

全局-实例特征对齐域适应检测方法及系统设计

刘源 娄亚鑫 张平 杨一帆 李亚伟 伍凌帆 张弘

Global-instance feature alignment domain adaptation detection method and system design

LIU Yuan, LOU Yaxin, ZHANG Ping, YANG Yifan, LI Yawei, WU Lingfan, ZHANG Hong

引用本文:

刘源, 娄亚鑫, 张平, 等. 全局-实例特征对齐域适应检测方法及系统设计[J]. 应用光学, 2024, 45(5): 946–955. DOI: 10.5768/JAO202445.0502002

LIU Yuan, LOU Yaxin, ZHANG Ping, et al. Global-instance feature alignment domain adaptation detection method and system design[J]. Journal of Applied Optics, 2024, 45(5): 946–955. DOI: 10.5768/JAO202445.0502002

在线阅读 View online: <https://doi.org/10.5768/JAO202445.0502002>

您可能感兴趣的其他文章

Articles you may be interested in

基于自适应像素级注意力模型的场景深度估计

Depth estimation based on adaptive pixel-level attention model

应用光学. 2020, 41(3): 490–499 <https://doi.org/10.5768/JAO202041.0302002>

基于改进二阶段检测网络的长时跟踪重检测方法

Re-detection method for long-term tracking based on improved two-stage detection networks

应用光学. 2023, 44(4): 768–776 <https://doi.org/10.5768/JAO202344.0402001>

基于全局与局部多尺度上下文的电表数据检测

Electric meter data detection based on global and local multi-scale context

应用光学. 2024, 45(4): 804–811 <https://doi.org/10.5768/JAO202445.0403006>

基于多尺度残差注意力网络的水下图像增强

Underwater image enhancement based on multiscale residual attention networks

应用光学. 2024, 45(1): 89–98 <https://doi.org/10.5768/JAO202445.0102003>

基于注意力残差编解码网络的动态场景图像去模糊

Image deblurring of dynamic scene based on attention residual CODEC network

应用光学. 2021, 42(4): 685–690 <https://doi.org/10.5768/JAO202142.0402008>

基于注意力机制与图卷积神经网络的单目红外图像深度估计

Depth estimation of monocular infrared images based on attention mechanism and graph convolutional neural network

应用光学. 2021, 42(1): 49–56 <https://doi.org/10.5768/JAO202142.0102001>



关注微信公众号, 获得更多资讯信息

文章编号: 1002-2082 (2024) 05-0946-10

全局-实例特征对齐域适应检测方法及系统设计

刘源¹, 娄亚鑫¹, 张平², 杨一帆¹, 李亚伟¹, 伍凌帆¹, 张弘¹

(1. 北京航空航天大学 宇航学院, 北京 102206; 2. 93129 部队, 北京 100036)

摘要: 在实际应用检测模型时, 由于真实场景和训练数据集间的差异, 检测算法的效果受到较大影响。为了在目标场景中获得更好的检测效果, 通常需要采集、标注数据后训练, 不仅成本高昂且流程复杂。提出基于注意力机制的全局-实例域适应检测算法与系统, 仅需采集部分真实场景数据即可进行迁移学习, 实现模型快速训练和边缘端-云端结合的远程部署。该域适应检测算法中, 基于注意力机制的全局特征对抗学习算法可减弱背景特征在迁移学习中的负作用; 基于字典学习的实例级特征对齐方法, 对实例级特征进行高精度对齐。经过实验对比, 本文的方法达到了接近 SOTA(state-of-the-art) 的水平, 并通过消融实验定量地证明了本方法对于域适应检测效果的提升。本文将提出的域适应检测技术与具有数据传输链路的边缘端系统结合, 在实际场景中使检测效果提升近 10 个点。

关键词: 域适应检测; 实例级; 注意力; 边缘端设备

中图分类号: TP391

文献标志码: A

DOI: 10.5768/JAO202445.0502002

Global-instance feature alignment domain adaptation detection method and system design

LIU Yuan¹, LOU Yaxin¹, ZHANG Ping², YANG Yifan¹, LI Yawei¹, WU Lingfan¹, ZHANG Hong¹

(1. School of Astronautics, Beihang University, Beijing 102206, China; 2. Unit 93129 of PLA, Beijing 100036, China)

Abstract: When actually applying the detection model, due to the difference between the real scene and the training data set, the effect of the detection algorithm is greatly affected. In order to obtain the better detection effect in the target scene, it is usually necessary to collect and label data and then train, which is not only costly but also complicated. The proposed global-instance domain adaptation detection algorithm and system based on the attention mechanism only needed to collect part of the real scene data to perform transfer learning, realizing rapid model training and remote deployment of edge-cloud integration. In this domain adaptation detection algorithm, the global feature adversarial learning algorithm based on the attention mechanism could reduce the negative effect of background features in transfer learning; the instance-level feature alignment method based on dictionary learning could align instance-level features with high precision. After experimental comparison, the proposed method reached a level close to SOTA(state-of-the-art), and the ablation experiment was quantitatively proved the improvement of the domain adaptation detection effect of this method. The proposed domain adaptation detection technology is combined with an edge system with data transmission links, improving the detection effect by nearly 10 points in actual scenarios.

Key words: domain adaptation detection; instance level; attention; edge device

收稿日期: 2023-09-28; 修回日期: 2023-10-23

基金项目: 国家自然科学基金 (62002005)

作者简介: 刘源 (1999—), 男, 硕士, 主要从事域适应检测与图像处理研究。E-mail: liuyuan8216@buaa.edu.cn

通信作者: 张弘 (1966—), 女, 教授, 主要从事模式识别与图像处理研究。E-mail: dmrzhang@buaa.edu.cn

引言

目标检测是计算机视觉中的一项热门任务,旨在从输入的视频或图像中获取感兴趣目标的信息^[1]。近年来,随着深度学习发展与硬件性能提升,针对通用场景数据集(如 MSCOCO 等)下的目标检测算法性能已经达到了较高的水平^[2]。然而,真实环境往往与现有的数据集存在较大差异,如光线、视角、天气等,导致源域(训练数据)与目标域(真实场景)存在域偏移,严重影响了检测算法在真实场景下的检测效果。

现有缓解域偏移技术途径分为两类。其一,使用少量标注目标域数据对模型进行微调,由于真实场景的实例标注是困难的,不具备实际可行性。其二,采用域适应算法改进的方式,主要包括基于特征对齐、半监督和数据增强的方法^[3]。CHEN Y 等人^[4]设计的 DAF(domain adaptive faster R-CNN)采用了基于对抗学习的深度域适应检测方法,首次提出了跨域目标检测任务;HE Z 等人^[5]在 DAF 的基础上提出了多层级特征对齐网络 MAF,后续被广泛使用;KIM T 等人^[6]提出的算法从多样性的角度学习多个域之间的特征不变性表示解决跨域检测问题;CHEN C 等人^[7]提出的 HTC(hierarchical transferability calibration network)算法将信息熵作为判别准则,对样本的迁移性进行筛选;VS V 等人^[8]从知识蒸馏的角度提出了一种均值教师网络用于域适应检测。目前从特征对齐角度出发,基于对抗学习的算法仍是域适应检测研究的主流。

然而,现有对抗学习算法仍然存在三方面问题。第一,目前大多数对抗学习方法没有考虑背景特征在迁移学习中的负作用。不加区分的对抗学习导致不对齐的背景特征也被强制对齐,使得前景与背景信息在检测中区分困难^[9]。第二,现有对抗学习算法难以对齐实例特征。实例特征具有丰富多变的特征,将实例特征利用对抗学习方法对齐会导致检测性能的下降。第三,现有方法少有对域适应检测系统的研究。如何利用边缘端设备采集到的图像,边缘端-云端结合的快速域适应训练与部署仍是尚未讨论的问题^[10]。

针对以上问题,本文提出了一种基于注意力机制的全局-实例域适应检测算法与系统,结合域适应算法与系统两方面的研究,实现快速训练、远程部署与较高精度的域适应检测系统。

一方面,提出了一种基于注意力机制的全局特

征对抗学习算法,能够减弱背景特征在迁移学习中的负作用。同时,还提出了基于字典学习的实例级特征对齐方法,对实例级特征进行高精度的对齐。实现更细致的特征迁移,从而提高检测模型的性能。

另一方面,提出了一种基于 FPGA+BPU(field programmable gate array+brain processing unit)单板多处理器协同架构的域适应模型可在线更新的检测系统,通过免标注实现快速训练,通过模型动态更新和远程部署,实现现场可重构域适应检测系统。总的来说,本文贡献主要有以下几个方面:

1) 提出了一种基于注意力机制的全局-实例域适应算法,使用注意力缓解全局特征对抗中背景知识负迁移的影响;通过基于字典学习的实例级特征对齐,能在实例级层面进行更加细致的对齐。

2) 提出了一种可实现免标注、场景数据回传、模型在线更新的域适应检测系统。将本方法应用到实际场景中,增强了在具体场景中的泛化性,无需标注,快速训练部署,并在端侧实现实时目标检测效果。

3) 所提方法在域适应检测数据集上进行了大量实验,实验结果表明,与现有方法相比,本文方法达到了较高水平。此外,还进行了消融实验,证明了本文所提各个模块的有效性。本文所提出的域适应模型可在线更新的检测系统,相比于无法更新的静态模型的检测系统具有更好的泛化性和场景适应能力,相比于先采集数据后标注训练的传统边缘端检测系统具有更快的响应速度和更好的检测结果。

1 域适应检测网络结构设计

域适应迁移学习的本质是将“源域”和“目标域”在特征空间中进行对齐。当“源域”和“目标域”在特征空间中的距离越接近时,就可以认为它们的分布差异不再存在,这意味着“源域”上的知识也适用于“目标域”,从而实现了迁移学习的目标。

1.1 基于注意力机制的全局-实例域适应检测算法

以 Faster R-CNN^[11] 为基准算法进行域适应检测的修改,如图 1 所示为本文算法的整体框架图。上下两条支路为参数共享的孪生网络,上方为源域训练支路,输入来自于源域,即有标注的数据集;下方为目标域训练支路,输入来自于目标域,即应用场景下的图片。输入通过骨干网络进行特征提

取,得到图像级特征图 F ,然后结合生成的注意力图,实现对前景区域的关注。将特征图通过一个梯度反转层,使用全局特征判别器 D_G 对特征图的来源域进行判断,通过梯度反转层与判别器之间的对抗学习,实现全局特征的对齐。经过RCN模

块后提取得到实例级特征 k ,来自源域和目标域的实例级特征共同组成了实例级特征字典,将检测头得到的分类结果作为伪标签,与相同类别的实例级特征进行靠近,不同类别的实例级特征进行疏远,达到在特征空间中精细化特征对齐的目的。

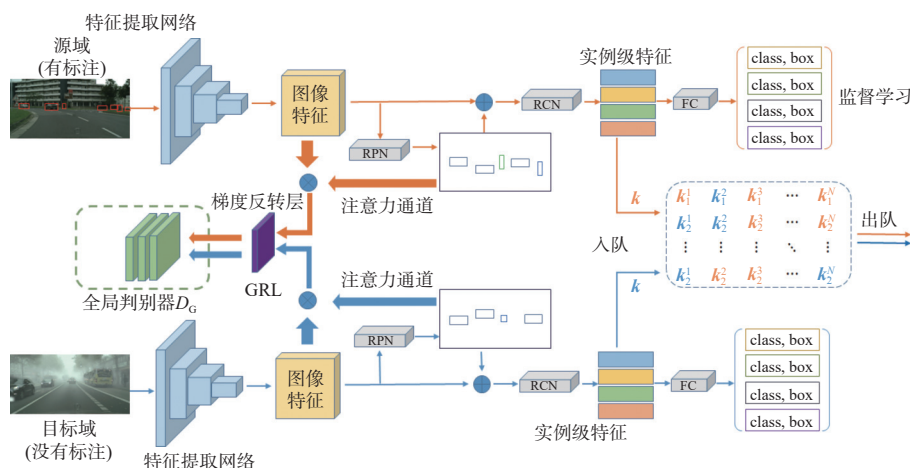


图1 基于注意力机制的全局-实例域适应检测算法框图

Fig. 1 Block diagram of global-instance domain adaptation detection algorithm based on attention mechanism

1.2 基于注意力机制的全局特征对抗学习

基于对抗学习的域适应检测方法利用域判别器进行对抗训练,在梯度反传时经过梯度翻转层使得源域和目标域特征尽量相似^[12]。实验证明,基于卷积网络的目标检测算法本质上是关注可能包含目标物体的局部区域^[13],基于这一前提,本文在全局特征的基础上加入了注意力模块,通过注意力挖掘生成包含目标的前景区域,再与全局特征结合,在前景区域上进行特征的域对齐。背景特征抑制的全局特征如图2所示,根据图片中的目标分布得出注意力图,将背景区域的图像级特征进行抑制。在抑制背景特征后,使用图像级特征对齐,源域和目标域前景特征的空间分布如图2右侧所示,将背景特征进行疏远,前景特征进行聚集。

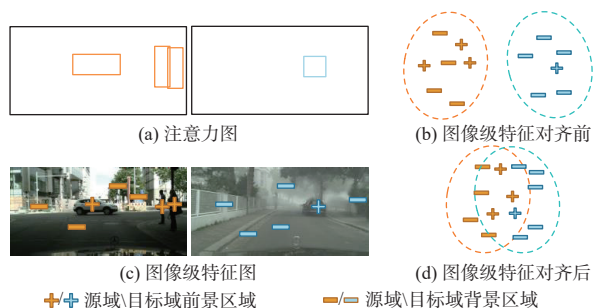


图2 背景特征抑制的全局特征对齐示意图

Fig. 2 Schematic diagram of global feature alignment for background feature suppression

域适应检测最早在分类任务中取得进展,DANN^[8](domain adversarial neural networks)网络提出将梯度反转层(gradient reversal layer, GRL)与卷积神经网络结合起来,图像分类网络与域分类网络共享特征提取网络参数,梯度反转层用于在梯度反向传播过程中实现梯度的取反。梯度反转层与域分类器在训练过程中对抗学习,使得输入域分类器的特征尽可能不具有领域的区分度。后面的研究者将梯度翻转层与域分类器应用到检测任务的迁移学习中,且取得了较好的效果。

若使用 I_s 代表源域中的图片, I_t 代表源域中的图片,经过特征提取网络 G 后得到图像级特征图 $F = G(I)$,在不加入注意力图时,域判别器的损失函数计算式(1)所示:

$$L_{GD} = -(Y_{DG} \log(D_G(F_s))) + (1 - Y_{DG}) \log(1 - D_G(F_t)) \quad (1)$$

式中: Y_{DG} 为特征图的分类标签,源域特征图的分类标签 Y_{DG} 为0,目标域特征图的分类标签是1。

整个对抗学习的原理是域判别器通过以上的损失函数极小化分类损失,而经过梯度反传层后的特征提取网络则通过极大化该损失来进行优化,当判别器无法判断特征图的来源时就达到了特征对齐的目的^[8]。

原有的将注意力图应用到域适应检测中的方

法有 EPM 网络^[14], EPM 的研究是在 FCOS 的网络结构上开展的, 利用了 FCOS 具有的目标提取层, 可以将特征图对应位置上的前景和背景区域进行分离。但是对于其他的检测结构来说, 难以直接获取注意力图。在基于锚框的检测网络中, 本文根据最终的检测结果作为伪标签生成注意力图。注意力图 AttenMap 的生成如下所示:

$$\text{AttenMap} = \sum_{i=0}^n N(x_{i_c}, w_i) \times N(y_{i_c}, h_i) \quad (2)$$

式中: n 为检测得到的框的总个数; N 为一维正态分布, 分别在检测框水平方向中心位置 x_c 和竖直方向中心位置 y_c , 以检测框的宽 w 和高 h 为方差构建高斯分布。将所有的检测结果分布求和, 得到最终注意力图。

1.3 基于字典学习的实例级特征对齐

源域和目标域不同类别的实例特征在空间分布如图 3 所示, 不同类别存在混叠现象, 容易导致类别的错分类等。全局特征对齐只能大致对齐源域和目标域特征, 无法进行更加细致的特征对齐。为解决特征对齐粗糙的问题, 本文将特征字典的方法引入到实例级特征对齐中, 进行更加细致的特征对齐。

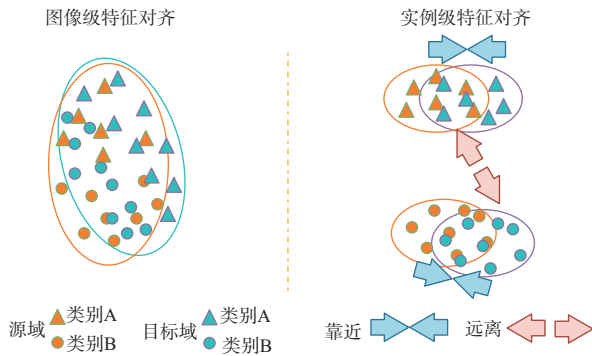


图 3 图像级特征对齐与实例级特征对齐示意图

Fig. 3 Schematic diagram of image-level feature alignment and instance-level feature alignment

本文将实例特征对齐视为字典查询操作, 构建了具有类别感知的特征字典, 其中包括源域和目标域的样本, 并赋予其类别伪标签, 使得实例级特征可以查询与类别字典中相同类别的特征相似度, 以及与不同类别之间的特征相似度。将相同类别的特征在空间中的表达尽量靠近, 不同类别特征的表达进行疏远^[15]。

特征字典使用队列数据结构, 采取先进先出的数据更新机制, 特征字典中储存着高维特征向量 k , 其中黄色标识的特征来自于源域, 蓝色标识来

自于目标域。源域和目标域的实例级特征组成混合特征字典, 由于数据集中不同类别实例个数相差较大, 所以针对这一问题, 本文提出了动态长度的特征字典 FD_C^N 。特征字典构建如图 4 所示。

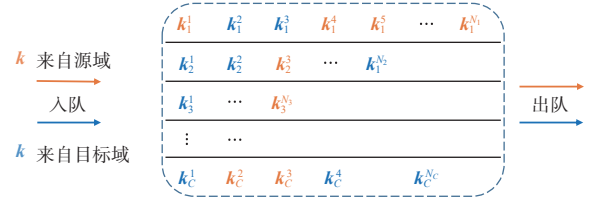


图 4 多种类变长度特征字典示意图

Fig. 4 Schematic diagram of multiple variable-length feature dictionaries

如图 4 所示, 其中 N 代表设定的每一类目标队列的最大长度, C 代表该数据集中目标的类别个数, k_c^n 代表第 c 类、第 n 个特征向量。考虑到检测性能以及算法实现上的可行性, 以域适应检测领域中常用检测效果较好的 Faster R-CNN 为基准模型。Faster R-CNN 经过 Align 模块得到特征向量 f_q , 以及每个特征向量所对应的属于每一类的概率 $\text{class}_{\log it}$ 。先将背景和低分特征向量进行滤除, 设置置信度阈值 $\text{score}_{\text{thresh}}$, 将前景目标的低分特征向量进行滤除, 得到高分前景目标特征向量索引值 fore_{idx} , 经过实验, $\text{score}_{\text{thresh}}$ 取 0.3 时最终的域适应检测效果最好, 此时能够正确分类且样本数量较多。

$$\text{fore}_{\text{idx}} \leftarrow \text{class}_{\log it} > \text{score}_{\text{thresh}} \quad (3)$$

把输入的特征向量送入一个特征编码网络, 并进行 L_2 正则化得到前景目标的字典特征 encoder_q :

$$\text{encoder}_q = L_2(\text{Encoder}(f_q[\text{fore}_{\text{idx}}])) \quad (4)$$

将 encoder_q 中的元素集合看作 X , 其中单个元素看作 x_i^c , c 是根据检测头得到的伪类别标签, i 是元素的索引值, 设元素总个数为 m , $i \in [0, 1, \dots, m-1]$ 。每个元素的特征字典损失计算如式(5)所示:

$$L_{\text{FD}}(x_i^c) = -\log_2 \left[\frac{\exp(x_i^c \cdot \text{FD}_c / \tau)}{\sum_{k \neq c} \exp(x_i^c \cdot \text{FD}_k / \tau) + \exp(x_i^c \cdot \text{FD}_c / \tau)} \right] \quad (5)$$

式中: τ 是温度系数, 温度系数决定了损失函数对困难负样本的关注程度, 本文取 $\tau = 0.07$ ^[15]; $x_i^c \cdot \text{FD}_c$ 中的 x_i^c 代表 encoder_q 第 i 个实例特征所属第 c 个类别, FD_c 代表特征字典中第 c 类的特征队列, “ $\cdot$ ”代表对两者进行相似度求解, 相似度范围是 $[-1, 1]$, 越接近 1 代表两者越相似, 越接近 -1 代表两者差

异越大。由于每次 encoder_q 中的元素个数不确定,所以对其中每个元素的损失函数进行取平均,以保证训练过程的稳定性,总的实例特征对比损失计算如式(6):

$$L_{\text{FD}}(X) = \frac{1}{m} \cdot \sum_{i=0}^{m-1} L_{\text{FD}}(x_i) \quad (6)$$

每次计算完成实例特征对比损失后都要使用 encoder_q 对特征字典 FD_C^N 进行特征元素的更新, encoder_q 元素所属目标根据分类头的伪标签进行分类,将对应类别的旧元素出列、新元素进列,维持实例特征字典元素的动态更新,保持特征一致性。训练过程中进行自动计数,动态调整各个类别字典的长度。

2 系统实现

嵌入式检测系统中在生活中发挥着越来越重要的作用^[16]。为满足实际场景需求,本文设计了如图5所示的系统框架,整个系统分为数据采集、图像回传、域适应模型训练和模型更新4个部分。视频采集阶段依赖端侧设备搭载的相机进行实现,将视频数据压缩后回传给服务器端,将视频帧作为目标域数据,利用服务器端进行域适应模型训练,将训练后的模型导出为板端支持的推理格式,并通过数据链路进行回传,更新端侧设备中的检测模型,然后继续执行相应的检测任务。

2.1 系统结构与

系统设计方面,本系统包括相机、FPGA、BPU

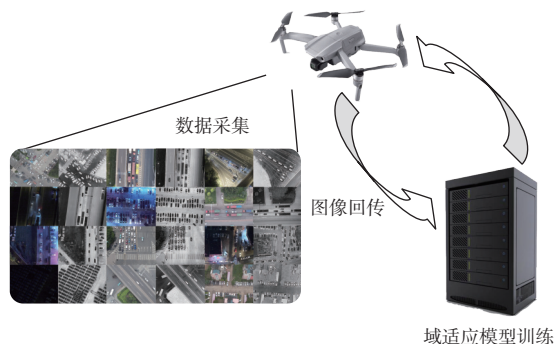


图 5 域适应检测系统设计结构图

Fig. 5 Structure diagram of domain adaptation detection system design

等。采用了FPGA+BPU单板多处理器设计,具有强大图像处理及神经网络推理能力,能够在边缘端进行图像实时处理和目标检测。本文结合域适应方法,设计了可在线更新的边缘端检测硬件系统,提出了实时在线更新的域适应目标检测方法。

利用平台以太网接口,结合无线通信实现系统与控制终端的通信链路,当硬件平台到达目标区域时,利用板端的相机对周围环境以及目标进行采集,将视频画面压缩之后通过数据链路回传到服务器端。服务器端对视频流进行处理,将采集到的画面作为目标域数据进行域适应检测训练,训练结束后将模型转换为板端支持的格式,再通过数据链路传输给硬件平台,实现模型的更新部署,并自动执行检测任务,以达到在目标场景中更好的检测效果。系统结构如图6所示。

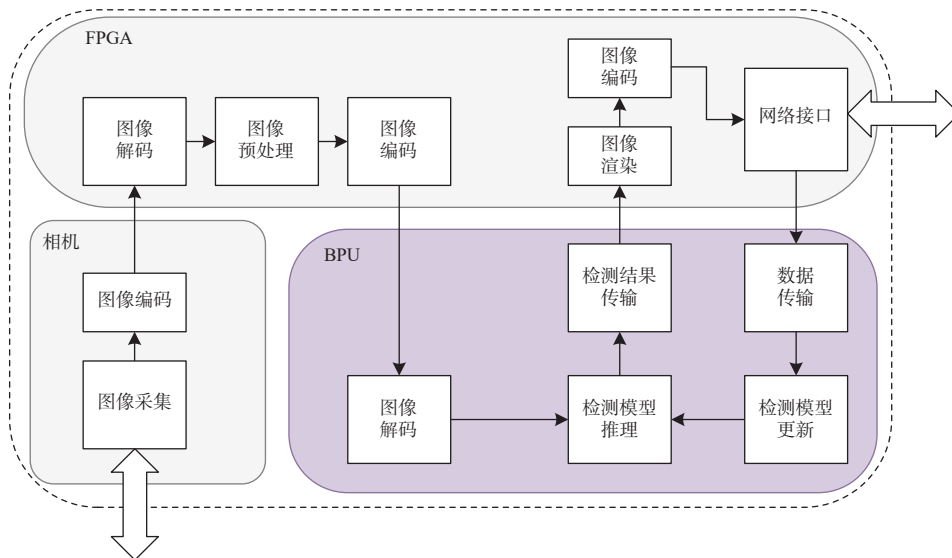


图 6 端侧系统设计结构图

Fig. 6 Structure diagram of embedded system design

与其他方法相比,本系统的优势在于:多处理器异构平台设计充分利用处理器优势,BPU对目标进行检测,FPGA负责图像渲染,达到了实时的检测效果;结合域适应检测的方法,本系统通过实时进行采集图像回传的方式进行模型的实时训练与在线更新,对于目标场景检测识别更加精准,效率更高。

2.2 端侧模型部署

本文以地平线 X3 国产化芯片为板端部署平台,其智能计算单元 BPU 具有 5 TOPs 的模型推理算力。由于板端芯片算力的限制,选择实时性较好的单阶段 YOLOv8 作为部署的检测算法。将本文提出的基于注意力机制的全局-实例级域适应方法应用到 YOLOv8 框架中,从 YOLOv8 的检测结果中可以得到注意力图的分布,从特征提取网络得到的注意力图中提取得到实例级特征,从而实现 YOLOv8 框架下的域适应检测。

在训练得到最终模型后,去除特征辅助对齐结构,将 YOLOv8 模型导出,配合 AI 工具链,经过浮点模型准备、模型验证、模型转换、性能评估和精度评估等操作,将模型转换成板端运行格式。

模型后处理部分包含很多算子,只能够在 CPU 上运行,这会导致量化时连续计算的部分被拆分为多个子图。多个子图计算会极大消耗处理器资源及存储器带宽,导致整体的推理速度极大降低。为此,我们从检测头的输出部分将模型进行划分,检测头之前的部分经过量化可以完全部署在 BPU 上进行推理,后处理部分则使用 CPU 进行解析,通过构建近似的激活函数等操作,节省推理时间。推理时间与推理效果的分析可以在实验与分析部分看到。

3 实验与分析

本章节详细描述使用的测试数据集、评价指标,本文方法的实验配置与超参数,以及与现有方法的性能比较,模块的消融实验与结果可视化。

3.1 实验数据集

域适应检测中经常使用 Cityscapes^[17]、Foggy Cityscapes^[18] 等数据集。Cityscapes 是一个在多个城市的街道场景以及季节采集得到的数据集,包括多个目标种类,针对语义分割、目标检测等任务都含有完整的标注信息。Foggy Cityscapes 也是一个城市场景的数据集,是在 Cityscapes 的基础上合

成的雾天的场景,标注信息与 Cityscapes 完全一致,Cityscapes 向 Foggy Cityscapes 的域适应检测是最常用的数据集。

3.2 实验评价指标

为了评估检测效果的精度,本文使用目标检测领域通用的平均精度(average precision, AP)作为模型性能评价指标。其中 mAP 指的是 IoU 在 0.50 到 0.95 区间每隔 0.05 时的 AP 取值的平均值。

3.3 实验配置与超参数

本文方法使用 PyTorch 1.9.0 框架实现,硬件平台为 Nvidia RTX A5000,PC 机操作系统为 Ubuntu 18.04。本文应用的训练策略包括:使用 SGD(stochastic gradient descent) 作为优化器来训练网络参数,训练时长为 30 epoch,初始学习率为 5×10^{-3} ,权重衰减为 0.000 1, batch size 设置为 8,使用阶段衰减的学习策略,分别在第 10 epoch、20 epoch 将学习率衰减至之前的 0.3 倍。本文使用标准的数据增强策略,包括随机翻转、随机裁剪、多尺度训练等。

3.4 性能比较

将本文所提出的基于注意力机制的全局-实例特征对齐域适应检测网络在多个数据集上进行了测试,其中以 Cityscapes 为源域、以 Foggy Cityscapes 为目标域进行多个类别的迁移学习,在目标域测试集上的检测性能如表 1 所示。

表 1 中 ConfMix 使用了图像裁剪拼接的方式,对图像级特征进行对齐,对齐的层次不够细致;SC-UDA 使用图像风格转换的方式进行域适应检测,由于数据集数量较少以及图像转换细节与真实存在差异等问题,导致仍存在域间隙;MeGA-CDA 使用了记忆模块对每个类别的实例特征进行动量更新,由于其学习的数据量较小,所以也无法实现较好的实例级域适应对齐效果;MGA 从多级特征对齐出发,将门控机制应用到对抗学习模块中,实现了多层级特征对齐;SSOD 从相似度的角度出发,通过源域和目标域特征聚类进行特征对齐,但是也会限制特征的表达。通过表 1 可以看出,本文所提方法在实例个数较多的类别上(行人、车辆)的检测指标达到了领先水平,在实例个数较少的类别上的检测指标表现一般,综合表现接近于 SOTA 水平。本文使用注意力的方式实现了实例特征与背景特征的区分,对实例特征进行图像级对齐可以减少背景特征负迁移带来的影

响,使用类别感知的动态实例级特征字典对源域和目标域实例级特征进行更加细致的特征对齐,

从而使得特征对齐更加全面和精细化,从而取得更好的跨域检测效果。

表1 本文的模型与其他域适应方法在 Cityscapes 向 Foggy Cityscapes 迁移性能对比

Table 1 Comparison of migration performance between proposed model and other domain adaptation methods from Cityscapes to Foggy Cityscapes dataset

方法	检测网络	Person	Rider	Car	Truck	Bus	Train	Motorcycle	Bicycle	mAP
Source only	Faster R-CNN	34.8	37.6	48.7	14.3	30.1	8.8	14.6	28.1	27.1
ConfMix ^[19]	YOLOv5	45.0	43.4	62.6	27.3	45.8	40.0	28.6	33.5	40.8
SC-UDA ^[20]	Faster R-CNN	38.5	43.7	56.0	27.1	43.8	29.7	31.2	39.5	38.7
MeGA-CDA ^[8]	Faster R-CNN	37.7	49.0	52.4	25.4	49.2	46.9	34.5	39.0	41.8
MGA ^[21]	Faster R-CNN	43.9	49.6	60.6	29.6	50.7	39.0	38.3	42.8	44.3
SSOD ^[22]	Faster R-CNN	38.8	45.9	57.2	29.9	50.2	51.9	31.9	40.9	43.3
Ours	Faster R-CNN	46.6	49.2	61.8	30.1	49.8	39.8	36.7	39.4	44.2
Oracle	Faster R-CNN	52.2	55.0	72.1	31.2	52.7	45.0	33.9	50.0	49.0

3.5 消融实验

为证明本文所提各个模块的有效性,并验证各个模块的提升作用,本文设计如表2所示的消融实验。

表2 在 Cityscapes 向 Foggy Cityscapes 数据集迁移学习的消融实验

Table 2 Ablation experiment of transferring learning from Cityscapes to Foggy Cityscapes dataset

方法	注意力(损失)	注意力(特征)	字典学习	mAP50	性能提升
Baseline				27.1	
Baseline	√			38.5	+11.4
Baseline		√		39.6	+12.5
Baseline			√	40.2	+13.1
Baseline		√	√	44.3	+17.2

从表2可以看出注意力机制和实例级特征字典在 Cityscapes 向 Foggy Cityscapes 迁移学习任务上表现出的有效性,且由于注意力机制与字典学习所用的区域不同,所以综合使用两者能够使检测性能再次得到提升。综合使用最多能够使检测效果相比于 baseline 提升 17.2,使得检测网络的检测效果在目标域有较大提升。

3.6 结果可视化

为了直观展示特征字典的作用,将训练过程中的实例特征使用非线性降维技术 t-SNE^[23] 进行降维可视化。对实例级特征对齐前后第1、10、20、30轮的实例特征进行可视化得到图7。

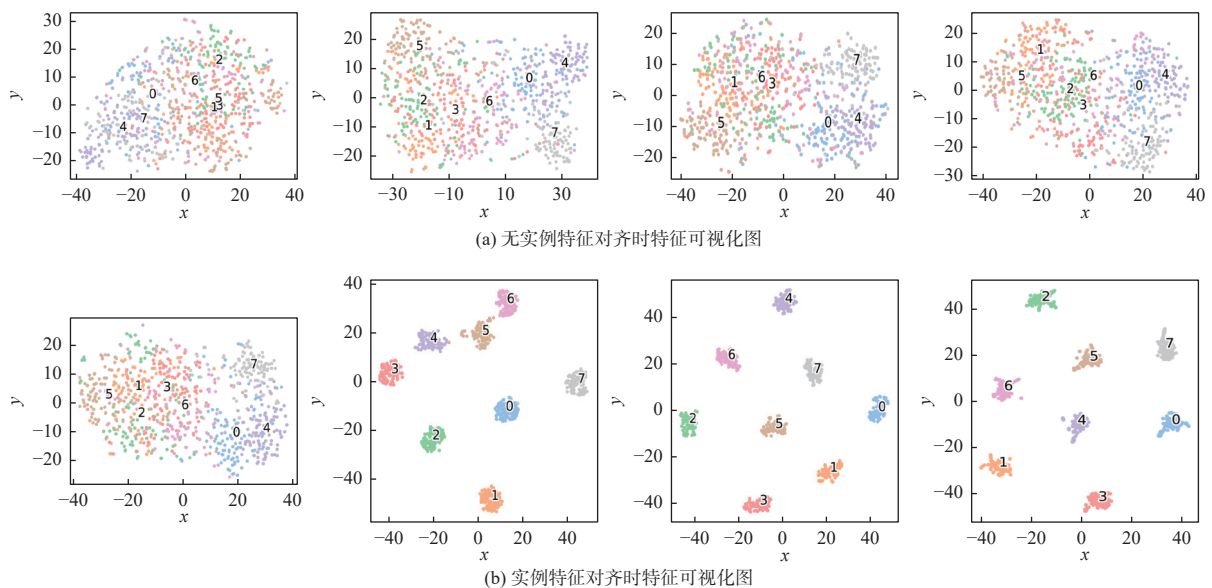


图7 实例级特征对齐前后第1、10、20、30轮的实例级特征可视化图

Fig. 7 Visualization diagram of instance-level features in epoch 1, 10, 20, and 30 without/with instance-level feature alignment

由于 Cityscapes 数据集具有 8 类目标, 从图 7(a) 中可以看出每类目标的实例特征分布较松散, 不同种类的特征之间混叠在一起。只在图像级对齐的策略下, 随着训练的进行, 不同类别的实例特征无法进行区分, 容易导致网络的错检。使用特征字典进行实例级特征对齐后, 随着训练的进行, 如

图 7(b) 所示, 不同种类之间的特征逐渐分离, 相同种类之间的特征逐渐聚集, 能起到在实例级特征进行对齐的作用。

在 Cityscapes 数据集向 Foggy Cityscapes 验证集进行迁移学习时, 使用域适应前后在 Foggy Cityscapes 上的检测效果对比图, 检测结果见图 8。

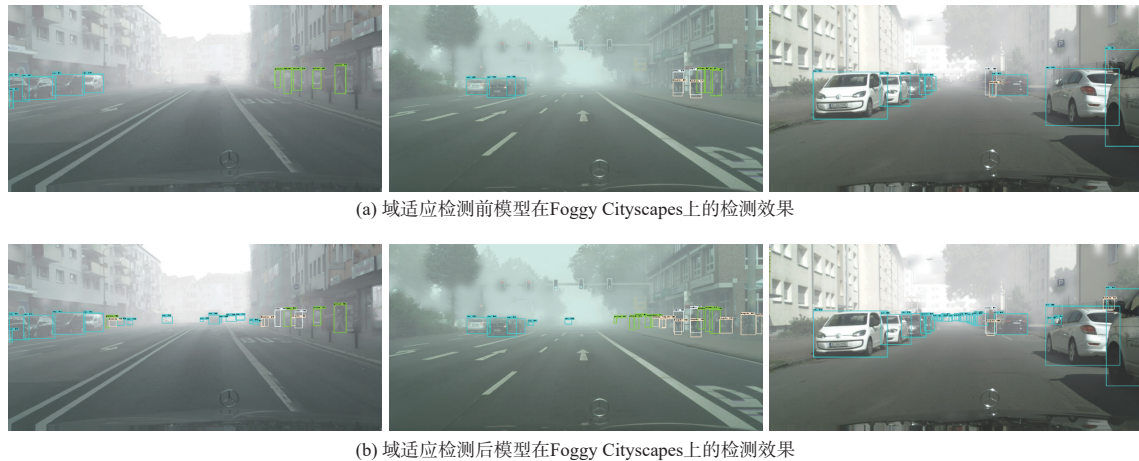


图 8 Cityscapes 训练的模型在 Foggy Cityscapes 数据集上的检测结果图

Fig. 8 Detection results of model trained by Cityscapes on Foggy Cityscapes dataset

从图 8(a) 可以看出, 未使用域适应训练得到的检测模型对于有雾遮挡的中小目标检出率较差, 大量远方目标无法被检出, 也体现了一般检测方法的局限性, 只能够在与训练数据集相似的场景中发挥较好的检测性能, 对于不同的检测场景, 检测性能下降较大。使用相同的网络结构加入域适应模块进行训练得到的模型, 在 Foggy Cityscapes 验证集上进行测试, 如图 8(b) 所示, 对于中小目标的检测效果有了很大的提升, 即使目标被雾所干扰, 网络也能够准确检出并正确分类, 可以看出本方法提出的域适应检测的有效性。

3.7 模型部署

以实时性较好的单阶段 YOLOv8 作为部署的检测算法, 结合本文所提方法, 在自有数据集上进行迁移学习的测试, 测试结果如表 3 所示。

表 3 在自有数据集上域适应前后的检测性能对比

Table 3 Comparison of detection performance before and after domain adaptation on self-owned datasets

方法	骨干网络	mAP
不进行迁移	C2fDarknet	36.5
本文方法	C2fDarknet	44.2
监督学习	C2fDarknet	46.3

从表 3 可以看出, 在不进行域适应检测的情况下, 仅利用监督学习的方法在源域上进行训练, 在目标域上进行测试的检测效果较差。在使用了本文的域适应检测方法后, 检测效果具有明显提升, 达到了接近监督学习的效果。

将转换之后的模型文件传输到板端进行推理, 推理一张 640×640 像素大小的图片, 在未去除后处理的情况下, 耗时达到 300 ms 左右。去掉检测后处理网络部分的模型部署到 BPU 上, 后处理部分使用 CPU 单独处理, 相比于直接将整个检测网络部署在 BPU 上推理大约能够快 4 倍, 达到 75 ms 左右。

为了进一步提高边缘端的推理速度, 我们参考 ZHANG Q 等人^[24] 的研究, 在网络中引入可变分组卷积 (VarGNet), 引入了可变分组卷积之后的网络模型精度、大小和边缘端推理时间, 如表 4 所示。

添加可变分组卷积之后, 编译器能够更好地针对 BPU 异构处理器进行优化, 网络中操作的计算强度更平衡, 存储器带宽也得到了更有效的利用。添加可变分组卷积之后, 虽然检测精度稍有下降, 域适应检测的精度 mAP 提升 9 个点, 但是推理时间得到了显著下降。因此实际测试中, 在边

缘端设备上的模型推理速度有了明显的上升,如表 4 所示,达到了实时的检测性能。

表 4 引入可变分组卷积前后网络的检测性能以及参数量
Table 4 Detection performance and number of parameters of network before and after using variable group convolution

方法	骨干网络	mAP	参数量/M	推理时间/ms
不进行迁移	C2fDarknet	31.5	3.2	74.2
本文方法	C2fDarknet	41.8	3.2	74.2
本文方法	C2fDarknet+VarGNet	40.5	2.5	23.4

为了验证域适应检测在实际场景中的检测效果,在采集得夜晚场景作为目标域进行训练,并将数据集视频输入边缘端设备进行推理,发现能够较好地检测到场景中的目标,最终域适应前后的检测结果如图 9 所示。



(a) 域适应前检测效果



(b) 域适应后检测效果

图 9 嵌入式端的检测效果

Fig. 9 Embedded board end detection effect

从图 9(a)可以看出,在白天场景训练的检测模型在夜晚场景的检测效果较差,对于弱特征、较小目标车辆容易出现漏检,在进行域适应训练后,得到如图 9(b)所示的检测效果,对于周围的小目标车辆都能够得到较好的检测结果。所以本方法所提出的全局-实例级域适应检测系统对于现实场景有着非常重要的应用价值和意义,可以显著提高在目标场景的检测效果。

4 结论

本文提出了一种基于注意力机制的全局-实例级域适应检测算法与系统,通过算法和系统两方面的研究,实现了一套具有快速训练、远程部署和较高精度的域适应检测算法与系统。本文将注意力机制应用到全局特征对齐中缓解了背景特征负迁移的影响,构建的特征字典使用相似度度量,实现了特征在实例级的精细化对齐。基于本文改进的检测网络在域适应检测数据集上,达到了接近 SOTA 的性能水平,并通过消融实验证明了本文方法的有效性。本文提出的免标注、场景数据回传、模型在线更新的域适应检测系统,将本方法应用到实际场景中,增强了在具体场景中的泛化性,且无需进行标注,可快速训练部署,并在嵌入式平台上实现实时目标检测效果。

参考文献:

- [1] 李小宁, 雷涛, 钟剑丹, 等. 基于改进 SSD 的车辆小目标检测方法[J]. 应用光学, 2020, 41(1): 150-155.
LI Xiaoning, LEI Tao, ZHONG Jiandan, et al. Detecting method of small vehicle targets based on improved SSD[J]. Journal of Applied Optics, 2020, 41(1): 150-155.
- [2] 廖延娜, 姚亮. 改进 YOLOX 的桥梁病害检测识别[J]. 应用光学, 2023, 44(4): 792-800.
LIAO Yanna, YAO Liang. Bridge disease detection and recognition based on improved YOLOX algorithm[J]. Journal of Applied Optics, 2023, 44(4): 792-800.
- [3] ZAIDI S S A, ANSARI M S, ASLAM A, et al. A survey of modern deep learning based object detection models[EB/OL]. (2021-05-12)[2023-10-10]. <http://arxiv.org/abs/2104.11892v1>.
- [4] CHEN Y, LI W, SAKARIDIS C, et al. Domain adaptive faster R-CNN for object detection in the wild[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. New York: IEEE, 2018: 3339-3348.
- [5] HE Z, ZHANG L. Multi-adversarial faster-rcnn for unrestricted object detection[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. New York: IEEE, 2019: 6668-6677.
- [6] KIM T, JEONG M, KIM S, et al. Diversify and match: a domain adaptive representation learning paradigm for object detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New

- York: IEEE, 2019: 12456-12465.
- [7] CHEN C, ZHENG Z, DING X, et al. Harmonizing transferability and discriminability for adapting object detectors[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2020: 8869-8878.
 - [8] VS V, GUPTA V, OZA P, et al. Mega-cda: Memory guided attention for category-aware unsupervised domain adaptive object detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2021: 4516-4526.
 - [9] XING W, ZHANG H, CHEN H, et al. Feature adaptation-based multipeak-redetection spatial-aware correlation filter for object tracking[J]. *Neurocomputing*, 2022, 488: 299-314.
 - [10] ZHANG H, LI Y, LIU H, et al. Learning response-consistent and background-suppressed correlation filters for real-time UAV tracking[J]. *Sensors*, 2023, 23(6): 2980-3001.
 - [11] 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 and Machine Intelligence*, 2015, 39(6): 1137-1149.
 - [12] ZHANG H, XING W, YANG Y, et al. SiamST: siamese network with spatio-temporal awareness for object tracking[J]. *Information Sciences*, 2023, 634: 122-139.
 - [13] GANIN Y, LEMPITSKY V. Unsupervised domain adaptation by backpropagation[C]//International conference on machine learning. [S. l.]: PMLR, 2015: 1180-1189.
 - [14] HSU C C, TSAI Y H, LIN Y Y, et al. Every pixel matters: center-aware feature alignment for domain adaptive object detector[C]//Computer Vision—ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part IX 16. Berlin: Springer International Publishing, 2020: 733-748.
 - [15] HE K, FAN H, WU Y, et al. Momentum contrast for unsupervised visual representation learning[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2020: 9729-9738.
 - [16] 娄树理, 王岩, 郭建勤, 等. 改进 YOLOX-S 的红外舰船目标检测算法[J]. *应用光学*, 2023, 44(5): 1054-1060.
 - LOU Shuli, WANG Yan, GUO Jianqin, et al. Infrared ship target detection algorithm based on improved YOLOX-S[J]. *Journal of Applied Optics*, 2023, 44(5): 1054-1060.
 - [17] CORDTS M, OMRAN M, RAMOS S, et al. The cityscapes dataset for semantic urban scene understanding[EB/OL]. [2023-10-10]. <http://download.visinf.tu-darmstadt.de/papers/2016-cvpr-cordts-cityscapes-preprint.pdf>.
 - [18] SAKARIDIS C, DAI D, GOOL L V. Semantic foggy scene understanding with synthetic data[EB/OL]. [2023-10-10]. <https://www.baidu.com/link?url=IX4kFbBqDNKzIJWpmt1YfKLcd0QtmaoMLIB3IEoRT2vaiaZjIYyd7CMRHwb9yjeM&wd=&eqid=c3bc4e3c00f50c840000000666cee097>.
 - [19] MATTOLIN G, ZANELLA L, RICCI E, et al. ConfMix: unsupervised domain adaptation for object detection via confidence-based mixing[C]//Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. New York: IEEE, 2023: 423-433.
 - [20] YU F X, WANG D, CHEN Y P, et al. Style and content gaps aware unsupervised domain adaptation for object detection[C]//Proceedings of IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). New York: IEEE, 2022: 1061-1070.
 - [21] ZHOU W Z, DU D W, ZHANG L B, et al. Multi-granularity alignment domain adaptation for object detection[C]//Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2022: 9581-9590.
 - [22] FARZANEH R, RAKSHITH S, RAHAF A, et al. Seeking similarities over differences: similarity-based domain alignment for adaptive object detection[C]// Proceedings of IEEE/CVF International Conference on Computer Vision. New York: IEEE, 2021: 9204-9213.
 - [23] VAN D M L, HINTON G. Visualizing data using t-SNE[J]. *Journal of Machine Learning Research*, 2008, 9(11): 2579-2605.
 - [24] ZHANG Q, LI J, YAO M, et al. Vargnet: variable group convolutional neural network for efficient embedded computing[EB/OL]. [2023-10-10]. <http://arxiv.org/pdf/1907.05653>.