪剑,曲海成. 基于双重优化的卷积神经网络图像识别算法 [J]. 模式识别与人工智能, 2016, 29(09): 856-864.
- [15] 朱虹,翟超,吕志,等. 基于 Faster-RCNN 的智能家居行人检测系统设计与实现 [J]. 工业控制计算机, 2018, 31(04): 68-70.