

结构主导目标检测中的纹理随机化*

王梓^{1,2}, 毕道明^{1,3}, 周颀鑫^{1,2}, 孙晓亮^{1,2}, 于起峰^{1,2}

(1. 国防科技大学空天科学学院, 湖南长沙 410073;

2. 国防科技大学图像测量与视觉导航湖南省重点实验室, 湖南长沙 410073;

3. 沈阳飞机设计研究所, 辽宁沈阳 110035)

摘要:已有基于卷积神经网络的目标检测算法倾向于提取目标纹理特征,而非结构特征;因此,已有方法不能实现变纹理目标的可靠检测。针对此问题,提出基于纹理随机化的结构主导目标检测方法,采用仿真纹理随机化方法减弱网络模型对纹理特征的拟合,实现基于结构特征的变纹理目标可靠检测。利用目标的三维模型,借助 Blender 渲染引擎,完成纹理随机化仿真训练数据集的生成。仿真及真实图像实验测试结果表明:该方法能够实现基于目标结构特征的变纹理目标可靠检测。

关键词:纹理随机化;仿真数据;目标检测;三维模型

中图分类号:TN391.41 文献标志码:A 文章编号:1001-2486(2021)04-024-07

Texture randomization in structure-dominant object detection

WANG Zi^{1,2}, BI Daoming^{1,3}, ZHOU Jiexin^{1,2}, SUN Xiaoliang^{1,2}, YU Qifeng^{1,2}

(1. College of Aerospace Science and Engineering, National University of Defense Technology, Changsha 410073, China;

2. Hunan Key Laboratory of Videometrics and Vision Navigation, National University of Defense Technology, Changsha 410073, China;

3. Shenyang Aircraft Design & Research Institute, Shenyang 110035, China)

Abstract: Region-based detectors with convolutional neural networks tend to learn textural rather structural feature and thus face substantial difficulties in detecting objects with various textures. To tackle this problem, the texture randomization to augment the synthetic training image dataset was employed and a novel method for structure-aware object detection was proposed. The texture-randomized simulation data were generated by rendering 3D model with varied textures using Blender. Experiments on synthetic and real images indicate that the proposed method is capable of robustly detecting texture-varied objects based on structural information.

Keywords: texture randomization; simulation data; object detection; 3D model

目标检测是计算机视觉中的基础问题,在机器人抓取^[1]、遥感图像处理^[2]等领域具有广泛应用。基于卷积神经网络(Convolutional Neural Network, CNN)的目标检测算法得益于大规模高质量标注的训练数据集,取得了当前最优的检测性能。本文关注变纹理目标的可靠检测。受工艺加工、应用需求、复杂光照等因素的影响,同一目标在纹理上可能存在巨大差异(如图1所示),已有基于CNN的目标检测算法倾向于提取目标纹理特征,而非目标的结构特征。因此,相关方法不能有效地实现图1所示变纹理目标的可靠检测。

目标结构特征是目标检测中的重要依据,表征目标结构特征的梯度方向直方图(Histogram of Gradient, HoG)特征^[3]已被广泛应用于目标检测

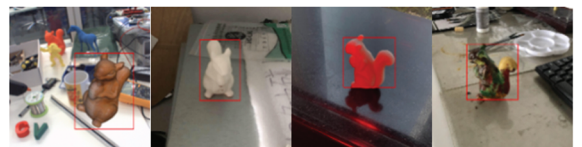


图1 变纹理目标图像示例

Fig. 1 Sample images of a texture-varied object

中^[4-6]。Aubry等^[4]采用HoG特征提取目标结构特征,并结合线性分析(Linear Discriminant Analysis, LDA)方法,在仿真图像数据集上完成目标检测器的训练,实现了仅基于目标结构特征的可靠目标检测,对光照变化、纹理差异等干扰表现出来较强的鲁棒性。在文献^[5]的基础上,Lim等^[6]引入目标局部结构区分度评价及可变形部件思想,进一步提升目标检测性能。Hariharan

* 收稿日期:2019-12-20

基金项目:湖南省自然科学基金资助项目(2019JJ50732)

作者简介:王梓(1994—),男,山东菏泽人,博士研究生,E-mail:wangzi16@nudt.edu.cn;

孙晓亮(通信作者),男,讲师,博士,E-mail:alexander_sxl@nudt.edu.cn

等^[7]对 HoG 特征进行改进,提出白化方向直方图(Whitened Histograms of Orientations, WHO)特征,并应用于目标检测中。

深度特征表征方法在特征表示能力上优于传统特征表征方法。因此,基于 CNN 的目标检测算法^[8-9]取得了明显优于传统方法的目标检测性能。但 Geirhos 等^[10]指出,已有基于 CNN 的目标检测方法倾向于提取目标的纹理特征,而非目标的结构特征,限制了已有目标检测模型的泛化性能,尤其是针对本文关注的变纹理目标情况。Zaech 等^[11]通过生成式对抗网络(Generative Adversarial Network, GAN)^[12]丰富训练数据中的纹理类型,有效降低了语义分割模型对纹理的拟合并成功应用于无人驾驶任务中。

相比于纹理特征,目标结构特征对于干扰更具有鲁棒性,有益于实现可靠的目标检测。受文献^[4, 10-11]的启发,本文针对变纹理目标的检测,提出一种新的基于纹理随机化的结构主导目标检测方法,依据目标三维模型,借助 Blender 渲染引擎,通过修改目标模型纹理贴图,构建纹理随机化仿真训练数据集,并训练目标检测器,降低模型对纹理特征的拟合程度,实现目标结构特征主导的目标检测。

1 相关工作

本文借助 Blender 渲染引擎构建纹理随机化仿真训练数据集,实现结构主导的变纹理目标检测。接下来,对与本文密切相关的已有工作进行简要梳理。

1.1 传统基于结构特征的目标检测方法

目标结构特征被广泛应用于目标检测,如被大家所熟知的 HoG 特征。Aubry 等^[4]以椅子为研究对象,采用 HoG 特征及 LDA,在仿真图像数据集完成检测器训练,依据目标结构特征实现图像中椅子目标的可靠检测,并给出了粗略的位姿估计结果,方法对纹理差异、光照变化等干扰表现出了较强的鲁棒性。Lim 等^[5]同样采用 HoG 特征提取目标结构特征,引入区分度量,筛选出目标模型上区分度高的结构特征用于分类器的训练,并依据目标不同部位结构特征的几何约束,剔除误匹配,提升算法的目标检测性能。在文献^[6]中,作者进一步引入可变形部件处理思想,提升了算法目标检测的准确率及效率。为适应光照变化等干扰,Hariharan 等^[7]基于梯度方向信息提出 WHO 特征,Choy 等^[13]进一步引入阈值处理,提出 NZ(non-zero)-WHO 特征,以消除仿真训练图

像中背景区域带来的干扰,并采用共轭梯度算法高效求解线性分类器权值。

1.2 基于形状和纹理的特征描述

形状描述子抽取形状的特征参数以描述形状,以实现形状检测、配准以及目标识别、检测、跟踪等。基于形状边界或形状二值图像的形状描述子,将形状视为基函数的线性组合,例如 Fourier 描述子^[14]、Chebyshev 描述子^[15]、Zernike 矩^[16]等。基于线性组合的形状描述子表达了目标的整体形状属性,缺少对目标空间关联属性的描述。基于空间关联关系的形状描述子充分利用了形状的局部细节信息,但是计算量较大,且难以重构出原始形状。

纹理是物体表面的内在属性,纹理特征也常被用于目标表示中。结构法、统计法和频谱分析法是纹理特征提取的主要方法。结构法将纹理视为一系列纹理基元的规律性排列,主要适用于周期性纹理的提取和分析;统计法以纹理图像灰度值空间分布的统计信息作为纹理特征,例如灰度共生矩阵^[17]等;频谱分析法通过 Gabor 变换、小波变换等方法,分析纹理图像的频谱信息。

基于形状和纹理的特征描述方法具有一定的特征提取和表达能力,在手工视觉特征领域具有重要地位。近年来随着对深度卷积神经网络的研究的不断深入,研究人员发现卷积网络中较低层的卷积类似于边缘、纹理提取算子,随着网络层数的不断增加,捕捉到的语义特征更加抽象、复杂。因此,使用深度卷积网络可以获得比手工描述子更加丰富的目标特征。

1.3 纹理偏移及纹理随机化

Wohlhart 等^[18]针对基于目标三维模型的目标检测问题,对比分析了 CNN 模型提取的深度特征及 HoG 特征等传统特征。实验结果表明:深度特征的表征能力优于 HoG 特征,取得了更优的目标性能。基于 CNN 的目标检测算法取得了当前最优的目标检测性能,相关算法一般采用端到端的处理流程,内部特征提取部分无法获知,对于基于 CNN 的目标检测模型中特征提取的解释主要有外观和纹理两个假设^[10]。Geirhos 等^[10]在 ImageNet 数据集^[19]上针对目标检测任务进行测试。实验结果显示:CNN 模型偏重于提取目标纹理特征,存在纹理偏移,甚至仅利用纹理特征即可完成目标识别任务。

纹理偏移严重限制了基于 CNN 的目标检测方法

随机性降低模型对纹理特征的依赖。Tobin 等^[20]针对工业机器人抓取应用中的目标检测需求,利用目标三维模型,借助仿真引擎生成仿真图像训练数据集,在生成数据集的过程中,选用 1 000 种纹理,增加了纹理的随机性,显著提高了模型的泛化能力。Ren 等^[21]采用相似的处理策略,提升了模型的检测率及位姿估计的精度。Zaech 等^[11]将纹理随机化策略应用到无人驾驶场景语义分割中,并在真实图像数据上取得了满意的分割结果。文献[20-21]针对机器人抓取任务,在仿真图像生成过程中,采用的纹理贴图较为简单,尚不能满足复杂环境下变纹理目标检测的需求。本文在文献[20-21]的基础上,着眼于仅利用主要结构特征实现目标检测,以适应目标纹理变化的情况,提出更具针对性的纹理随机化方法。

2 方法

提出基于纹理随机化的结构主导目标检测方法,消除纹理偏移,使目标检测器基于目标结构特征实现变纹理目标的可靠检测。

2.1 目标检测模型

采用 Faster RCNN 检测器^[8-9]作为目标检测基本框架,如图 2 所示。Faster RCNN 是第一个端到端的检测器,具有接近实时检测的性能。其主

要组件包括用于提取特征的主干网络、用于生成锚框的锚框生成器、用于似物性检测的区域网络 (Region Proposal Network, RPN)、用于提取特征图的 ROI Align 以及对 proposal 进行分类和回归的预测头。

在训练时将每个视角下的目标作为一个类别,并将视角标签编号作为分类分支监督信号,在测试时,完成输入图像中目标实例的检测,并输出对应的视角标签。

2.2 纹理随机化仿真数据集

为构建纹理随机化仿真图像训练数据集,利用目标的三维模型,借助 Blender 渲染引擎,通过改变目标模型纹理图像,生成变纹理仿真图像数据集,如图 3 所示。图 3(a)中,纹理图像随机选取 MS COCO 数据集^[22]中的自然图像。图 3(b)中,虚拟相机指向目标中心,通过离散采样方位角 α 和高度角 β 得到不同视角下目标投影图像。

将每一视角下的目标作为一个类别,在构建仿真图像数据集时,采用如图 3 所示的视角采样方式,通过离散采样虚拟相机的方位角 α 和高度角 β 得到不同视角下的目标投影图像。为了仿真目标在图像平面内的运动,随机化目标模型在相机坐标系下的位置 $\mathbf{T} = [T_x, T_y, T_z]^T$ 。除了保存采样视角作为仿真图像的分类标签,在仿真图像

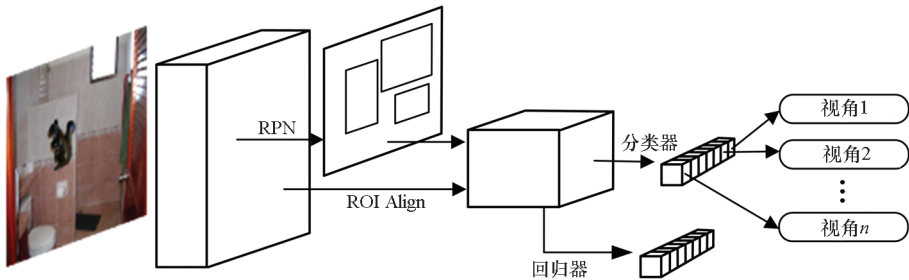


图 2 Faster RCNN 检测器结构图
Fig. 2 Structure of Faster RCNN detector

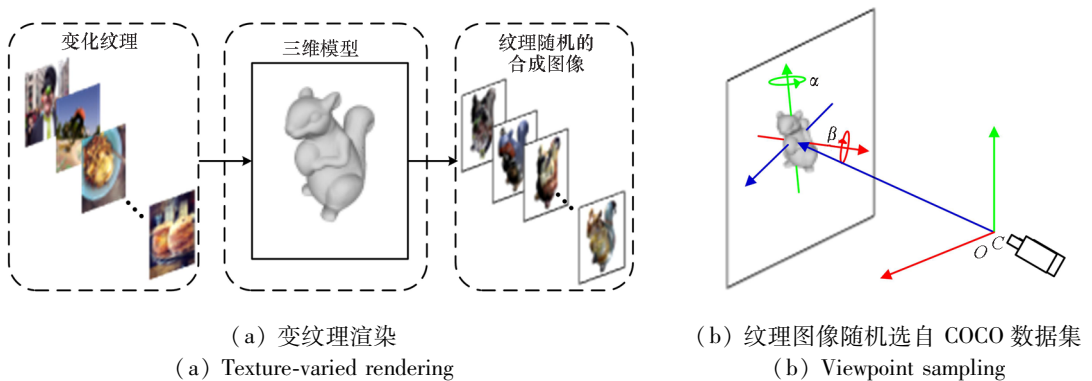


图 3 基于 Blender 的变纹理模型渲染及视角采样
Fig. 3 Blender based texture-varied 3D model rendering and viewpoint sampling

渲染过程中,同样保存每张仿真图片对应的深度图,用以生成精确的目标包络框与目标掩膜。

在仿真图像生成过程中,固定虚拟相机参数设置,每幅仿真图像中含一个目标。不同于文献[2, 4]仿真图像训练数据集生成中采用纯净背景的方式,本文对背景贴图 and 光照位置、能量进行了随机化处理。

3 实验结果与分析

3.1 实验设置

实验中目标模型为 RBOT 数据集^[23]中的“squirrel”模型,如图3所示进行视角采样,方位角 α 和高度角 β 的采样范围分别为 $[0^\circ, 360^\circ)$ 和 $[11.25^\circ, 78.75^\circ]$,采样步长分别为 22.5° 及 11.25° ,共得到96个视角采样,对应96个目标类别。从MS COCO数据集中随机选择 K 张不同的图片作为目标三维模型纹理,每个视角下得到 K 张包含不同纹理目标的仿真图像。背景图像从SUN数据集^[24]中随机选择,仿真图像尺寸为 640 像素 \times 512 像素。基于上述设置完成纹理随机化仿真图像数据集生成,记为 S^K 。

为测试本文方法性能,在仿真及真实图像数据上进行测试,并给出定量衡量指标。在配备4块1080Ti GPU的电脑上完成模型训练,批量大小设置为12,优化器为随机梯度下降(Stochastic Gradient Descent, SGD)算法,基础学习率设为0.01,迭代次数设置为50 000,在20 000和40 000时以比例因子0.1降低学习率,采用线性预热策略,其中预热长度为400步,优化器的权重衰减和冲量分别设为 5×10^{-4} 和0.9。此外,在训练过程中采用以下数据增强方法:①几何变换包括尺度变换、旋转变换和透视变换;②图像变换包括高斯模糊、运动模糊、限制对比度自适应直方图均衡(Contrast Limited Adaptive Histogram Equalization, CLAHE)、对比度调整、RGB值偏移与通道调换、随机遮挡。

3.2 评价指标

采用平均准确率(Average Precision, AP)定

量衡量检测器性能。AP被定义为不同召回率下的平均检测精度,通过计算P-R(precision-recall)曲线下的面积得到AP值越高代表检测性能越高。具体方法是,将检测器在测试数据集上的所有检测框按得分从高到低排序,根据设定的交并比(Intersection over Union, IoU)阈值判断每个检测结果是否正确,得到P-R曲线,然后计算P-R曲线下的面积。有多种P-R曲线下面积计算方法,VOC09^[25]将召回率平均分为11等份,每个召回率下的准确率为该召回率下的最大值。IoU阈值越高,对检测框的定位要求越高,相应的AP值越低。

$$AP = \frac{1}{11} \sum_{r \in \{0.0, 0.1, 0.2, \dots, 1.0\}} p_{\text{interp}}(r) \quad (1)$$

$$p_{\text{interp}}(r) = \max_{\tilde{r}: \tilde{r} > r} p(\tilde{r}) \quad (2)$$

3.3 仿真图像数据实验

在仿真数据集上进行算法性能测试。如前所述,每一视角下的目标为一个类别,仿真数据集中每一张仿真图像包含纹理及视角两个标签。设置 $K = 11\ 000$,分别对应单一纹理情况及本文的纹理随机化情况,对应仿真数据集分别为 S^1 及 $S^{1\ 000}$, $S^{1\ 000}$ 包含 9.6×10^4 张仿真图像。为保证可比性,在 S^1 数据集的生成中,对于设定的唯一纹理,同样在每一个采样视角下生成1 000张仿真图像,由于光照等因素的随机化设置,该1 000张仿真图像之间可能存在一定差异。以比例0.7对数据集 S^K ($K = 11\ 000$)中的视角和纹理标签进行训练集和测试集分割,得到4个互不相交的数据集,分别记为训练集(S_{train}^K)、纹理测试集(S_{testT}^K)、视角测试集(S_{testV}^K)和视角-纹理测试集(S_{testVT}^K),其图像数量分别为 4.69×10^4 , 2.01×10^4 , 2.03×10^4 和 0.87×10^4 。记在 S_{train}^K 训练的模型为 D^K ,图4给出 $D^{1\ 000}$ 模型在纹理测试集 S_{testT}^K 上的部分检测结果示例, S_{testT}^K 中的纹理未包含于训练集 S_{train}^K 中,且目标纹理存在明显差异,图中矩形框标识被成功检测的目标。本文方法都实现了可靠检测。

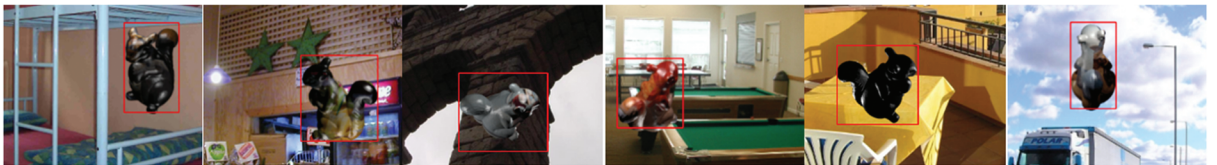


图4 $D^{1\ 000}$ 模型在 $S_{\text{testT}}^{1\ 000}$ 上的部分检测结果

Fig. 4 Partial detection results of $D^{1\ 000}$ model on $S_{\text{testT}}^{1\ 000}$

采用 AP 值定量衡量方法性能,测试结果如表 1 所示。在 $IoU \geq 0.9$ 的设置下, D^{1000} 在 S_{testT}^{1000} 上的 AP 值达到 96.9%, D^1 在 S_{testT}^{1000} 上的 AP 值也达到 74.5%, 这是因为在仿真图像生成中引入了光照等因素的随机化, 且在训练中引入了数据增强处理, 因此 D^1 也具备一定的纹理泛化性能。但本文方法通过纹理随机化处理, 有效降低了目标检测模型对纹理特征的偏移, 取得了更高的 AP 值。 D^{1000} 在 S_{testV}^{1000} 上取得了 79.6% 的 AP。 S_{testV}^{1000} 中的采样视角不同于 S_{train}^{1000} , 得益于如图 3 所示的视点采样策略, 本文模型对视角也具备较好的泛化能力。另外, D^{1000} 在 S_{testVT}^{1000} 上取得了与在 S_{testV}^{1000} 上相近的测试结果, S_{testVT}^{1000} 中的纹理未出现在 S_{testV}^{1000} 中, 这证明本文方法通过纹理随机化实现了结构主导的变纹理目标可靠检测。

表 1 D^1, D^{1000} 在仿真数据集上的目标检测平均准确率
Tab.1 Evaluation AP of D^1 and D^{1000} on simulation data
单位: %

模型	S_{testT}^{1000}	S_{testV}^{1000}	S_{testVT}^{1000}
D^1	74.5		60.0
D^{1000}	96.9	79.6	79.8

为进一步验证本文方法实现结构主导的可靠

目标检测, 抽取 S_{testV}^{1000} 和 S_{testVT}^{1000} 中目标的轮廓, 轮廓宽度设置为 3 个像素, 且随机赋值, 背景同样随机选自 MS COCO 数据集, 共得到 2.9×10^4 张仅包含目标轮廓信息的仿真图像, 记为仿真轮廓数据集 S_C^{1000} 。采用在 S_{train}^{1000} 上训练得到的目标检测模型 D^{1000} 对 S_C^{1000} 中的目标进行检测, 部分检测结果如图 5 所示, 图像中仅包含目标轮廓信息, 且目标内部区域与周围背景较为相似, 甚至一致, 这给目标检测带来极大挑战。本文方法可靠地完成目标检测, 在 $IoU \geq 0.5$ 的设置下, D^{1000} 取得了 45.8% 的 AP 值, D^1 对应的 AP 值为 9.7%。这说明通过纹理随机化, 本文方法有效降低了检测器的纹理偏移, 实现了可靠的结构主导目标检测。

3.4 真实图像数据实验

采用真实图像数据对本文方法性能进行测试, 目标检测模型完全基于纹理随机化仿真图像数据集训练得到, 目标实物依据目标三维模型打印得到, 并人工涂绘纹理, 与训练数据集中纹理存在明显差异。在纹理随机化仿真图像数据集 S^{1000} 上完成目标检测模型的训练, 在真实图像上进行目标检测。图 6 给出了部分目标检测结果示例, 图中包含多目标、不同纹理、复杂光照等情况。本文方法实现了可靠检测, 有效验证了方法性能。



图 5 D^{1000} 模型在 S_C^{1000} 数据集上的部分检测结果

Fig. 5 Partial detection results of D^{1000} model on S_C^{1000} dataset



图 6 真实图像数据目标检测部分结果

Fig. 6 Partial target detection results on real images data

为定量分析纹理随机化程度对目标检测模型性能的影响, 同前述 S^1 及 S^{1000} 仿真数据集的生成设置, 分别设置 $K=1, 10, 100, 1000$, 得到对应的仿真图像数据集, 每个仿真图像数据集包含 9.6×10^4 张图像, 并以此训练目标检测模型, 对搜集的 417 张真实图像中的目标进行检测。选择了

3 种具有代表性的目标检测模型, Faster RCNN, SSD^[26] 和 FoveaBox^[27], 分别代表两步检测器、一步检测器和无锚点检测器。在 $IoU > 0.5$ 的设置下, 图 7 展示了 3 种目标检测模型在不同纹理数量下的 AP 得分。首先, 随着纹理数量的增加, 3 种检测器的得分都得到了提高, 说明本文方法可

以作为一种有效的数据增强方法,提高不同类型的目标检测器对纹理随机目标的检测性能;然后,当纹理数量从 100 增加到 1 000 时,可以观察到 3 种目标检测器的 AP 得分几乎没有增长;最后,对比 3 种目标检测器的得分增幅,可知本文方法对基于锚点的两步检测器 Faster RCNN 的提升最大。

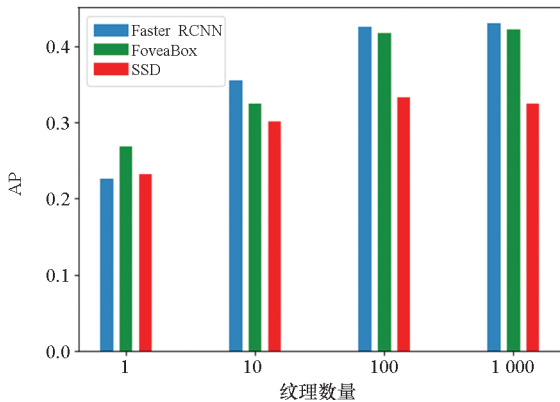


图7 纹理随机化程度对3种目标检测模型性能的影响

Fig. 7 Detectors performance increase with the number of unique texture.

仿真图像及真实图像上的测试结果验证了本文方法有效实现了变纹理目标的可靠检测。但在实验中也发现当输入图像中存在多个目标时可能会出现漏检。另外,在仿真图像上的检测结果优于真实图像,如表1及图7中所示,这是由于仿真图像与真实图像之间存在特性差异,仅使用仿真图像训练得到的目标检测模型推广到真实图像上依然存在不足。

4 结论

本文针对变纹理目标检测问题,提出一种新的基于纹理随机化的结构主导目标检测方法,利用目标三维模型,借助 Blender 渲染引擎,得到不同纹理的目标渲染,并随机化相机视角、光照等参数生成纹理随机化仿真数据集,并以此训练目标检测器,降低模型对纹理特征的偏移,实现结构主导的可靠变纹理目标检测。仿真及真实图像数据实验测试结果显示:本文方法可以有效地减弱网络模型对纹理的偏移,增强模型对目标结构特征的利用,实现变纹理目标的可靠检测。

本文方法仅使用纹理随机化的仿真图像数据完成目标检测模型的训练,依然受到仿真图像与真实图像之间差异的影响,在下一步工作中,可考虑引入域适应策略,提升该方法在真实图像上的

性能。此外,可考虑将纹理随机化处理思路推广至基于单帧图像的目标 6D 位姿估计中,探索目标结构主导的单目 6D 位姿估计。

参考文献 (References)

- [1] REDMON J, ANGELOVA A. Real-time grasp detection using convolutional neural networks [C]// Proceedings of IEEE International Conference on Robotics and Automation, 2015: 1316 - 1322.
- [2] SOMMER L W, SCHUCHERT T, BEYERER J. Fast deep vehicle detection in aerial images [C]// Proceedings of IEEE Winter Conference on Applications of Computer Vision, 2017: 311 - 319.
- [3] DALAL N, TRIGGS B. Histograms of oriented gradients for human detection [C]// Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005: 886 - 893.
- [4] AUBRY M, MATURANA D, EFROS A A, et al. Seeing 3D chairs: exemplar part-based 2D - 3D alignment using a large dataset of CAD models [C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2014: 3762 - 3769.
- [5] LIM J J, PIRSIIVASH H, TORRALBA A. Parsing IKEA objects: fine pose estimation [C]// Proceedings of IEEE International Conference on Computer Vision, 2013: 2992 - 2999.
- [6] LIM J J, KHOSL A, TORRALBA A. FPM: fine pose parts-based model with 3D CAD models [C]// Proceedings of European Conference on Computer Vision, 2014: 478 - 493.
- [7] HARIHARAN B, MALIK J, RAMANAN D. Discriminative decorrelation for clustering and classification [C]// Proceedings of European Conference on Computer Vision, 2012, 7575: 459 - 472.
- [8] REN S Q, HE K M, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137 - 1149.
- [9] HE K M, GKIOXAR G, DOLLÁR P, et al. Mask R-CNN [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(2): 386 - 397.
- [10] GEIRHOS R, RUBISCH P, MICHAELIS C, et al. ImageNet-trained CNNs are biased towards texture, increasing shape bias improves accuracy and robustness [C]// Proceedings of Seventh International Conference on Learning Representations, 2018.
- [11] ZAECH J, DAI D, HAHNER M, et al. Texture underfitting for domain adaptation [C]// Proceedings of IEEE Intelligent Transportation Systems Conference, 2019: 547 - 552.
- [12] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial networks [J]. Communications of the ACM, 2020, 63(11): 139 - 144.
- [13] CHOY C B, STARK M, CORBETT-DAVIES S, et al. Enriching object detection with 2D - 3D registration and continuous viewpoint estimation [C]// Proceedings of IEEE

- Conference on Computer Vision and Pattern Recognition, 2015: 2512–2520.
- [14] 章志勇, 潘志庚, 张明敏, 等. 基于多尺度通用傅里叶描述子的灰度图像检索[J]. 中国图象图形学报, 2005, 10(5): 611–615.
ZHANG Zhiyong, PAN Zhigeng, ZHANG Mingmin, et al. Multiscale generic Fourier descriptor for gray image retrieval[J]. Journal of Image and Graphics, 2005, 10(5): 611–615. (in Chinese)
- [15] MUKUNDAN R, ONG S H, LEE P A. Image analysis by Tchebichef moments [J]. IEEE Transactions on Image Processing, 2001, 10(9): 1357–1364.
- [16] 夏婷, 周卫平, 李松毅, 等. 一种新的 Pseudo-Zernike 矩的快速算法[J]. 电子学报, 2005, 33(7): 1295–1298.
XIA Ting, ZHOU Weiping, LI Songyi, et al. A novel algorithm for fast computation of Pseudo-Zernike moments[J]. Acta Electronica Sinica, 2005, 33(7): 1295–1298. (in Chinese)
- [17] HARALICK R M. Statistical and structural approaches to texture [J]. Proceedings of the IEEE, 1979, 67(5): 786–804.
- [18] WOHLHART P, LEPETIT V. Learning descriptors for object recognition and 3D pose estimation [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2015: 3109–3118.
- [19] RUSSAKOVSKY O, DENG J, SU H, et al. ImageNet large scale visual recognition challenge [J]. International Journal of Computer Vision, 2015, 115(3): 211–252.
- [20] TOBIN J, FONG R, RAY A, et al. Domain randomization for transferring deep neural networks from simulation to the real world [C]//Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems, 2017: 23–30.
- [21] REN X, LUO J, SOLOWJOW E, et al. Domain randomization for active pose estimation [C]//Proceedings of International Conference on Robotics and Automation, 2019: 7228–7234.
- [22] LIN T, MAIRE M, BELONGIE S, et al. Microsoft COCO: common objects in context [C]//Proceedings of European Conference on Computer Vision, 2014: 740–755.
- [23] TJADEN H, SCHWANECKE U, SCHÖMER E, et al. A region-based Gauss-Newton approach to real-time monocular multiple object tracking [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019, 41(8): 1797–1812.
- [24] XIAO J, HAYS J, EHINGER K A, et al. SUN database: Large-scale scene recognition from abbey to zoo [C]//Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2010: 3485–3492.
- [25] EVERINGHAM M, VAN GOOL L, WILLIAMS C K, et al. The pascal visual object classes (VOC) challenge [J]. International Journal of Computer Vision, 2010, 88(2): 303–338.
- [26] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot MultiBox detector [C]//Proceedings of European Conference on Computer Vision, 2016: 21–37.
- [27] KONG T, SUN F C, LIU H P, et al. FoveaBox: beyond anchor-based object detection [J]. IEEE Transactions on Image Processing, 2020, 29: 7389–7398.