天津大学

本科生毕业设计（论文）任务书

题目：复杂场景下的目标检测算法研究

学　　院　电气自动化与信息工程学院

专　　业　通信工程

年　　级　2018

姓　　名　高翔

学　　号　3018234335

指导教师　周圆

一、原始依据

随着信息技术的蓬勃发展，在互联网上每天都会产生海量的图片和视频数据，针对爆炸式递增的数据进行分析和处理关系到网络空间的安全保障以及用户体验的提升。而目标检测是计算机视觉和多媒体应用的基础技术之一，在实际的工程中有广泛的应用需求，具有非常重要的研究意义。

目标检测任务是在给定的图片中识别所有目标的类别，同时给出其坐标位置框，并使用一个外接矩形框来定位所识别目标的位置。由于同一张图片中可能存在多个目标，目标之间类别不一，尺度不同，位置也可能相互遮挡等，因此相比于一般的图像分类任务而言，针对多个目标的识别和定位具有更大的难度和挑战性。虽然近年来随着深度学习的发展，目标检测算法有了突破性进展，但是在面对复杂场景下的跨域目标检测、恶劣天气下的目标检测等实际问题上依然较多尚未解决的问题和挑战。因此，如何提高基于现有的深度学习算法来提高复杂场景下的目标检测的精读具有十分重要的研究价值。

实验室长期以来在计算机视觉，特别是基于深度学习的图像和视频处理领域进行过大量的研究，能够为课题的开展提供良好的实验环境。所在课题组长期致力于使用深度学习进行目标识别等研究工作，具有一定的经验积累。目前实验室已有的装置和各种实验条件，为本科毕业设计提供了良好的工作环境，促进项目的顺利进行。

二、参考文献

[1] Ganin Y, Lempitsky V. Unsupervised Domain Adaptation by Backpropagation[J]. 2014.

[2] Chen Y, Li W, Sakaridis C, et al. Domain Adaptive Faster R-CNN for Object Detection in the Wild[C]// 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE, 2018.

[3] Lecun Y, Bengio Y, Hinton G . Deep learning[J]. Nature, 2015, 521(7553):436.

[4] Papageorgiou C P, Oren M, Poggio T. General framework for object detection[C]// Computer Vision, 1998. Sixth International Conference on. IEEE, 1998.

[5] Viola. Robust Real-time Object Detection[J]. International Journal of Computer Vision, 2001, 57(2):87.

[6] Lowe D G. Distinctive Image Features from Scale-Invariant Keypoints[J]. International Journal of Computer Vision, 2004, 60(2):91-110.

[7] Dalal N , Triggs B . Histograms of Oriented Gradients for Human Detection[C]// IEEE Computer Society Conference on Computer Vision & Pattern Recognition.

IEEE, 2005.
[8] Felzenszwalb, Pedro, F, et al. Object Detection with Discriminatively Trained Part-Based Models.[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2010, 32(9):1627-1645.
[9] Everingham M, Eslami S, Gool L V, et al. The Pascal Visual Object Classes Challenge: A Retrospective[J]. International Journal of Computer Vision, 2015, 111(1):98-136.

三、设计（研究）内容和要求

本课题拟研究基于 Fast R-CNN 算法，探索通过改进卷积特征表达等方法提高目标检测在复杂场景下的识别效果，并考虑如何通过改善注意力机制有效提高模型的对图像上下文信息的提取，以及如何通过域自适应来实现域迁移后的目标识别精读。要求学生查阅和学习最新资料与经典文献，了解的目标检测的发展历程，学习基于深度学习的目标检测算法。具备 python 编程能力，熟悉 Pytorch 深度学习框架的应用，最后通过卷积神经网络实现复杂场景下目标识别的任务，并基于 PyTorch 实现且对比、评估方案性能。

指导教师（签字）

年　　月　　日

审题小组组长（签字）

年　　月　　日

天津大学

本科生毕业设计（论文）开题报告

题目：复杂场景下的目标检测算法研究

学　　院　电气自动化与信息工程学院

专　　业　通信工程

年　　级　2018

姓　　名　高翔

学　　号　3018234335

指导教师　周圆

一、课题的来源及意义

随着信息技术的蓬勃发展，互联网每天都会产生数以亿计的图片和视频，针对这些爆炸式递增的数据进行分析和处理关系到网络空间的安全保障以及用户体验的提高。目标检测使得计算机能够识别图像中的关键信息，是视觉理解的基础技术之一，具有非常重要的研究意义。目标检测任务是在给定图片中识别出所有目标的类别，同时给出其坐标的位置框，并利用一个外接矩形框来定位所识别目标的位置。目标检测技术在实际的工程中也有非常广泛的应用需求，通常会作为一种基础性的技术与其他任务结合发挥多种多样的功能，例如，智能安防监控，自动驾驶，图像视频检索，行人检测等等。

得益于深度学习[1]技术的发展，目标检测在最近几年取得了巨大的成功，一大批高效的检测算法被提出。现有的基于深度卷积神经网络（Convolution Neural Networks, CNN）的目标检测算法主要分为两大类:单阶段检测器与两阶段检测器。这些检测算法充分利用了神经网络来提取丰富的图像特征，极大的提高了目标检测的性能，但是由于互联网中的图像与视频数据的爆炸式增长，用户拍摄的图像往往具有更加复杂的场景，目标的尺度和遮挡情况对检测产生了更大的挑战。因此，针对这些复杂场景和类别多样的目标检测任务，提出泛化性能更好的检测算法是计算机视觉技术发展的迫切需求。

二、国内外发展状况

早期的目标检测算法采用精心设计的手工特征来进行模板匹配与识别，通过滑动窗口提取目标特征并编码，与标准模板进行匹配来识别目标。Papageorgiou 等人[2]提出了一种通用的目标检测框架，该框架提取 Haar-wavelets 特征，并利用支持向量机进行检测，直接从样本中学习特征，不需要任何的先验知识、模型或运动分割，在应对背景环境多变的人脸检测取得了较大的提升。Girshick 等人[3]是最早探索将卷积网络用于一般目标检测的人之一，并开发了 R-CNN 检测框架。现有的基于深度神经网络的目标检测方法可以大致分为两大类：两阶段法和单阶段法。（1）两阶段法又称基于区域推荐的方法，以 R-CNN 系列的算法为代表，在多个目标检测相关的任务中都取得了领先的检测精度。由于 R-CNN 需要对每一个推荐区域提取图像特征并分类，速度慢、优化困难一直困扰着该算法，因此 Girshick 又提出了 Fast R-CNN，大大提高了算法的处理速度；任少卿等在此基础上更进一步开发了 Faster R-CNN，将区域推荐过程也整合到检测模型中，实现了端到端的目标检测算法框架，速度又快了一个数量级。（2）单阶段法将分类和位置回归整合到一个前馈网络中，减少了区域推荐环节，网络结构更加简洁，计算量通常更小，在速度上相比两阶段法有较大优势。Sermanet 等人[4]提出的 OverFeat 被认为是第一个基于卷积深度网络的单级目标检测器，虽然具有一定的速度优势，却大幅度牺牲了检测精度。

Redmon 等人[5]开发了 YOLO 检测框架，将目标检测重新定义为一个回归问题。YOLO 的推理速度很快，但容易产生定位误差。刘伟等人[6]提出了单发多级目标检测器 SSD，利用不同尺度的特征图预测不同尺度的目标，极大的提高了定位精度。目前，对于复杂场景下的目标检测，尤其涉及到跨域检测，大多基于 R-CNN 系列算法进行改进，且研究尚处于起步阶段。

三、研究目标、研究内容与研究方法

通过对现有文献的调研，准确的识别并定位图片中的关键目标是一件非常具有挑战性的任务。因为这不仅需要考虑到待测图片的环境（光照，角度等）变化，还和目标本身的特性（刚性、尺度等）有关。为了更高效的识别和定位目标，本课题旨在研究目标检测在复杂场景的自适应性问题与跨域目标检测问题，着力寻找可以增强目标特征的方法，使增强后的特征可以更加关注相关目标，抑制不相关目标引起的干扰信息，有效地提高检测精度；此外研究如何对不在数据集下的真实场景进行有效目标检测；探索在场景时空变换下，待测目标特性变化时，如何提高目标检测的有效性。研究框架与方法包括但不限于：无监督学习、R-CNN 算法及其改进版本、域自适应的方法等，其中在目标检测方向的域自适应方法[7]包括，基于差异的域自适应、基于对抗的域自适应、基于重构的域自适应以及基于混合的域自适应。本课题基于 Fast R-CNN 算法方法进行改进，从而达到提高目标检测在复杂场景下的检测性能。

四、进度安排

（1）2021 年 12 月 1 日-2022 年 1 月 11 日，阅读相关文献，初步学习了解基于深度学习的目标检测算法和域迁移知识，撰写开题报告；

（2）2022 年 1 月 12 日-2022 年 2 月 18 日， 阅读国内外关于目标检测算法和域迁移的文献，学习总结基本原理和实现方法等，阅读深度学习在目标检测领域应用的文献，学习总结具体的建模思想和实验方案；

（3）2022 年 2 月 19 日-2022 年 3 月 1 日，根据已学知识，将基于深度学习的目标检测算法与域迁移等技术进行结合，尝试提出具体的实现方法与框架，以实现复杂场景下目标检测性能的提升，并与老师进行交流；

（4）2022 年 3 月 2 日-2022 年 4 月 2 日，进行实验仿真，在数据集上进行训练与测试，不断完善实现方案；

（5）2022 年 4 月 3 日-2022 年 5 月 10 日，整理实验数据，撰写毕业设计论文初稿；

（6）2022 年 5 月 11 日-2022 年 5 月 30 日并在老师指导下修改论文，准备后续的查重与答辩。

五、研究或实验方案的可行性分析

理论基础方面：近年来，目标检测方法都比较成熟，有很多适用的网络或模型。将深度学习和目标检测相结合的研究，也已有很多成熟的算法，如R-CNN 系列算法、YOLO、SSD 等。此外，近年来，在域迁移的目标检测领域也提出了较多可行的模型和方法，为进一步研究改进提供了基础理论依据。

实验条件方面：在目前流行的两个目标检测数据集 ASCAL VOC 和 MS COCO 上对所提出的方法进行了验证，这两个数据集分别包括 20 类和 80 类目标。此外，实验室具备支持研究的开发平台和硬件设备。

六、主要参考文献

[1]Lecun Y, Bengio Y, Hinton G . Deep learning[J]. Nature, 2015, 521(7553):436.

[2]Papageorgiou C P, Oren M, Poggio T. General framework for object detection[C]// Computer Vision, 1998. Sixth International Conference on. IEEE, 1998.

[3] Girshick R, Donahue J, Darrell T, et al. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation[C]// IEEE Computer Society. IEEE Computer Society, 2013.

[4] Neural D , Related S , Networks M , et al. OverFeat: Integrated Recognition, Localization and Detection using Convolutional Networks .

[5] Redmon J , Divvala S , Girshick R , et al. You Only Look Once: Unified, Real-Time Object Detection[J]. IEEE, 2016.

[6] Liu W , Anguelov D , Erhan D , et al. SSD: Single Shot MultiBox Detector[J]. Springer, Cham, 2016.

[7]Ganin Y, Lempitsky V. Unsupervised Domain Adaptation by Backpropagation [J]. 2014.

[8]Viola. Robust Real-time Object Detection[J]. International Journal of Computer Vision, 2001, 57(2):87.

[9]Lowe D G. Distinctive Image Features from Scale-Invariant Keypoints[J]. International Journal of Computer Vision, 2004, 60(2):91-110.

[10]Dalal N , Triggs B . Histograms of Oriented Gradients for Human Detection[C]// IEEE Computer Society Conference on Computer Vision & Pattern Recognition. IEEE, 2005.

[11]Felzenszwalb, Pedro, F, et al. Object Detection with Discriminatively Trained Part-Based Models.[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2010, 32(9):1627-1645.

[12]Everingham M, Eslami S, Gool L V, et al. The Pascal Visual Object Classes Challenge: A Retrospective[J]. International Journal of Computer Vision, 2015, 111(1):98-136.
[13]Chen Y, Li W, Sakaridis C, et al. Domain Adaptive Faster R-CNN for Object Detection in the Wild[C]// 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE, 2018.

选题是否合适： 是□ 否□

课题能否实现： 能□ 不能□

指导教师（签字）

年 月 日

选题是否合适： 是□ 否□

课题能否实现： 能□ 不能□

审题小组组长（签字）

年 月 日

天津大学

本科生毕业设计

题目：复杂场景下的目标检测算法研究

学　　院　电气自动化与信息工程学院

专　　业　通信工程

年　　级　2018 级

姓　　名　高翔

学　　号　3018234335

指导教师　周圆

独创性声明

本人声明：所呈交的毕业设计（论文），是本人在指导教师指导下，进行研究工作所取得的成果。除文中已经注明引用的内容外，本毕业设计（论文）中不包含任何他人已经发表或撰写过的研究成果。对本毕业设计（论文）所涉及的研究工作做出贡献的其他个人和集体，均已在论文中作了明确的说明。本毕业设计（论文）原创性声明的法律责任由本人承担。

论文作者签名：高翔

年　　月　　日

本人声明：本毕业设计（论文）是本人指导学生完成的研究成果，已经审阅过论文的全部内容。

论文指导教师签名：周圆

年　　月　　日

摘　要

通常，针对目标检测的任务需求，一贯假设需要训练的数据和被测试的数据源于相同的分布，但是实际并不总是如此。因此产生的分布的相互不匹配现象将导致模型的性能显著降低。本文介绍的域自适应 Fast R-CNN 模型以提升跨域问题的目标检测任务的鲁棒性为宗旨。本文为处理域偏移带来的性能降低问题在两方面提出改进：1）图像层级的偏移，例如光照、色彩等， 2）实例层级的偏移，例如检测目标的大小、形状等。基于当前最优秀的 Faster R-CNN 模型，本文在特征提取网络采用并加入了局部强对齐模型，对纹理和颜色等局部特征进行强对齐且不会改变类别级别的语义。此外，我们基于 H-散度理论采用并增添了图像层级和实例层级两个领域自适应组件，从而削减域偏移。组件为习得领域分类器均经由对抗训练的策略来达成目标。最后，经由一致性正则化组件深化两个层级的域分类器，从而习得具有领域不变性的原始模型中的区域提议网络。我们采用 Cityscapes 数据集和 Foggy Cityscapes 数据集评估此方法的可靠性和实用度。最终的结果证实了本文所采用的方法在跨领域的复杂场景中的目标检测任务的鲁棒性和有效性。

关键词：目标检测，域自适应，对抗训练，局部强对齐

ABSTRACT

Object detection usually assumes that the training and test data come from the same distribution, but this is not always the case. The mismatch of distributions will lead to significant performance degradation. The Domain-Adaptive Faster R-CNN model introduced in this paper aims to improve the robustness of cross-domain object detection. In this paper, domain shifts are addressed at two levels: 1) the image-level shifts, such as image style, lighting, etc., and 2) the instance-level shifts, such as object appearance, size, etc. Based on the current state-of-the-art Faster R-CNN model, two domain adaptation components, image level and instance level, are adopted and added in this paper to reduce the domain discrepancy. The two domain adaptation components are based on the H-divergence theory and are implemented by learning domain classifiers in an adversarial training manner. The domain classifiers at different levels are further enhanced by consistent regularization to learn the domain invariant region proposal network (RPN) in the Faster R-CNN model. In addition, we adopt a local strong alignment model in the feature extraction network to strongly align local features such as texture and color without changing the semantics at the category level. We evaluate this approach using the Cityscapes dataset and the Foggy Cityscapes dataset. The results demonstrate the effectiveness of our adopted method for robust object detection in domain transfer scenarios.

KEY WORDS: Object Detection, Domain Adaptation, Adversarial Training, Strong Local Feature Alignment

目 录

第一章　绪论

1.1　研究背景与意义

计算机领域中一个具有挑战性的研究方向是计算机视觉，旨在使计算机能够像人类一样提取和识别图像信息。随着计算机视觉的发展，自动驾驶、安全检测、人脸识别、医学诊断、交通监测等计算机视觉相关应用已经渗入到我们的生活中。而应用的背后是图像分割、定位和检测等视觉识别技术的成熟运用。伴随深度学习的井喷式发展，更多地研究者正在试图使用深度学习来提高计算机视觉任务算法的性能。基于深度学习的视觉算法与传统视觉算法相比在准确度上有明显提升，但为了达到这一目标，需要对大量地优质图像数据的进行采集和标注，从而来训练出泛化能力更好的深度学习模型。相应的，由于对视觉任务准确度的要求的逐渐提升，我们对标注准确、详尽的图像数据需求也在与日俱增，这也逐渐成为相关技术的一大瓶颈。

目标检测任务是计算机视觉领域中的本原性问题之一。它旨在鉴别并确定目标图像中某些需求类别的所有目标实例的准确位置。在卷积神经网络（CNN）[1]极速发展的助力下，许多基于卷积神经网络的目标检测模型被创造出来，故而极大地提高了目标检测的性能[2-4]。然而模型的训练依靠浩繁的带标注的图像数据。除此之外，与图像分类等任务相比，目标检测任务对于数据的标注要求更为严苛，除物体类别外，还需额外标注每一物体的边界框。又因重新标注一未知目标检测应用场景的数据集必然是耗时且代价昂贵的过程。另外，目前并无除人工作业外其他取得标注信息的途径。面临技术困境，无监督学习的应用使得不依靠人工标注信息成为可能，然则因其缺乏关键的监督信息，在实际应用中仍处于较低的性能表现，数据集的差异通常会降低他们对于新数据集的泛化能力。经调研，部分研究者尝试通过风格迁移将无标签和有标签的两种数据集进行风格转化来实现数据标注工作的替代操作。然而，这一过程需要训练两对判别和生成网络来实现数据集的转化，任务的计算复杂度显著提高，况且其转化质量很大程度地连带影响检测性能，未来的发展前景不明。

随着迁移学习的快速发展，域自适应也逐渐成为近几年兴起的研究方向，它与目标检测的结合一定程度上能够解决当前的问题。迁移学习定义为依靠两个或多个需学习的任务间的各要素的类似性，将旧领域中习得的知识迁移到新领域中并得以运用的一类学习进程。域自适应是通过在标注丰富的数据集上学习过的模型直接转移到未标注信息的数据集上，期望其在标注匮乏的数据集上依旧表现出

优异性能的迁移学习中的一类。此间，源域数据集代表标注丰富的数据集，目标域数据集是标注匮乏的数据集。在数据分布上两个域存在显著差异，也称为域偏移。域自适应的核心任务就是消除域偏移。由于不同数据集的光照、背景、角度、分辨率、复杂度等因素的不尽相同，域偏移普遍存在于其中，因而当把在源域上训练完成的目标检测模型运用于目标域时，目标检测的性能表现会大幅降低。且在实际场景中，误检、漏检现象频出，实际效果与在公开数据集上的表现相差甚远。域自适应为完成语义监督知识的迁移，将不包括域偏移的特征提取，从而让两个域中的特征映射于相同的特征空间。然而，针对复杂场景下的目标检测任务仍较少应用这一技术，主要仍用于图像分割、图像分类等任务需求。故此，在复杂场景中应用基于域自适应的跨域目标检测方法兼具研究价值和操作难度。

1.2 国内外研究现状与问题

本文主要研究内容是复杂场景下的目标检测算法，复杂场景特指跨域情景，算法主要基于域自适应技术。故此，本节主要介绍三部分内容，基于卷积神经网络的目标检测模型类别、域自适应在计算机视觉领域的发展状况以及基于域自适应的目标检测的研究现状。依据发展脉络进行梳理，并明确现在存在主要技术难关和瓶颈。

1.2.1 国内外研究现状

通常将目标检测的任务分为两类，一则在指定图片中锁定目标，二则辨别并标记目标种类。关键是要确定分类和定位对象的目标特征的信息，譬如纹理、外形和色彩等。继 2012 年 AlexNet 网络模型赢得 ImageNet Large Scale Visual Recognition Challenge 的桂冠后[1]，将卷积神经网络应用于视觉任务的方式大行其道。相比于传统目标检测利用人工设计的特征，即当目标发生变化时必须重新设计，且只能检测特定的目标，缺乏泛化能力，在深度学习大行其道的时期，由卷积神经网络构成的深度神经网络模型，以其强有力的泛化和特征提取能力，能够将目标检测从人工设计的特征的技术束缚中解放出来，并且持续提高目标检测算法的有效性。

当前，从网络架构分类，基于卷积神经网络的目标检测模型公认被划分为两大类：按照分类问题处理任务的双阶段目标检测算法以及按照回归问题处理的单阶段目标检测算法。双阶段目标检测算法通常以基于区域的卷积神经网络（Regionbased Convolutional Neural Networks, R-CNN）系列算法[5, 6]为代表性算法模型结构，单阶段目标检测算法则普遍以“只查看一次”（You Only Look Once, YOLO）系列算法[7-9]为代表性算法模型结构。按照“区域提议”方式来处理问题

的双阶段目标检测算法的第一次提出是 R-CNN[5]算法，它初次把深度学习运用于目标检测任务。相较此前的基础的检测算法，它开拓性地运用卷积神经网络这一工具来对需检测图像实行特征提取这一操作，从而使得检测性能表现极大上升。具体而言，是先把由选择性搜索算法生成的候选区域转化为一定大小的图片，其次将这些图片作为输入传入卷积神经网络，最后将提取出的特征经支持向量机进行分类。但是此方法存在耗时颇久、精确度低、重复计算等问题。之后 Fast R-CNN[10]应运而生，通过对全幅目标图片提取生成特征图，继而经感兴趣区域的池化操作将候选区域位置映射到全幅目标图片的特征图上，规避掉了候选区域特征重复计算的问题。此后，Ren 等人[6]还在此基础上创新性地加入了区域建议网络（Region Proposal Networking, RPN）来生成候选区域从而解决了耗时长、精度低的问题，这便是著名的 Faster R-CNN 模型。之后，具有代表性的 Cascade R-CNN[11]、特征金字塔网络（Feature Pyramid Networks, FPN）[12]等大部分双阶段算法均在 Ren 的这一版的 Faster R-CNN 根基上演化、发展而成。单阶段目标检测算法较于双阶段的目标检测算法，略去了独立提取候选区域的过程，在生成候选区域的同时，便对其进行分类和回归操作，从而提升了目标检测的速度。代表性的算法有 Redmon 等人[7]提出的 YOLO 系列算法、Liu 等人[4]提出的 SSD 算法、RetinaNet 网络[13]等。得有所失，伴随着检测速率的提高，精确度和检测的召回率都会有所降低。

域自适应问题普遍是指源域和目标域拥有着几乎一致的种类和特征，但具有不同的特征分布时，如何利用信息富裕的源域样本提高信息匮乏的目标域模型的性能[14]。起初，域自适应的研究多集中于图像分类、分割等领域。例如：DUANL 等人[15]提出的域迁移多核学习、KULIS 等人[16]提出的非对称度量、GOPALAN 等人[17]提出的子空间插值、SUN 等人[14]提出的协方差矩阵对齐等等。此外，Ghifary 等人[18]提出的域自适应神经网络（Domain Adaptive Neural Network, DANN）通常被认为是深度域自适应网络的开端。但因其设计较为简易，特征提取算力较弱，域偏移未得到很好解决。文献[19]中提出深度混淆网络（Deep Domain Confusion, DDC）使用 AlexNet 来提高精度和特征提取能力。此后，Long 等人[20]又提出了应用多核最大平均差异（Multi-Kernel Maximum Mean Discrepancy, MK MMD）的深度自适应网络（Deep Adaptation Networks, DAN）来改良 DDC 中最大平均差异度量表现能力弱的情况，进一步缩减了域间的差异性。受到生成对抗网络（Generative Adversarial Net, GAN）[21]启示，Ganin 等人[22]提出了利用域间对抗训练的梯度反向传播层（Gradient Reversal Layer, GRL）来降低域间差别。Yu 等人[23]在跨域的语义分割中运用了自训练策略和注意力判别器机制，在特征对齐上取得了优异的结果。括而言之，域自适应在图像分类和分割中的应用大多将着力点放在全幅图片的语义信息上，测算域间图片特征分布的距离，进而努力

实现距离最小化以缩减域偏移。然而，目标检测更侧重研究实例级别的语义信息而非全幅图片，即只对感兴趣目标区域进行语义分析和域偏移的缩减。从难度级别上分析，实例级的域自适应更胜一筹。

需求的敦促使得域自适应这项技术越来越多地应用于目标检测。当前，按主流的研究思路可以分为三大类，一是基于重建的域自适应目标检测算法，二是基于对抗学习的域自适应目标检测算法，三是混杂算法方案。

基于重建的算法主要依赖通过重建两个域的数据时，利用循环生成对抗网络（CycleGAN）[24]生成数据，获取的过程结果进行训练从而达到提高域迁移的学习表现。Arruda 等人[25]利用 CycleGAN 将日间场景数据集转化为夜间场景数据集，将模型在合成的数据集上训练来解决日夜交替下的跨域检测域自适应问题。Lin 等人[26]在此基础上经由多域生成众多结构一致的转化图片改进了合成的数据集，达到域自适应的目的。此类算法更多借鉴了风格迁移的方法，卓有成效地提高了性能，却很难构建端到端训练的检测器，其准确率也很大程度上取决于风格迁移的效果。

近年来基于对抗的算法逐渐崭露头角。Chen 等人[27]提出的 DAF（Domain adaptive Faster R-CNN）网络分别加入了图片级别和实例级别的域判别器来减少图片级别和实例级别的域偏移。还添加一致性正则化模块来使得特征仍具有域不变性。Zhu 等人[28]从自适应的关键任务出发来提高网络的鲁棒性，提出了选择性交叉域对齐网（Selective Cross-Domain AlignmentNet, SDA）来发现需要对齐的区域以及判断如何去对齐，再应用权重估计器来决定不同区域的权重。Saito 等人[29]在 DAF 基础上贡献了弱对齐模型，将对抗性对齐的损失集中于全局相似的图像上，而弱化对齐全局不同的图像，提出了 SWF（Strong-weak Faster R-CNN）网络。Yu 等人[30]提出了通过对图像级类别正则化的方式来取得核心分类信息，为测算难分样本还提出了分类一致性正则化方法，从而取得良好效果。

基于混杂的算法结合了包括以上两种算法的多种机制以期望达到域自适应需求。Rodriguez 等人[31]将基于重建和软标签训练相结合提出了两阶段域自适应网络，对于底层特征进行风格迁移，而对于高层特征进行软标签训练，从而实现域自适应。

1.2.2 存在的问题与不足

在卷积神经网络的推动下，在检测的准确度和速率上，目标检测都有了迅猛地发展，且在基准数据集[32, 33]上取得了优异的性能，这是因为通常预设所训练的数据集与被测试数据集保持独立同分布。然而，在现实世界中，这种预设在绝大多数情况下无法被满足，目标检测仍旧面临着由于照明、视角、图像分辨率、目标外观、背景环境等因素的巨大分布差异所带来的挑战，这些都会导致相当大的

所训练数据集与被测试数据集之间的域偏移，进而使得检测器的检测准确度下降。

譬如，无人驾驶场景中，特定汽车所使用的摄像头品类和预设参数很大概率上与收集训练数据所用的摄像头的品类和参数设置不尽相同，另外由于汽车一直处于移动状态，都市场景在不断变换，目标物体的外观将会改变。再者，系统期望在不同的气象要求中（如：雨雪雾天气）稳定地作业，然而，与之矛盾的是绝大多数训练数据是在可见度良好、空气干燥的晴朗日间收集所得。

因而，由于视觉与现实世界的不匹配，这对最近通常使用人工合成的数据集来进行训练深度卷积神经网络任务的趋势提出了较大的挑战。已进行的实验表明这种域偏移会导致性能的显著下降[17]。尽管收集更多的训练数据会缩减域偏移的影响，但标注边界框的过程耗时且代价昂贵。通过调研，应用域自适应来使目标检测模型自主适应与训练域在视觉上不同的新域或许是一个良好的解决办法。但由于感兴趣区域自身与原区域存在较大偏差，在仅有源域给予标签信息的条件下，也很难消除域间域偏移，这使得跨域目标检测的实际表现不佳。此外，误检、漏检、定位不准的问题也会因此增多。再者，对于哪些区域进行何种级别的特征对齐也是影响域偏移的一大难题。因此，基于域自适应的目标检测算法模型如何更好地缩减域偏移以及保持提取特征的域不变性是目前跨域目标检测所面临的关键性改进困境和难题。

1.3 本文的主要研究内容和组织结构

本文主要研究了基于域自适应的跨域目标检测 Faster R-CNN 算法，具体而言是指：采用基于目前最经典的 Faster R-CNN 模型[6]构建的输入图片后的神经网络的输出为判断指令的深度学习网络模型。本文采用了两个层级的领域自适应组件分别来减缓图像层面的和实例层面的域偏移问题，还进一步采用了一致性正则化组件来巩固这一方法。此外，还采用并集成了局部强对齐模块到 Faster R-CNN 模型中。

本文根据研究进程和思路划分章节，各章节组织结构安排如下：

第一章：绪论，首先对复杂场景下的目标识别算法的研究背景进行了概述，主要介绍基于域自适应的跨域目标检测算法的发展历程和各时期的理论应用和意义，阐明当前国内外基于域自适应的目标检测算法的特征和问题缺陷。

第二章：相关技术，主要介绍了目标检测算法相关的基础技术和发展过程中应用到的技术手段。从深度卷积神经网络的历史、组成结构，到两类经典的目标检测算法模型框架。之后，详细介绍本文所用基线模型——双阶段目标检测算法中的 Faster R-CNN 的主要组件。其次，介绍本文所主要研究方向——跨域目标

检测的关键性技术：领域自适应。最后，还针对域差异的量化问题阐述了衡量指标——H 散度。

第三章：基于域自适应的目标检测 Faster R-CNN 算法，先是主要介绍了针对目标检测任务的域自适应的可实现性和不同级别的自适应。之后对本文所采用的基于域自适应的复杂场景要求下的目标检测算法的基本原理和四个组件进行了梳理和理论介绍。最后，对于模型的损失函数进行了总结。

第四章：实验设计与结果分析，首先，主要介绍了本实验所采用的 Cityscapes 和 Foggy Cityscapes 实验数据集相关信息；之后，介绍了实验开发环境和相应的评价指标 mAP；最后，展示了实验结果的同时，还对实验的结果进行了细致地分析，对部分组件的作用也进行了验证分析。

第五章：总结与展望，归纳了本文在复杂目标检测，特别是跨域目标检测领域所做的任务。阐述了当前此领域仍旧困扰的难题，且对未来的研究可能性作出展望。

第二章　相关技术

本文研究方向是复杂场景下的目标检测算法，主要研究内容是基于域自适应的双阶段跨域目标检测算法，该算法将经典 Faster R-CNN 目标检测算法当作根基框架。因而将本章分为三部分进行相关理论介绍。2.1 节主要介绍深度卷积神经网络的基本原理，逐层介绍其内部原理；2.2 节主要介绍以深度学习网络为基础的两类经典目标检测算法模型，双阶段目标检测经典模型之一的 Faster R-CNN 的重点详细介绍则放在了第三章；2.3 节分别介绍域自适应和使用 H 散度的分布对齐的理论知识。

2.1　深度卷积神经网络基本原理

2.1.1　神经网络

二十世纪四十年代，有关生物神经元的递质和突触的研究成果启迪了 Mcculloch 和 Pitts[34]来运用学科交叉的思维衍生提出了“M-P 神经元”模型，从而为神经网络奠定了初代的理论模型，可以看作是开创第一代人工神经网络的先河。五十年代末，F • Rosenblatt[35]设计出了“感知机器”，具体而言是一个包含众多层级的神经网络，第一次把人工神经网络的研究从理论分析转化为工程应用，能够切实有效地处理数据，算是神经网络的第一次发展高潮。图 2-1 展示了通俗化的简易神经网络层结构，可见同一层级神经元相互不连接，上层级与下一层级间每一个神经元均需要相互衔接。

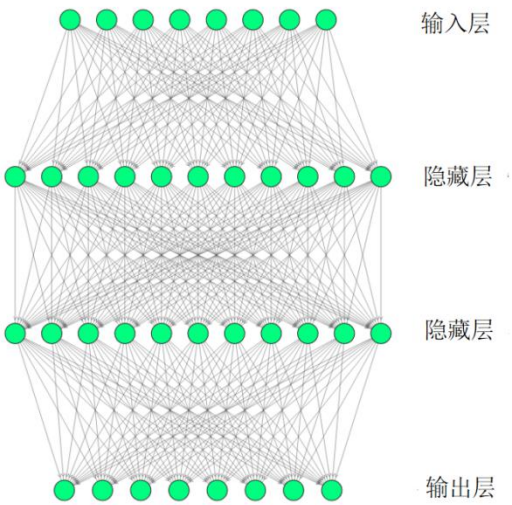

图 2-1　简易神经网络层结构

第一代神经网络仅能对形状简单的图形进行分类操作，是期望计算机达到类人去感知、识别、学习。然而，其感知机中特征提取的参数依靠人工调整这一做法限制了它的发展，且单层结构过于简单，许多知识都超出其学习能力。第一个实际可应用的卷积神经网络是 1998 年 Lecun 等人[36]的 LetNet-5 CNN，由于该网络对于手写阿拉伯字母的良好识别，网络的基本架构被应用改进至今。

第二代人工神经网络的开端是 1985 年，Hinton 使用 Back-Propagation 算法[37]去计算网络参数，并用复合隐藏层替代感知机中的单层结构。1989 年，LeCun 等人[38]用深度神经网络实现了计算机识别手写字母的任务，且识别率达到商业应用级别。但仍存在训练数据过慢的问题。21 世纪初期，随着 FPGA、GPU 等期间的升级迭代，这类具备多核并行计算能力的硬件设备常被用于高性能计算，从而大幅缩减深度学习的训练时间，为深度卷积神经网络的快速发展打下了可持续发展的硬件基础。2010 年，美国国防部高级研究计划局对深度学习领域的开发提出了初次的赞助。2011 年，谷歌的语言研究团队和微软研究院研究员选取多层感知机技术将语音的错误识别概率较传统方法下降至 20%-30%。2012 年，Krizhevsky 等人[1]提出的 AlexNet 网络以极大的优势击败了其它非神经网络的算法，并凭借远超人类预测能力得到表现在 ImageNet[39]大规模视觉识别挑战赛上拔得头筹。他们在 ImageNet 图片分类问题上把概率向量从大到小排列的头五名中包含正确概率即为正确预测的错误率降至 15%，改进一半有余。在这一创举之前，神经网络这一概念一直处于不被工业界认可的状态。此后，深度卷积神经网络的发展便如潮涌之势，迭代出一大批性能极佳的神经网络模型，例如：Google 的人脸识别系统 FaceNet[40]、VGGNet[41]、ResNet[42]等等。深度卷积神经网络模型凭借其应用广泛的特性在各个领域大放异彩。相较于国外的领跑，国内对此领域的研究也在不断加速，华为先声夺人率先建设的“诺亚方舟实验室”、百度亦步亦趋设立的“深度学习研究院”（IDL）、由腾讯开始创设的 Mariana、阿里的 DTPAI 人工智能平台都致力于对此领域深耕细作。

深度卷积神经网络相较于其他网络被公认为图像学习内容的最适配技术之一，当前多数数字图像处理领域任务的竞赛桂冠都是基于深度卷积神经网络架构的思路而建构出的模型方法，在图像分类[1]、目标检测[5, 9]、分割[43, 44]等任务中体现出优异的性能。这样的特性很大程度上依赖于其任一单次运算都涵盖卷积范围内的多维输入数据，因而最大化利用了数据的多维度空间信息。通常，经典的卷积神经网络架构涵盖以下数个模块：依次迭变相连的卷积层和池化层、其次是激活函数、末尾一般是数个全连接层或全局平均池化层来进行收尾，图 2-2 便展示了普适的卷积神经网络架构。为进一步提高网络性能，有时还会加入正则化单

元，但本节只针对基本组件进行阐述。具体而言，为提炼得到目标图像的高层次的语义特征，经卷积层提取出的特征需要再次经过池化层进行降采样操作并防过度拟合，再通过激活函数增加非线性特性，之后在全连接层对经此前步骤训练习得的语义特征一一对应到相应的样本空间，最后提取出的特征作为输入在目标函数中进行损失计算，再经反向传播算法[37]回传迭代网络参数。

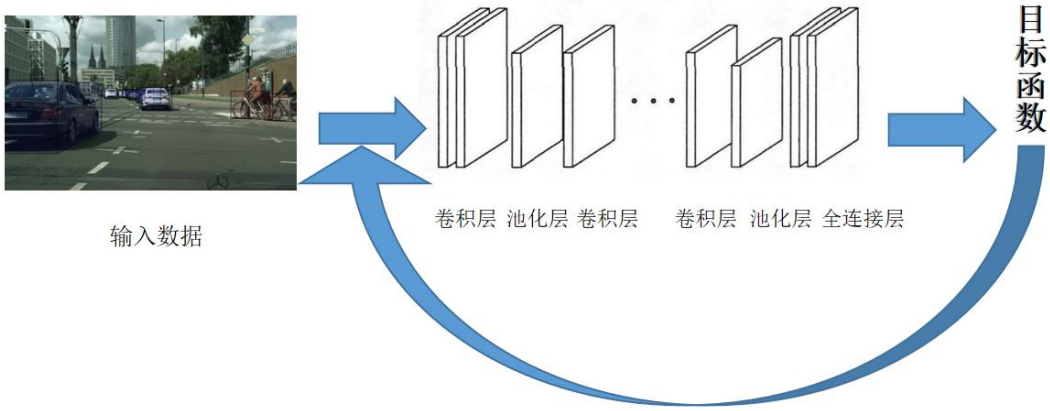

图 2-2　典型的卷积神经网络结构

本文所主要研究的工作基于域自适应的 Faster R-CNN 目标检测算法也正是以深度卷积神经网络为基础所衍生出来的。

2.1.2　卷积层

卷积层（Convolutional layer），卷积操作从通俗的数术操作上解释即被输入矩阵和作为卷积核的矩阵实行相应的元素乘积操作后再相加求和的数学运算。在卷积神经网络中，卷积核一般被当作卷积运算中的最小单元，每层卷积层在数个卷积核的堆砌下构成，卷积单元的数量通常称作通道数或深度。此外，该模块还运用逆向传播算法使得每一卷积核中的运算参数都达到优化态，且所得参数量与卷积单位尺寸相同。卷积操作从计算机专业角度分析可以看作是对图像的二维离散卷积操作，是数字图像处理任务中的滤波方法的应用。卷积运算的主要作用是将输入图片的各类特征提取出来，且作用于局部图像，公式如 2-1 所示。单一层的卷积层因其提取能力限制仅能将一些如线条、边界等底层特征提炼而出，越多层的结构可以逐渐优化出更高层的特征。

$$I(x,y) * \quad k(s,t) = \sum_{i=0}^{m} \sum_{t=0}^{n} k(s,t) I(x-s, y-t) \tag{2-1}$$

其中：k（s, t）代表卷积核所在坐标(s,t)处的参数值，I（x-s, y-t）代表输入图像中与卷积核相乘的像素值，公式所得为卷积操作后在（s, t）处的像素值，I（x-s, y-t）代表卷积之前的图像，k（s, t）表征卷积核。

卷积层的应用广泛与不可或缺源于其两个重要的特征：参数共享与局部感受野。在设置卷积层时，通常只需设置三个参数，即零填充、卷积核的数量和尺寸、步长。参数共享就是因为参数量的大小只决定于卷积单元的数量和大小，与输入数据特征无关，故而经卷积运算后的数据都会产生一个特征图。卷积神经网络因而大幅缩减参数量，降低了运算复杂度。局部感受野则是指特征图中每一个值通常都对应上一层中的一定大小的区域，通过迭代使得网络学习逐渐学习到高层特征，既符合人类管中窥豹、以点带面的认知方式，又充分利用了图像的空间结构，如图 2-3 所示。

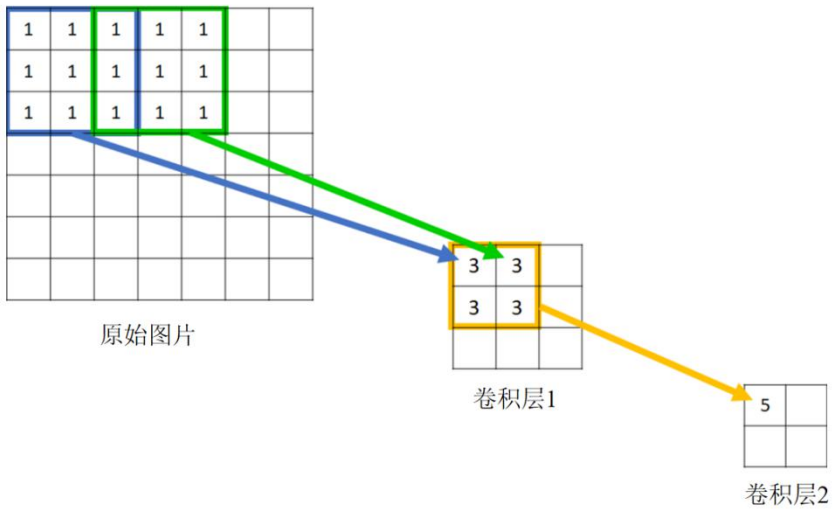

图 2-3　卷积层中的局部感受野

2.1.3　池化层

在普适的卷积神经网络中，通常会在卷积层中依次交叉存在着池化层，从而起到保持特征不变性、特征降维、防止过度拟合、扩大感受野、实现非线性等作用。池化层实际上是一种降采样，也可称作下采样，思想上仿照人类视觉理解中对视觉对象的降维和抽象化理解。经降采样运算后，输出的特征值经非线性池化函数处理后每一值都映射为输入的一块矩形区域，且单次运算后仅输出一个值。常见的池化层有：最大池化、平均池化、全局平均池化、全局最大池化。以最大池化为例，其公式如 2-2 所示，最大池化是将矩形选中区域的最大值选取再输送为结果；以平均池化为例，其公式如 2-3 所示，它是将矩形区域中的数值相加求平均值输出为结果。如下图 2-4 是两种典型池化方法的计算示意图。

$$\begin{aligned}\alpha_i &= \text{argmin}\,\min_{\alpha} L(\alpha,D) \triangleq \ \|x_i - D\alpha\|_2^2 + \beta\ //\ \alpha //_1,\\ h_{m,j} &= \max_{i\in N_m} \alpha_{i,j}, \text{ for } j = 1,\ldots,K,\\ \alpha_i &\in \{0,1\}^K, \alpha_{i,j} = 1 \text{ iff } j = \underset{k\le K}{\text{argmin}} \|x_i - d_k\|_2^2,\end{aligned} \tag{2-2}$$

$$h_m = \frac{1}{|N_m|}\sum_{i\in N_m} \alpha_i, \tag{2-3}$$

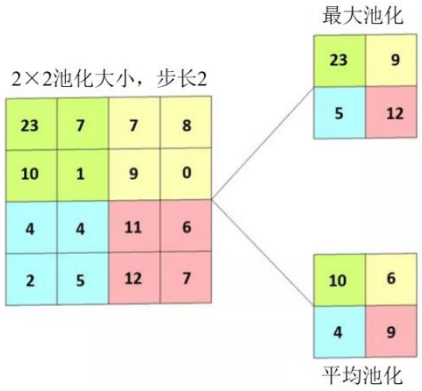

图 2-4　最大池化和平均池化

与卷积层不同的是，池化层通常只有两个参数：步长和池化核大小。且池化层不需要学习参数的步骤，仅根据池化函数的类型来决定在反向传播时梯度的大小取值。以最大池化为例，规定梯度只在区域最大值处传播，剩余区域梯度值取 0 值。在大多数卷积神经网络中，最大池化层是首选之举。

2.1.4　激活函数

由于神经网络本身便是受生物神经学科启发而来，激活函数便是受神经元的“激活状态”启发而设置的。将网络中数据的传递类比于神经元的电位传递，当数据经卷积层等运算后均会遇到激活函数这一节点，激活函数对传递来的数据进行判别后，决定其是否继续转化、传递。常见的激活函数主要分为两类：饱和激活函数（Saturated Neurons）和非饱和函数（One-sided Saturations）。典型的饱和激活函数有 Sigmoid 型和 Tanh 双曲正切型等，而典型的非饱和激活函数有校正线性单元函数（Rectified Linear Units, ReLU）以及其变种等。下图 2-5 展示了常见的激活函数。

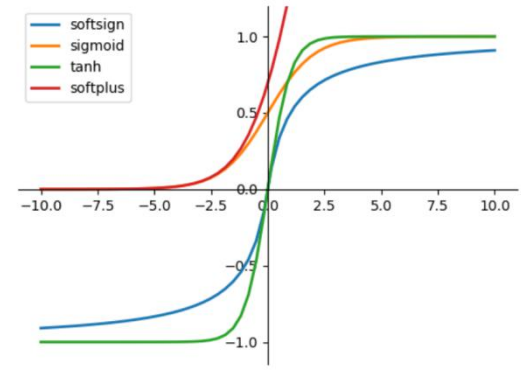

图 2-5　常见的深度学习激活函数

激活函数的作用包括：增强网络的非线性从而能够处理非线性学习任务，减少网络冗余特征、进一步加强拟合能力等。相比于饱和激活函数，非饱和激活函数可以针对梯度消失这一难题进行一定程度上的减轻，且能够对收敛过程起到加速作用。以 ReLU 函数为例，梯度值规定只取两个值：0，1。当输入＜0 时，梯度取 0 值；当输入＞0 时，梯度取 1 值。因而其梯度的连乘不会收敛至 0，连乘的结果只取两个值：0，1 。若值为 1，梯度保持当前值不变继续向前传播；但若值为 0 ,梯度在当前位置停止传播。ReLU 激活函数的公式如 2-4 所示：

$$f(x)=\begin{cases}x, f(x)\geq 0\\ 0, f(x)<0\end{cases}$$

$$f(x)=\max(0,x) \tag{2-4}$$

2.1.5　全连接层

全连接层（Fully Connected Layers, FC）通俗而言可以当作“分类器”放置于整个网络结构的末尾，将之前网络中提取出的特征进行综合整理后多分类，最终得出预测结论。其计算公式如 2-5 所示。

$$Out_i=\sum_j Weight_{ij}\cdot I_j+\mathrm{Bias}_i \tag{2-5}$$

其中：I_j 代表上一层所传递的数值，$Weight_{ij}$ 表示对于第 i 个神经元 I_j 的权重参数，衡量这一数值的重要程度，$Bias$ 是偏置参数，为了调整结果而设置，Out 表示全连接层的输出值。显而易见，全连接层的参数值具有多且重要的特性，一般可占用全网络的 80%左右，因而计算量庞杂且容易产生过度拟合现象。近年来的研究因全连接层对输入数据的大小的严苛要求，常用卷积层来替代且在多个视觉任务中展现出了良好表现，因而这一结构变得不再不可或缺。

2.2 基于深度学习的目标检测算法

目标检测因其源远流长，溯根及源，典著良多，方法浩繁。最为经典的表述[45, 46]通常将目标检测描述为滑动窗口分类问题。深度卷积神经网络（CNN）[1]的兴起便是起源于目标检测，其成功的案例导致了经典范式的迅速改良发展。拥有强大的主干特征提取器是准确检测目标的关键。当前主流的检测网络可以分为两类：双阶段目标检测算法和单阶段目标检测算法。

2.2.1 双阶段目标检测算法

双阶段目标检测通常将任务分为两个进程，第一阶段先大致将感兴趣候选区域选取出来，第二阶段具体判别感兴趣区域中的类别并进行回归操作。在众多提出的方法[47, 48]中，Girshick 等人创意性地将卷积神经网络模型基于区域进行预测（RCNN）[5, 49]，此方法因其出色的有效性而受到了研究者的广泛关注。由基于区域的深度卷积神经网络[5]开创的检测流程具体如下，它首先从目标图片中将感兴趣的建议区域选取出来出，并将一个被训练而出的网络来孑然地对每个感兴趣区域（Region Of Interest, ROI）实施分类操作。该思想已被论文[10, 50]延伸至在悉数的建议区域之间公共分享卷积特征图。但此方法由于要求图片的尺寸大小必须固定，因而造成了图像拉伸或裁剪而丢失部分信息的缺点。Faster R-CNN[5]使用区域建议生成网络 （Region Proposal Network, RPN）生成目标建议，且取得了最优异的结果，不仅使得检测速度加快，还让检测精准度更上一层楼。这为之后的繁多的研究夯实了根底[51, 52]。Faster R-CNN 也非常灵活且泛化能力强，能够延展到相似的同领域工作中，譬如图像分割[53]等。He 等人在文献[54]中提出了 Mask R-CNN 网络为实现语义分割任务设计出全卷积网络分支来对 Mask 目标进行提取操作。不仅解决了量化误差问题，还提高了网络对于小目标的检测性能。

2.2.2 单阶段目标检测算法

单阶段的目标检测算法是基于回归的思想省略了双阶段算法的粗略提取感兴趣候选区域的步骤，为了将目标检测的速度提高，通过主干网络得到特征图的同时，对目标对象实施分类操作和回归操作，但相应牺牲了精度。Redmon 等人[7]提出的 YOLO 网络和 Liu 等人[4]提出一种多尺度特征检测模型 SSD 均是单阶段目标检测算法的先驱，后人在此基础上发展出一系列改进网络机制。除检测精度和速度上的差异，单阶段检测算法由于直接在特征图中生成感兴趣候选区域，因而存在正负样本失衡现象，负样本过多但对训练作用较小，正样本因比重太小不能充分表现性能，从而导致算法无法得到最佳的训练结果。相反，双阶段算法则

由于第一阶段已筛选出感兴趣区域不会产生样本不平衡的后果，从而保证了训练效果。

2.3 域自适应

域自适应是在将训练数据集上习得的知识迁移到测试数据集上时，当测试数据集无标签且两个数据集分布不一致时，针对如何成功迁移学习所提出的方法。各种视觉领域任务（例如图像分类和语义分割）已经针对研究了弥合域之间差距的问题[19, 55]。为了解决这个问题，大量的方法利用了训练和测试域之间的特征分布匹配。基本思想是测量不同域的特征分布之间的某种距离，并训练特征提取器以最小化该距离。已经提出了各种测量距离的方法[22, 56]。目前主流的两种距离度量标准是最大均值差异（Maximum Mean Discrepancy, MMD）[57]和 H 散度（H-divergence）[58]。受理论结果[58, 59]的启发，各种方法利用域分类器[19, 22]来测量域差异。这些方法均以对抗的方式训练域分类器和特征提取器，就像训练 GAN 网络[21]所做的那样。此类方法旨在将目标的特征分布与源的特征分布严格对齐。此外，Long 等人[60]设计了域分类器的损失函数，以完全匹配域之间的特征从而来应用于图像分类。

域自适应现今已被普遍运用钻研于计算机视觉领域中的图像分类等任务[21, 61]。通常普适的学习模型涵盖：不对称度量学习[16]、子空间对齐[54]、子空间插值[17]、协方差矩阵对齐[14]、测地线流式核[62]、域迁移多核学习[63]等。近期的研究任务大多是为了将深度卷积神经网络的域自适应性有所提升，包括众多经典研究[20, 64]。与分类任务不同，本文所专研的目标检测问题，由于需检测目标的坐标方位和其种类均需要进行判断测算，因此愈加考验算法的有效性，为研究者提出诸多难题。迩来，有研究者大胆提出了一些办法来在源数据和目标数据之间实施未匹配图片的变换，从某种角度而言，这也能够当作是像素级的域自适应[24, 65]。然而，在实际中，根据自动驾驶等实际应用中的客户需求来产出高分辨率的真切图片依旧是一个极具难度和复杂性的问题。

2.3.1 域自适应在检测方向中的应用

与分类任务的域自适应的研究相比，研究者对其他计算机视觉任务的域自适应的关注则要少得多。最近，逐渐兴起对语义分割[66]，细粒度识别[67]等任务的研究活动。针对检测方向的研究工作，文章[31]提出经由援引自适应 SVM 模型（DPM）来降低基于可变形部分的域偏移难题。在近期的一项研究[68]中，Namboodiri 等人应用 R-CNN 模型作为网络中的特征提取器模块，继而利用子空

间对齐的策略将提取出的特征与之对齐。还生发了由另外的出处习得训练检测器的研究作业，譬如图片与视频的连接过渡[69]、3D 模型[70]或人工制造模型[71]。但目前大多数研究工作所实现的域自适应目标检测仍停留于或者无法实现以端到端的策略实施训练模型，抑或是只能专注于某些指定示例的阶段。

2.3.2 使用 H 散度的分布对齐

Ben-David 等人[58]提出了度量目标域和源域的分布差异的一种主流测量标准——H 散度（H-divergence）。即 H 散度[58]旨在衡量拥有不同分布的两组样本之间的散度。若令 $\mathbf{x}$ 表征一个特征向量，源域样本以 $\mathbf{x}_S$ 来表征，目标域样本以 $\mathbf{x}_T$ 来表征。令 h 表征域分类器：$\mathbf{x} \to \{0, 1\}$，为了使得预测源样本 $\mathbf{x}_S$ 等于 0，目标域样本 $\mathbf{x}_T$ 等于 1。倘使 H 为满足要求的域分类器的齐集，则界说 H 散度的源域和目标域之间的距离公式如下：

$$d_{\mathcal{H}}(\mathcal{S},\mathcal{T}) = 2(1 - \min_{h\in\mathcal{H}}(\mathrm{err}_{\mathcal{S}}\,(h(\mathbf{x})) + \mathrm{err}\,r_{\mathcal{T}}(h(\mathbf{x})))) \qquad (2\text{-}6)$$

其中：err_S 和 err_T 分别是 $h(\mathbf{x})$在源域和目标域样本集上的测算偏差。上述界说表明两域间的距离 $d_H(S, T)$与域分类器 h 的错误率被看作成反比。换而言之，当最优秀的域分类器的偏差很大时，即目标域和源域难以辨别，就会导致两个域所得的值十分接近，从而降低了网络性能，反之亦然。

在深度卷积神经网络中，特征向量 $\mathbf{x}$ 普遍涵盖其中一层紧接其后的激活函数。用 f 表示产生 $\mathbf{x}$ 的网络。因旨在使得目标域和源域对齐，务必使得 f 网络输出最小化的两域间的距离 $d_H(S, T)$ [22]的特征向量，这将会导致：

$$\min_f d_{\mathcal{H}}(\mathcal{S},\mathcal{T}) \Leftrightarrow \max_f \min_{h\in\mathcal{H}}\{\mathrm{err}\,\mathcal{S}(h(\mathbf{x})) + \mathrm{err}_{\mathcal{T}}\,(h(\mathbf{x}))\} \qquad (2\text{-}7)$$

这使得网络能够得以以对抗训练的策略逐步实施改进完善。Ganin 等人[22]成功实现了梯度反向层（GRL），并将其集成到卷积神经网络中，用于无监督的域自适应情景中的图像分类任务。

第三章　基于域自适应的 Faster R-CNN 目标检测算法

3.1　Faster R-CNN 目标检测算法

本节简要回顾了 Faster R-CNN[52]模型，它是本文研究工作中使用的基线模型。在本文的模型算法中，拣选最先进的 Faster R-CNN 作为的基线检测器，并采用域自适应模块提高其在陌生目标领域中目标检测的对于新样本的适应能力。

Faster R-CNN 模型可以看作是一个双阶段目标检测器，一般当作是三个主要模块组成：共享基底卷积层、区域建议网络（RPN）以及基于感兴趣区域（ROI）的分类器。下面是对于这三部分模块的详细介绍。如图 3-1 展示的是此模型的结构图。

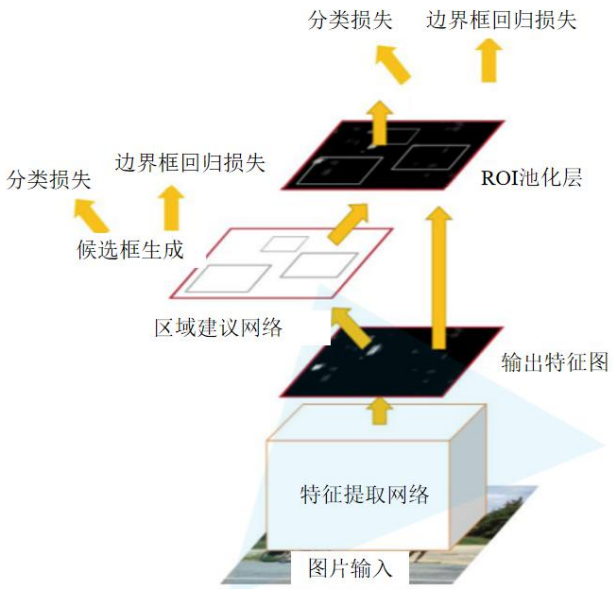

图 3-1　Faster R-CNN 目标检测器

3.1.1　特征提取网络

共享基底卷积层也称为特征提取网络，这一组件是所有算法的基础，之后的两个模块均在此基础上进行。任何网络都是将目标图像输入后最先体现为由同一分享的基底卷积层产生的卷积特征图。优秀的检测算法的性能上限常常受共享底部卷积层限制。目前，常见的特征提取能力优异的网络模型涵盖 VGGNet[41]、ResNet[42]、GoogLeNet[72]等等。一般而言，共享底部卷积层都会在特定数据集上

进行预训练操作以提高特征提取能力和便于后续的学习任务。本文中所采用的特征提取网络是 ResNet50。

3.1.2 区域建议网络

在基于共享基底卷积层生成的特征图上，设置好锚框的区域建议生成网络使用大小一定的卷积核在特征图上遍历来生成感兴趣候选对象建议，这是 Faster R-CNN 相较于 Fast R-CNN 改进的一点，这样用区域建议网络取代选择性搜索算法使得算法对于目标的检测速度大幅提升且提高了检测的精确程度。Faster R-CNN 算法共设置 9 种锚框使得单次遍历会生成 9 个候选区域，然而每个候选区域都伴随一个置信度的生成，不达阈值的区域将被筛选淘汰。此外，算法还采用了非极大值抑制方法来比较重叠度很高的区域的置信度，淘汰置信度低的区域，从而节省计算资源，提高模型检测效率。

3.1.3 基于感兴趣区域的分类器及损失函数

最后，ROI-wise 分类器从使用 ROI 池化获得的特征向量预测类别标签。具体而言，是把 RPN 选取的感兴趣候选区域逐一对应到共享基底卷积层生成的特征图上，再进而调整目标检测框的位置以确定目标类别标签。

这一网络的训练损失主要由区域建议网络的损失和基于感兴趣区域的分类器的损失组成，如公式 3-1 所示：

$$L_{det} = L_{rpn} + L_{roi} \tag{3-1}$$

RPN 和 ROI 分类器的训练损失都分别包含两个损失项：一类是用于概率测算的精准程度的分类损失，还有一类是用于更优秀的定位目标的框坐标的回归损失。

3.2 目标检测的域自适应

在深度卷积神经网络极大地提高了对象识别的精确性的当下，绝大多数研究依然有赖于标记丰富的训练数据集。尤其对于目标检测任务，标注工作尤为繁琐：每幅图像中的每个目标类别的每个实例都必须使用高精确度的边界框进行标注。因而从标签精良的域转移预训练模型便成为一个极富吸引力的解决方案，但数据集差异性通常会降低这种发放对新数据集的泛化能力。如今已经提出了各种无监督域自适应（Unsupervised Domain Adaptation，UDA）方法来解决数据集偏差问题[23, 61]。

依据域自适应中的一般性通用术语，本文所称作的源域指的是训练数据集的域，用 S 表征，而称作目标域的则是指测试数据集的域，表征为 T 。譬如，当

使用 SIM10k 数据集进行训练并使用 PASCAL VOC 数据集进行测试时，S 指代 SIM10k 数据集，T 指代 PASCAL VOC 数据集。本文遵循无监督的域自适应的典型配置，即有权限查看源域中的图片和并在源域中应用有监督策略（譬如：边界框和目标种类），然而在目标域中只有未标注的图片可以被利用。本文的工作主要是习得能够适应无标注的目标域的目标检测模型。

3.2.1 概率角度分析

目标检测问题一般而言能够当成旨在习得后验概率 $P(C,B\mid I)$，其中 I 代表图像表征，B 表征的是框选目标的边界框，$C\in\{1,\ldots,K\}$ 表示目标的具体类别（K 表示目标类别的总数）。

$P(C,B,I)$表征的是用于目标检测的训练样本的联合分布，并使用 $P_S(C,B,I)$ 代表源域的联合分布以及 $P_T(C,B,I)$ 来代表目标域的联合分布。需要注意的是，虽然在训练期间目标的边界框和类别标签注释（即 B 与 C）是未知的，但是我们在此处仍使用 $P_T(C,B,I)$来剖析域迁移的现象。如果域偏移现象在此处出现时，则令 $P_S(C,B,I)\neq P_T(C,B,I)$。

3.2.2 图像级的自适应

使用贝叶斯公式，联合分布可以分解为：

$$P(C,B,I)=P(C,B\mid I)\,P(I) \tag{3-2}$$

宛如分类任务一样，针对目标检测任务需要做出协变量偏移假设，即目标域和源域的条件概率 $P(C,B\mid I)$ 相同，边际分布 $P(I)$上的不同从而引发域间分布的偏移。换而言之，在目标域抑或源域中，目标检测器的结果之间是统一的：对于指定的某一张图片，无论目标对象最终隶属任一域中，检测的数据结果都理当是一致的。对于 Faster R-CNN 模型分析可得，基底共享卷积层的特征图的输出本质上由图像表征 I 表示。故而，图像级的自适应旨在缓解域偏移带来的后果，普遍使得目标域和源域的图片表征分布必须一致（即：$P_S(I)=P_T(I)$）。

3.2.3 实例级的自适应

另一方面，联合分布也可以分解为：

$$P(C,B,I)=P(C\mid B,I)\,P(B,I) \tag{3-3}$$

在协变量偏移假设下，即两个域的条件概率 $P(C\mid B,I)$ 相同，可以看出边际分布 $P(B,I)$ 带来了域分布的偏移。直观地说，这往往预示着域间的语义一致性：无论指定的涵盖某一目标的相同图片区域来自哪个域，它最终识别出的类别标签都应该是保持一致的。故而，实例级的对齐通常是指域间的实例级表征的分布也

必须被规定具有一致性（即：$P_S(B, I) = P_T(B, I)$）。

此处所谓的实例级表征（B，I）具体指是那些从每个例子的真实边界框中的图片区域中选取出的特征。尽管边界框的标注在目标域中无法被利用，但依旧能够经由 $P(B, I) = P(B \mid I)\, P(I)$ 求得，其中边界框预测器用 $P(B \mid I)$ 表征。$P(B \mid I)$是域不变的是上述情况的前提条件，这同样适用于我们下面采用的方法。

3.2.4 联合适应

在理想状态中，能够实现域对齐操作既在图像层级上，也能在实例层级上实施。思量到 $P(B,I) = P(B \mid I)\, P(I)$ ，而且通常预设条件分布 $P(B \mid I)$ 对于目标域和源域均是等同且非零的，因此会得到：

$$P_S(I) = P_T(I) \Leftrightarrow P_S(B, I) = P_T(B, I) \tag{3-4}$$

换句话说，如果两个域的图像级表征的分布相同，则实例级表征的分布也相同，并且反之亦然。然而，实际上完美地估计出条件分布 $P(B \mid I)$ 通常并非易事。原因有两个：（1）实际上可能很难完美地对齐边际分布 P（I），这意味着最先用于估计 $P(B \mid I)$ 的输入在一定程度上存在偏差；（2）边界框标签注释是仅适用于源域去训练数据集，因此 $P(B \mid I)$ 仅使用源域数据学习，进而学习结果容易偏向源域。

为此，本文采用的方法在执行域分布对齐时，不仅在图像层级上，还在实例级层级上，还为了达到削减估计 $P(B \mid I)$ 的偏差的目的使用一致性正则化来进一步优化。如 2.3 节所述，旨在对齐目标域和源域的分布，一个域分类器 $h(\mathbf{x})$ 亟须得到训练。在目标检测的上下文中，图像级表征 I 或实例级表征 (B, I)可以用 $\mathbf{x}$ 来代表。从概率论的角度来看，估计样本 $\mathbf{x}$ 的属于目标域的概率可以用 $h(\mathbf{x})$ 来表征。

故而，通过用 D 来表征域标签，图像层级的域分类器可以看作是推断 $P(D \mid I)$，而实例层级的域分类器可以看作是推断 $P(D \mid B，I)$。通过使用贝叶斯定理，可以得到：

$$P(D \mid B, I)\, P(B \mid I) = P(B \mid D, I)\, P(D \mid I) \tag{3-5}$$

尤其须注意的是，$P(B \mid I)$ 是域不变的边界框预测器，$P(B \mid D, I)$ 是域依赖的边界框预测器。复盘上述，在实践中，之所以我们只能学习领域可依赖的边界框预测器 $P(B \mid D, I)$，是因为我们没有目标域的边界框标签注释。故而，经由对两个域分类器之间的一致性的必要规定，即 $P(D \mid B, I) = P(D \mid I)$，可以逐渐学习使得概率上 $P(B \mid D, I)$来逼近 $P(B \mid I)$。

3.3 算法原理

算法基于最先进的 Faster R-CNN 模型[6]搭建了一个端到端的深度学习模型，采用了域自适应组件和局部强特征对齐组件来增强了 Faster R-CNN 基础架构，称为域自适应 Faster R-CNN 模型。网络结构如图 3-2 所示，其中 D_l 表示卷积核为 1 的全卷积网络，基础网络的特征提取器 F 被分解为 F_1 和 F_2，F_1 的输出是 D_l 的输入。

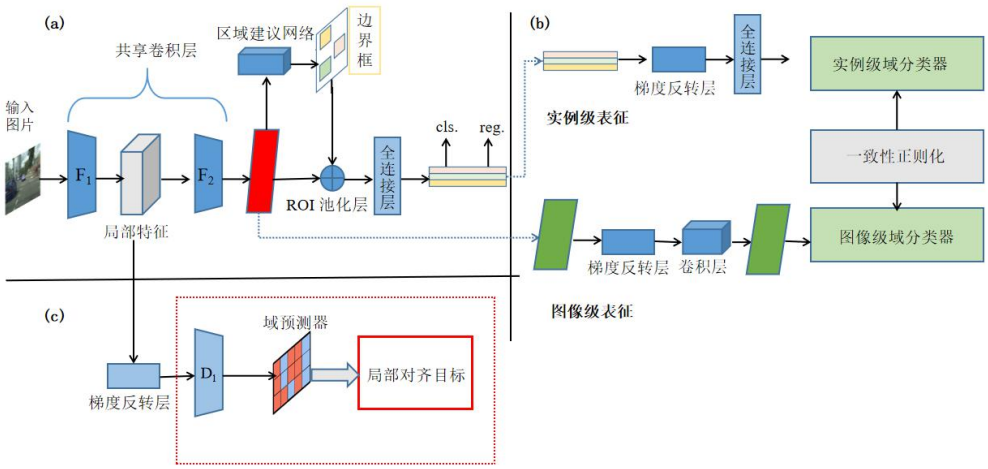

图 3-2　所采用网络架构模型图

(a)　基础 Faster R-CNN 模型；(b)　域自适应组件；(c)　局部域分类器网络

图 3-2 的左上方展示的是最初的 Faster R-CNN 基础模型。所有模块均共同享有基底卷积层。继而在顶部构建 RPN 和 ROI 池化层，而后为提取实例层级的特征用两个全连接层接续。

本文的域自适应 Faster R-CNN 中引入了四个新模块。第一个，图像层级的域分类器补充在末尾的卷积层之后。第二个，在 ROI-wise 特征的尾端增添实例层级的域分类器。上述二者域自适应模块是通过最小化两个域之间的 H-散度以用来解决域偏移。另外，在每一组件中都有训练域分类器并采用对抗方法来学习域不变性的鲁棒性。第三个，用一致性正则化来关联这两个分类器，以鼓励 RPN 保持域不变特性。第四个，我们采用在 RPN 之前从较低层提取局部特征，在底层特征空间中进行强局部特征对齐的方法。局部特征的强对齐会匹配不同域的纹理或颜色，并且在大多数情况下会提高性能。这是因为它不会改变类别信息的同时会减少域间隙。在本文中，“局部”并非指实例（目标）尺度，而是指代具有小感受野的纹理或颜色特征。

3.3.1 强局部特征对齐

局部的域分类器 D_l 的架构旨在着眼于局部特征并非全局特征。D_l 是一个内核大小等于 1 的全卷积网络。特征提取器 F 被分解为 $F_2 \circ F_1$，F_1 的输出是 D_l 的输入，如图 3-2 所示。F_1 输出宽度和高度分别为 W 和 H 的特征。D_l 输出与输入特征具有相同宽度和高度的域预测图。此方法依照论文[50]采用最小二乘损失来训练域分类器。这种损失函数有助于稳定域分类器的训练，并且经验证明对于对齐底层特征很有效。强局部对齐 L_{loc} 的损失函数总结为：

$$\begin{aligned}
\mathcal{L}_{loc_s} &= \frac{1}{n_s HW}\sum_{i=1}^{n_s}\sum_{w=1}^{W}\sum_{h=1}^{H} D_l\left(F_1\left(x_i^s\right)\right)_{wh}^2 \\
\mathcal{L}_{loc_t} &= \frac{1}{n_t HW}\sum_{i=1}^{n_t}\sum_{w=1}^{W}\sum_{h=1}^{H}\left(1 - D_l\left(F_1\left(x_i^t\right)\right)_{wh}\right)^2 \\
\mathcal{L}_{loc}(F, D_l) &= \frac{1}{2}\left(\mathcal{L}_{loc_s} + \mathcal{L}_{loc_t}\right)
\end{aligned} \tag{3-6}$$

其中 $D_l\left(F_1\left(x_i^s\right)\right)_{wh}$ 表示域分类器在每个位置的输出。损失旨在将特征的每个感受域与另一个域对齐。

3.3.2 图像级域分类器

在 Faster R-CNN 模型中，图像级表征具体是说基底卷积层的特征图输出（参考图 3-2 中的原始模型部分的红色四边形）。如图 3-2 的右下部分所示，为了解决图像级别的域分布偏移问题，本文采用了一个基于块区的域分类器。

特别要指出，我们在特征图的任一激活函数上都训练一个域分类器。由于任一激活函数的感受野均与输入图像 I_i 的一个图像块区相对应，因此域分类器实际上是对任一图片块区的域标签的预测。

这种做法的好处是双重的：（1）图像级对齐表征一般对降低源于图像风格、图片比例、照明等全局层面上的图片差别所导致的偏移有一定的作用。在最近关于风格迁移的研究工作[73]中相仿的基于块区的损失业已被实证是可行且有用的，它也可以处理全局转换问题。（2）在训练目标检测网络时，因为高分辨率输入的应用，所以批量大小（Batch Size）通常非常小。此类基于块区的架构对于提高训练域分类器的训练的样本数量大有裨益。

用 D_i 表征第 i 个训练图片的域标签，则对于源域 $D_i = 0$，对于目标域 $D_i = 1$。把 $\phi_{u,v}(I_i)$ 用于表征位于基底卷积层之后的第 i 个图像的特征图的坐标为 (u, v) 处的激活函数。把 $p_i^{(u,v)}$ 用于表征域分类器的输出并应用交叉熵损失，图像层级的自适应损失应表示为：

$$L_{img} = -\sum_{i,u,v}\left[D_i \log p_i^{(u,v)} + (1 - D_i)\log\left(1 - p_i^{(u,v)}\right)\right] \tag{3-7}$$

如第 2.3 节所述，若要使得域分布对齐，则应使得域分类器的参数同时得到优化处理以最小化上述的域分类器损失，另外使得基础网络的参数得以同时优化以最大化这种损失。为实现这一目标，本文采取了梯度反向层（GRL）[22]的方法，而在训练域分类器时则运用普通梯度下降。当经由 GRL 层使得基础网络改善时，梯度的符号会在这个过程中被反转。

3.3.3 实例级域分类器

实例级表征是指在输入最终种类分类器前基于 ROI 的特征向量（即图 3-2 中全连接层之后的矩形）。实例级对齐表征对于降低局部实例差别引起的域偏移大有裨益，例如观察视角、目标的外形、大小等。与图像级自适应相仿，为了使得实例级分布的特征向量对齐，本文中训练了域分类器。用 $p_{i,j}$ 来表征第 i 个图像中第 j 个区域建议的实例层级的域分类器的输出。因此，实例级自适应损失应写作：

$$\mathcal{L}_{ins} = -\sum_{i,j}\left[D_i \log p_{i,j} + (1-D_i)\log\left(1-p_{i,j}\right)\right] \tag{3-8}$$

此外，还增添了一个梯度反向层在域分类器之前来以便应用对抗训练学习的方法。

3.3.4 一致性正则化

如第 3.2 节所述，在不同级别上强制要求域分类器间的一致性对于学习边界框预测器（即区域建议网络模块）的跨域时的稳定性有一定的帮助。故而，模型内又增加了一个一致性正则化模块。鉴于图像层级的域分类器为图像级表征 I 的任一激活函数生成相应输出，故而把图像级的概率等同于图像中所有激活函数的均值。因而，一致性正则化模块的损失应写作：

$$L_{cst} = \sum_{i,j}\left\|\frac{1}{|I|}\sum_{u,v} p_i^{(u,v)} - p_{i,j}\right\|_2 \tag{3-9}$$

其中$|I|$表示特征图中的激活函数的总数，$\|\cdot\|$ 表示 ℓ_2 距离。

3.3.5 损失函数

基于域自适应的跨域目标检测网络架构的最终训练损失是每个单独部分的总和，损失函数可以写为：

$$L = L_{det} + L_{loc}(F, D_l) + \lambda\left(L_{img} + L_{ins} + L_{cst}\right) \tag{3-10}$$

其中，λ 是使得 Faster RCNN 的损失和本文所采用的域自适应组件和局部强对齐模块达到均衡的权重参数。网络能够利用标准的 SGD 算法以端到端的策

略进行训练。需要关注的是，应用的 GRL 层使得域自适应组件的对抗性训练学习得以成功实现，在传播过程中该层具有自动反转梯度符号的特性。此外，图 3-2 中的全局网络是用于训练进程中的。在预测判决阶段中，便可以移除两个域自适应模块，且仅仅应用到拥有自适应权重的最初的 Faster R-CNN 基础组织架构。

第四章　实验设计与结果分析

4.1　实验数据集

算法、数据、计算能力是计算机视觉任务性能表现最密切相关的三个关键因素，其中数据集的出现极大程度上帮助验证了算法的鲁棒性且一定程度上帮助其进行优化。本文中，为了检验基于域自适应的跨域目标检测算法的可靠性和有效性，通过 Cityscapes 数据集[74]和 Foggy Cityscapes 数据集[75]来比较分析算法检测性能。此外，本文中使用数据集时，被当作源域的数据集将会在利用数据样本的时候一齐利用到标注的数据，而被当作目标域的数据集将只是利用到其中的样本。下面是对这两个数据集的具体阐述。

4.1.1　Cityscapes 数据集

Cityspaces 是由奔驰推出的经典城市景观数据集，数据集图片视角是行驶中的汽车所摄，在自动驾驶、无人驾驶场景中的图像分割、3D 目标检测、语义分割等领域的任务中被普遍运用，且都涵盖相应的打标工细的子数据集，常被应用于对计算机视觉领域的算法模型在都市区域场景中语义理解的性能表现的评价。具体地，Cityscapes 数据集涵盖了 50 余个各类城市的多场景、全季节、复杂背景的城市街景，包含 30 类标注目标、5000 张标注完整拍摄所得的图片、以及标注粗略的 20000 张图片。

在评价体系上，数据集采用目标检测数据集 PASCAL VOC 标准的交并比（Intersection-Over-Union，IOU）得分对算法性能表现实行评估。此外，数据集对于标注完整和粗略的图像分别设置了 Fine 和 Coarse 两套评估标准。

数据集包含的文件 image base 和 annotation base 分别对应文件 LeftImg8bit（包括 5030 个项目，总共 11.6 GB，确切的含 5000 个项目）和 GtFine（30030 个项目，总共 1.1 GB）。其中均包含 train、val、test 这三个文件夹，分别对应测试图片 1525 张，训练图片 2975 张和验证图片 500 张，共计 5000 张标注完整的图片。其中 LeftImg8bit 中的 train 包含 18 个子文件分别对应 16 个德国城市和法国、瑞士各一个城市的图片文件；LeftImg8bit 中的 val 包含 3 个子文件分别对应 3 个德国的城市；LeftImg8bit 中的 test 包含 6 个子文件分别对应 6 个德国的城市；GtFine 中的 test 包含 18 个子文件对应 LeftImg8bit 中的 train 的子文件。另外，初期的数据集没有测试集，只有验证集，想得到测试结果需要提交模型在线测试。本文中采用标准 Cityscapes 8 bit 低动态范围格式、像素为 2048×1024 大小的数据集来研究基于域自适应的跨域目标检测任务，主要应用于 8 个常见城市场景目

标类别，涵盖：行人，骑行者，轿车，卡车，公交汽车，火车，摩托车，自行车。且均是取景于晴朗天气。

4.1.2 Foggy Cityscapes 数据集

Foggy Cityscapes 数据集是由 Sakaridis 等人[75]使用合成数据来理解雾景语义所制作的数据集。之所以通过在晴朗户外场景的真实图像 Cityscapes 数据集上应用雾合成来合成有雾图像，是因为在解决语义模糊场景的问题时，难以收集和注释有雾的图像。Foggy Cityscapes 数据集包含 3475 张图片，注释和数据拆分均沿用 Cityscapes 数据集，同样包含 8 个常见城市场景目标类别，非常适合研究天气条件引起的域偏移。该数据集是来自汽车行驶中采集的 Cityscapes 数据集的图像的深度渲染，因而有效地模拟了雾景且可与原数据集一起测试目标检测模型的对天气变化情况的跨域的泛化能力。在本文中，Cityscapes 数据集和 Foggy Cityscapes 数据集分别被当作跨域测试的源域和目标域，来检验目标检测模型的域自适应能力。

4.2 实验设置

4.2.1 实验开发环境

本文所采用的目标检测模型所需算力较高，因而实验所需的实验环境也要求颇高，详细软硬件配置见下表 4-1。

表 4-1 实验开发环境

环境	环境名称	配置
硬件环境	CPU	AMD Ryzen Threadripper 2950X 16-Core Processor
	内存	64GB
	显卡	Nvida GTX 1080Ti 11G x4
软件环境	操作系统	Ubuntu 20.04.3 LTS
	编程语言	Python 3.6
	深度学习框架	Pytorch 1.4.0
	第三方库	Cython, Matplotlib, Cuda 11.4

本文在运行实验中选取的是无监督域自适应协议。训练数据是两部分数据构成：源域中同时涵盖着图片信息和打标信息（即：目标种类和边界框）的训练数据，还有目标域中只涵盖着未打标的图片的训练数据集。

为了验证所采用的方法的鲁棒性和有效性，对于所有域迁移场景，本文展示了模型的最终训练结果以及通过组合不同组件（即图像级自适应、实例级自适应、一致性正则化、局部强对齐）的结果。本文采用原始的 Faster RCNN 模型作为基线模型，原模型训练时仅应用到了源域中的训练数据集，未曾考虑域自适应的场景。

4.2.2 实验评价指标

当论及目标检测模型的有效性和优劣，不仅仅要考虑算法和数据集的选取，还要考虑任务的需求，而合理的模型评价指标便是任务需求的体现。对于本文所有实验，实验指标主要采用平均精度（mAP）来衡量，阈值设定为 0.5 用于评估性能表现，故而，在实验中观察 AP50 的数值进行比较即可。

mAP 是常应用于目标检测的算法模型中的评价指标，英文全称为 Mean Average Precision，即多类平均精确度的再取平均值，下面都用 mAP 代称。为了具体了解这一指标的实际含义之前，需要知道以下名词的含义：精确率（Precision）、召回率（Recall）、交并比（IOU, Intersection Over Union）、平均精确度（AP, Average Precision）。

精确度也称为查准率，是用来衡量预测的准确性。换言之，正确预测的百分比，公式如 4-1 所示。召回率也称为查全率，是用来衡量检测到所有正例的能力，公式如 4-2 所示。譬如，某模型可以在前 K 个预测中检测到 80%的正例。IoU 是用来度量两个边界之间的重叠程度，公式如 4-3 所示。通常使用它来度量模型的预测边界与实际目标边界（ground truth）的重叠程度。在一些数据集中，通常预先定义了一个 IoU 阈值（比如 0.5）来分类预测是真阳性还是假阳性，本实验中 IoU 设定为 0.5。AP 是用来衡量模型对单一目标类别检测的平均精度，广义的定义是求出由精确度和召回率构成的曲线下的面积，公式如 4-4 所示，PR 曲线下的面积称为 AUC，AUC 面积的值越接近 1 代表模型性能越好。通俗意义上，曲线映射下的面积代表各自召回值下各类型准确度的平均值。狭义上，AP 和 mAP 有许多针对不同任务的计算方法，本文采用的是 PASCAL VOC2010-2012[76]中应用的计算方法。

$$\text{Precision} = \frac{TP}{TP+FP} \tag{4-1}$$

$$\text{Recall} = \frac{TP}{TP+FN} \tag{4-2}$$

$$IoU = \frac{\text{Pred} \cap GT}{\text{Pr}ed \cup GT} \tag{4-3}$$

$$AP = \int_0^1 p(r)dr \tag{4-4}$$

其中，TP 代表 True Positive，即正样本被判别为正确；FP 代表 False Positive，正样本经过判别为错误；TN 代表 True Negative，负样本经过判别为正确；FN 代表 False Negative，负样本经过判别为错误。Pred 代表预测出的边界框；GT 代表 Ground Truth，实际目标边界框；p(r) 代表精确度和召回率构成的曲线。

mAP 是用来衡量多种类检测任务的平均准确度。具体而言，是将所考虑的所有类别的 AP 进行计算，再求平均值，所得即为 mAP。

此外，除非另有说明，否则一切需测试和训练的图片大小均须事先调节，为与 GPU 内存相匹配略短一边的长度设定为 500 像素，另外凡实验中用到的 λ 均等于 0.1。本文的超参数设定依照的是 Ren 等人的论文[6]。简而言之，模型在实验时初始化的施行是依靠在 ImageNet 上预训练的权重。本文先以 0.001 的学习率微调网络来进行 5 万次迭代，然后以低至 0.0001 的学习率进行剩下的 2 万次迭代运算。每个批量（Batch）由 2 张分别来自目标域和源域的图像组成。本文的 L2 正则化中应用 0.0005 的权重衰减和 0.9 的以便更新模型中参数的超参数动量。

4.3 实验结果与分析

在本节中，详细评估了所采用的基于域自适应的域迁移目标检测 Faster R-CNN 模型，用于在不同领域场景中实施的目标检测：在恶劣天气下驾驶时，其中训练数据是在良好的天气条件下采集的 Cityscapes，而测试数据是源于雾天的图片。

本文经由对不同天气状况下的域迁移的探究来进行评估域自适应能力。域差异的主要来源之一便是气候状态，气候状态的转变会导致目标情境下的视觉观感的变迁。目标检测模型在气候状况变化迥异的时候是否能够高性能地完成目标检测任务对于可靠的自动驾驶需求尤为关键[75, 77]。下面，本节将会重点探究当把模型由气候晴朗转换成有雾场景时模型检测物体的本领。

4.3.1 消融实验的结果与分析

定量分析：表 4-2 罗列的是本文所采用的算法的检测结果和消融实验中各个模块添加与否的检测结果以及基线模型的结果。通过图像级和实例级的自适应模块、一致性正则化模块、局部强对齐模块的逐一累积添加到基线模型 Faster R-CNN 框架上，我们可以看到实验的评估指标 AP50 从最初的 24.9%逐步增加至 40.1%，相比于基线算法，本文所采用的数个模块共同作用下的算法的精确度提升了 15.2%。

定性分析：此外，由于本文所应用的两个数据集均涵盖着 8 个种类的检测目

标，AP50 的明显提升便意味着所采用的算法能够在绝大多数的类别检测中精确度有所提升，即提高了算法对新样本的适应能力。也相应证明了所采用算法能够减少不同目标类别之间的域差异，从而较好地适应了天气的变化与恶劣程度。

表 4-2　在基线模型基础上逐步增添模块后的检测精度

基线模型	Img	Ins	Cons	Loc	AP50
√					24.9
√	√				38.0
√		√			37.5
√				√	31.9
√	√	√			39.8
√	√	√	√	√	40.1

其中，img 代表图像级自适应组件，ins 代表实例级自适应组件，cons 代表一致性正则化组件，loc 代表局部强对齐组件，√代表该模型使用了该组件。

定量分析：在表 4-2 中，实验的检测结果表明两个自适应模块和局部强对齐模块均对于降低域偏移有不同程度的裨益。提升程度由低到高排列，局部强对齐模块的增加致使精度提升了 7.0%，实例级自适应模块的增加导致精度提升了 12.6%，图像级自适应模块的增加导致精度提升了 13.1%。

定性分析：不难发现局部强对齐模块对系统性能虽有所提升，但提升能力相比两个自适应模块较为薄弱。一方面有所提升的精度是由于在本文研究的数据集中域偏移主要是因为有雾气候引起的噪声所致，在域偏移中属于局部层级的偏移。故而，在两域中对应图片拥有几乎全部一致的图像分布组合和目标的个数的情况下，强有力的局部级别对齐模块一定程度缩小了这种偏移。因而，在此类数据情境下，不同域之间的局部强对齐相比于基线模型是更为有效的。然而，针对更为复杂且差异悬殊的数据集间，该模块的性能表现将预计无法与另外两个自适应模块相比。另一方面，本文主要实验代码的研究工作是围绕着基于域自适应模块展开的，因而对于这两个自适应模块的优先级更高和应用范畴更大，故此提升能力有所不及。

下面着重探究分析两个自适应模块作用。数据表明单独图像级的作用相比于单独实例级作用更大一些，对于改进基础模型检测准确度更为显著。为探究出这一现象的具体原因，本文分别为原始 Faster R-CNN 模型、仅具有图像级自适应的模型和仅具有实例级自适应的模型选择了 20, 000 个置信度最高的预测。采用 D. Hoiem 等人[78]的类似方法，本文将检测结果归纳为 3 种错误种类：（1）准确，即检测框与真值的相交大于 0.5；（2）未准确定位，即检测框与实际物体真实框

的重叠为 0.3 到 0.5；（3）背景，即检测框与真实框的交叠小于 0.3，这意味着检测器将背景错误地判别为正样本。图 4-1 展示了检测结果。

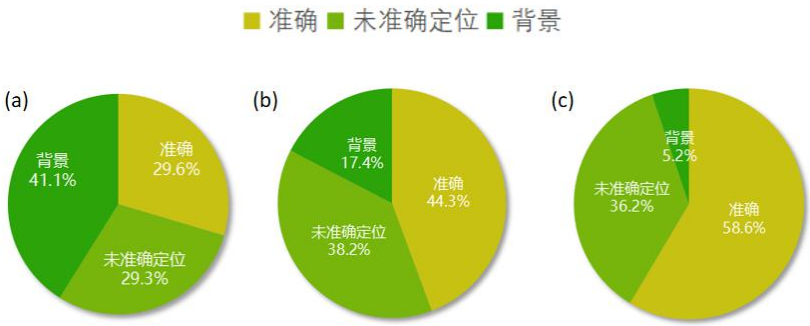

图 4-1　目标检测的错误分析

(a)　Faster R-CNN；(b)　单独采用实例级模块的模型；（c）单独采用图像级模块的模型

易观察得，两个模块任一添加至原始基线 Faster R-CNN 模型均会使得检测准确无误的比例增加，如上图所示，同时错误检测的比例明显下降。这一结果与之前的消融实验中的结果保持一致。另外，在“背景”这一项单独的实例级自适应的模块相比图像级自适应的模块与原始模型结合后产生的误差更大，正如消融实验中单实例级模块所提升的精度没有图像级模块提升大。分析缘由，一定概率是由于 RPN 的性能表现受到图像级对齐的影响而提升尤为明显，故而能够生成的建议区域拥有了更优异的定位能力。

如第 3.3 节所述，本文的一致性正则化模块施加在两个层级的域分类器上，以学习具有鲁棒性的区域建议网络。本文以 Cityscapes→Foggy Cityscapes 为例，为了将一致性正则化的好处更详尽展示使用，探究了一致性正则化模块应用前后区域建议网络的性能表现，结果展示在表 4-3 中，其中 Cons 代表一致性正则化模块。

表 4-3　三个算法的最大平均重叠值

评价指标	Faster R-CNN	无 Cons 模块算法	含 Cons 模块算法
mIoU	18.8	28.5	30.3

源于论文[27]启发，本文设置来自区域建议网络的前 300 个建议与真实情况当中能够实现的最大平均重叠作为测试对象。首先，基线的比较对象是原始的 Faster R-CNN 模型。如上表 4-3 所示，无一致性正则化模块的图像级和实例级自适应算法，mIoU（平均交并比）在此模型中比在基线 Faster R-CNN 模型中表现高出了 9.7%。在含一致性正则化模块的图像级和实例级自适应算法中，区域

建议网络的性能能够更进一步提升 1.8%，这表明一致性正则化模块能够促使区域建议网络更具有鲁棒性，很大程度上证明了本文所采用的模型可以有效降低目标检测的定位、误检、漏检概率问题。

4.3.2 同任务多类型算法比较与分析

在天气条件迥异时稳定的目标检测性能对于自动驾驶汽车等安全关键型应用至关重要。尤其是天气条件会引入可能对检测性能产生负面影响的图像伪影。故而为了评估所采用方法在恶劣天气中与其他近几年经典算法相比的有效性和优势所在，本文依旧利用 Foggy-Cityscapes 和 Cityscapes 作为目标域和源域。评估指标依旧采用 IoU 设定为 0.5 时，多类别检测任务的平均准确读 mAP 作为比较根据。一切参与测试、训练的图片也均需要调节大小，数值同上。超参数的设置仍然按照原始论文[6]所示。本文所比较的经典近几年算法涵盖如下所示：Domain Adaptive Faster R-CNN (DA)、Adapting Object Detectors (AOD)、Strong-Weak Distribution Alignment (SWDA)、Diversify and Match (DM)、Progressive Domain Adaptation (PDA)、Adaptive Object Detection (AODDMP)、Prior-based Domain Adaptive Object Detection (PDAOD)、Harmonizing Transferability and Discriminability (HTD)。

在表 4-4 中，展示了包括本文所采用算法在内的十种目标检测算法在跨域目标检测任务中的性能，且主要针对恶劣气候条件下的域自适应能力，定量结果分析是通过 mAP 这一指标来衡量。Ren 等人[6]提出的使用区域提议网络实现实时目标检测的 Faster R-CNN 算法在此任务中多目标平均检测精度为 24.9%；初代应用自适应模块的用于户外目标检测的域自适应 Faster R-CNN 算法[27]在此任务中平均检测精度比未添加自适应组件算法检测精度提升了 3.2%；2019 年 Zhu 等人的[28]通过选择性跨域对齐调整目标检测器则将平均检测精度提升至 33.8%，突破了 30%；同年的自适应目标检测的强弱分布对齐[29]的平均检测精度则更胜一筹，达到了 34.3%的检测精度；到了 2020 年在论文多样化和匹配：目标检测的域自适应表示学习范式[79]将平均精度提升至 34.6%；用于目标检测的渐进域自适应[80]则通过中间域的应用来弥合域差异从而将平均精度提升至 36.9%；具有双多标签预测的自适应目标检测[81]利用多标签预测结果使得平均精度提升到 38.8%；雾霾和多雨条件下基于先验的域自适应目标检测[82]则将在多雨和朦胧条件下的目标检测平均精度提升到 39.3%；协调适应对象检测器的可迁移性和可辨别性[83]则在此任务中达到了 39.8%的平均检测精度；在本文所采用的框架模型下性能表现提升至 40.1%，易观察得，本文所采用的算法在很大程度上优于近几年此领域中的大多数算法。此外，不难看出所采用的方法在所有类别中平均都表现良好，证明了泛化性能的突出优势和两种自适应模块、一致性正则化模块和局部强对齐

模块相结合应用的好处。

表 4-4　从 Cityscapes→Foggy-Cityscapes 数据集的天气域自适应的定量结果 (mAP)

实验方法	mAP
Faster R-CNN[6] (NIPS' 2015)	24.9
DAF R-CNN[27] (CVPR' 2018)	27.6
AOD[28] (CVPR' 2019)	33.8
SWDA[29] (CVPR' 2019)	34.3
DM[79] (ECCV' 2020)	34.6
PDA[80] (WACV' 2020)	36.9
AODDMP[81] (ECCV' 2020)	38.8
PDAOD[82] (ECCV' 2020)	39.3
HTD[83] (CVPR' 2020)	39.8
本文所采用模型	40.1

4.4　经济与社会影响

从经济角度考量，跨域目标检测这一技术如今的应用场景多为自动驾驶、交通等领域，而针对这一复杂系统中的功能要素，目标检测算法仅负责感知与识别物体的任务。当算法在精度和速度上均有所改善后，一定程度上可以降低整个系统的成本。但是否能够产生较大的经济效益，仍需决定于系统的各项组件。从长远的、发展的角度而言，是百利而无一害的。

从社会角度考量，能够适应各种环境和天气的目标检测算法的应用场景远非交通所限，在办公打卡、实时监控、社会治安等场景也有很多的发挥空间。本文所采用的算法在精度上的提升更对这项技术的广泛应用打下基础。

从法律角度考量，对于自动驾驶所产生的潜在风险责任归属问题仍是需要完关注的，相关规定也需要进一步跟随技术演化进步。

从环境角度考量，算法在交通中的应用可以辅助人类进行一定的规避和识别操作，能够节省一定的人力物力。但相应的会产生一定的能源消耗，也对于汽车所使用的燃料有了进一步的环保要求。多使用清洁能源、倡导共享经济等举措均会预防这一后果。

第五章　总结与展望

5.1　工作总结

在本文中，采用了基于域自适应的 Faster R-CNN 模型来完成跨域的目标检测任务。使用这种方法，能够在不利用一切额外标记数据的情况下为新域取得较为有效的目标检测器。该算法以当前最先进的基础 Faster R-CNN 模型为根基。基于对域迁移的目标检测的理论解析，本文分别采用了一个图像级域自适应模块和一个实例级域自适应模块，以减缓因域偏移引起的性能表现降低，基于 H-散度对自适应模块实施对抗训练。之后为学习拥有域不变性能的区域建议网络再使用一致性正则化模块。另外，还加入了强对齐模型，只查看特征图的局部感受野，进一步降低了域偏移。此外，该模型可以使用标准的梯度下降法优化技术进行端到端的训练。旨在检验该自适应算法的有效性和鲁棒性，本文采用的模型在恶劣天气的域转移场景数据集中得到检验，此算法不论是与基线 Faster R-CNN 算法相比，还是与同任务多类型算法比较均有不同程度的优势，与基线算法比较而言有更为显著地提升，故此证明了算法在域迁移任务中目标检测方面的有效性。

5.2　未来展望

本文中所采用的基于域自适应的域迁移目标检测模型是基于双阶段目标检测算法基础上进行改进和添加的。然而模型的较为优良的检测结果的精确度带来了检测速度的下降的劣势。特别地，在某些对于需要即时反馈的检测任务，该模型的应用就有些过犹不及，譬如，自动驾驶实时检测目标、商场的人流统计等等。目前，域迁移的目标检测任务中将域自适应方法应用于单阶段目标检测算法的研究仍比较少。考虑到基于回归的单阶段 YOLO 算法的检测速度较快，将域自适应的方法与单阶段目标检测算法相结合的技术也将在可预见的未来发展、研究起来。针对不同种任务需求，研究者也需要因势利导，以任务目标为驱动进行不同算法的融合与改进。

参考文献

[1] KRIZHEVSKY A, SUTSKEVER I, HINTON G. ImageNet Classification with Deep Convolutional Neural Networks [J]. Advances in neural information processing systems, 2012, 25(2).

[2] GIRSHICK R, DONAHUE J, DARRELL T, et al. rich feature hierarchies for accurate object detection and semantic segmentation tech report (v5) [J]. 2017.

[3] GIDARIS S, KOMODAKIS N. Object Detection via a Multi-region and Semantic Segmentation-Aware CNN Model [J]. IEEE Computer Society, 2015.

[4] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single Shot MultiBox Detector [J]. 2015.

[5] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation [J]. IEEE Computer Society, 2013.

[6] 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, 2017, 39(6): 1137-49.

[7] REDMON J, DIVVALA S, GIRSHICK R, et al. You Only Look Once: Unified, Real-Time Object Detection [J]. IEEE, 2016.

[8] REDMON J, FARHADI A. YOLO9000: Better, Faster, Stronger [J]. IEEE, 2017: 6517-25.

[9] REDMON J, FARHADI A. YOLOv3: An Incremental Improvement [J]. arXiv e-prints, 2018.

[10] GIRSHICK R. Fast R-CNN [J]. arXiv e-prints, 2015.

[11] CAI Z, VASCONCELOS N. Cascade R-CNN: Delving into High Quality Object Detection [J]. 2017.

[12] LIN T Y, DOLLAR P, GIRSHICK R, et al. Feature Pyramid Networks for Object Detection [J]. IEEE Computer Society, 2017.

[13] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal Loss for Dense Object Detection [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, PP(99): 2999-3007.

[14] SUN B, FENG J, SAENKO K. Return of Frustratingly Easy Domain Adaptation [J]. 2015.

[15] LIXIN, DUAN, DONG, et al. Visual Event Recognition in Videos by Learning from Web Data [J]. Pattern Analysis & Machine Intelligence IEEE Transactions on, 2012.

[16] KULIS B, SAENKO K, DARRELL T. What you saw is not what you get: Domain adaptation using asymmetric kernel transforms; proceedings of the Cvpr, F, 2011 [C].

[17] GOPALAN R, LI R, CHELLAPPA R. Domain adaptation for object recognition: An unsupervised approach; proceedings of the IEEE International Conference on Computer Vision, ICCV 2011, Barcelona, Spain, November 6-13, 2011, F, 2011 [C].

[18] GHIFARY M, KLEIJN W B, ZHANG M. Domain Adaptive Neural Networks for Object Recognition [J]. Springer International Publishing, 2014.

[19] TZENG E, HOFFMAN J, ZHANG N, et al. Deep Domain Confusion: Maximizing for Domain Invariance [J]. Computer Science, 2014.

[20] LONG M, WANG J. Learning Transferable Features with Deep Adaptation Networks [J]. JMLRorg, 2015.
[21] GOODFELLOWIAN, POUGET-ABADIEJEAN, MIRZAMEHDI, et al. Generative adversarial networks [J]. Communications of the ACM, 2020.
[22] GANIN Y, LEMPITSKY V. Unsupervised Domain Adaptation by Backpropagation [J]. JMLRorg, 2014.
[23] DONG H, YU F, ZHAO J, et al. Unsupervised Domain Adaptation in Semantic Segmentation Based on Pixel Alignment and Self-Training [J]. 2021.
[24] ZHU J Y, PARK T, ISOLA P, et al. Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks [J]. IEEE, 2017.
[25] ARRUDA V F, PAIXãO T M, BERRIEL R F, et al. Cross-Domain Car Detection Using Unsupervised Image-to-Image Translation: From Day to Night [J]. IEEE, 2019.
[26] LIN C T. Cross Domain Adaptation for on-Road Object Detection Using Multimodal Structure-Consistent Image-to-Image Translation; proceedings of the 2019 IEEE International Conference on Image Processing (ICIP), F, 2019 [C].
[27] CHEN Y, LI W, SAKARIDIS C, et al. Domain Adaptive Faster R-CNN for Object Detection in the Wild [J]. IEEE, 2018.
[28] ZHU X, PANG J, YANG C, et al. Adapting Object Detectors via Selective Cross-Domain Alignment; proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), F, 2019 [C].
[29] SAITO K, USHIKU Y, HARADA T, et al. Strong-Weak Distribution Alignment for Adaptive Object Detection; proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), F, 2019 [C].
[30] YU F, ZHANG M, DONG H, et al. Dast: Unsupervised domain adaptation in semantic segmentation based on discriminator attention and self-training; proceedings of the Proceedings of the AAAI Conference on Artificial Intelligence, F, 2021 [C].
[31] RODRIGUEZ A L, MIKOLAJCZYK K. Domain adaptation for object detection via style consistency [J]. arXiv preprint arXiv:191110033, 2019.
[32] GEIGER A, LENZ P, STILLER C, et al. Vision meets robotics: The kitti dataset [J]. The International Journal of Robotics Research, 2013, 32(11): 1231-7.
[33] LIN T-Y, MAIRE M, BELONGIE S, et al. Microsoft coco: Common objects in context; proceedings of the European conference on computer vision, F, 2014 [C]. Springer.
[34] MCCULLOCH W S, PITTS W. A logical calculus of the ideas immanent in nervous activity [J]. The bulletin of mathematical biophysics, 1943, 5(4): 115-33.
[35] ROSENBLATT F. The perceptron: a probabilistic model for information storage and organization in the brain [J]. Psychological review, 1958, 65(6): 386.
[36] LECUN Y, BOTTOU L, BENGIO Y, et al. Gradient-based learning applied to document recognition [J]. Proceedings of the IEEE, 1998, 86(11): 2278-324.
[37] RUMELHART D E, HINTON G E, WILLIAMS R J. Learning representations by back-propagating errors [J]. nature, 1986, 323(6088): 533-6.
[38] LECUN Y, BOSER B, DENKER J S, et al. Backpropagation applied to handwritten zip code recognition [J]. Neural computation, 1989, 1(4): 541-51.
[39] DENG J, DONG W, SOCHER R, et al. Imagenet: A large-scale hierarchical image database; proceedings of the 2009 IEEE conference on computer vision and pattern recognition, F, 2009 [C]. Ieee.

[40] SCHROFF F, KALENICHENKO D, PHILBIN J. Facenet: A unified embedding for face recognition and clustering; proceedings of the Proceedings of the IEEE conference on computer vision and pattern recognition, F, 2015 [C].
[41] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [J]. arXiv preprint arXiv:14091556, 2014.
[42] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition; proceedings of the Proceedings of the IEEE conference on computer vision and pattern recognition, F, 2016 [C].
[43] HE K, GKIOXARI G, DOLLáR P, et al. Mask r-cnn; proceedings of the Proceedings of the IEEE international conference on computer vision, F, 2017 [C].
[44] WANG X, ZHANG R, KONG T, et al. Solov2: Dynamic, faster and stronger [J]. arXiv e-prints, 2020: arXiv: 2003.10152.
[45] VIOLA P, JONES M. Rapid object detection using a boosted cascade of simple features; proceedings of the Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition CVPR 2001, F, 2001 [C]. Ieee.
[46] DALAL N, TRIGGS B. Histograms of oriented gradients for human detection; proceedings of the 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05), F, 2005 [C]. Ieee.
[47] DAI J, LI Y, HE K, et al. R-fcn: Object detection via region-based fully convolutional networks [J]. Advances in neural information processing systems, 2016, 29.
[48] SERMANET P, EIGEN D, ZHANG X, et al. Overfeat: Integrated recognition, localization and detection using convolutional networks [J]. arXiv preprint arXiv:13126229, 2013.
[49] ZHANG Y, DAVID P, GONG B. Curriculum domain adaptation for semantic segmentation of urban scenes; proceedings of the Proceedings of the IEEE international conference on computer vision, F, 2017 [C].
[50] HE K, ZHANG X, REN S, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition [J]. IEEE transactions on pattern analysis and machine intelligence, 2015, 37(9): 1904-16.
[51] LIN T-Y, DOLLáR P, GIRSHICK R, et al. Feature pyramid networks for object detection; proceedings of the Proceedings of the IEEE conference on computer vision and pattern recognition, F, 2017 [C].
[52] ZHANG L, LIN L, LIANG X, et al. Is faster R-CNN doing well for pedestrian detection?; proceedings of the European conference on computer vision, F, 2016 [C]. Springer.
[53] DAI J, HE K, SUN J. Instance-aware semantic segmentation via multi-task network cascades; proceedings of the Proceedings of the IEEE conference on computer vision and pattern recognition, F, 2016 [C].
[54] FERNANDO B, HABRARD A, SEBBAN M, et al. Unsupervised visual domain adaptation using subspace alignment; proceedings of the Proceedings of the IEEE international conference on computer vision, F, 2013 [C].
[55] SAENKO K, KULIS B, FRITZ M, et al. Adapting visual category models to new domains; proceedings of the European conference on computer vision, F, 2010 [C]. Springer.
[56] SAITO K, WATANABE K, USHIKU Y, et al. Maximum classifier discrepancy for unsupervised domain adaptation; proceedings of the Proceedings of the IEEE conference on computer vision and pattern recognition, F, 2018 [C].

[57] GRETTON A, BORGWARDT K, RASCH M, et al. A kernel method for the two-sample-problem [J]. Advances in neural information processing systems, 2006, 19.
[58] BEN-DAVID S, BLITZER J, CRAMMER K, et al. A theory of learning from different domains [J]. Machine learning, 2010, 79(1): 151-75.
[59] BEN-DAVID S, BLITZER J, CRAMMER K, et al. Analysis of representations for domain adaptation [J]. Advances in neural information processing systems, 2006, 19.
[60] LONG M, CAO Z, WANG J, et al. Conditional adversarial domain adaptation [J]. Advances in neural information processing systems, 2018, 31.
[61] GANIN Y, USTINOVA E, AJAKAN H, et al. Domain-adversarial training of neural networks [J]. The journal of machine learning research, 2016, 17(1): 2096-30.
[62] GONG B, SHI Y, SHA F, et al. Geodesic flow kernel for unsupervised domain adaptation; proceedings of the 2012 IEEE conference on computer vision and pattern recognition, F, 2012 [C]. IEEE.
[63] DUAN L, TSANG I W, XU D. Domain transfer multiple kernel learning [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 34(3): 465-79.
[64] LU H, ZHANG L, CAO Z, et al. When unsupervised domain adaptation meets tensor representations; proceedings of the Proceedings of the IEEE international conference on computer vision, F, 2017 [C].
[65] LIU M-Y, BREUEL T, KAUTZ J. Unsupervised image-to-image translation networks [J]. Advances in neural information processing systems, 2017, 30.
[66] CHEN Y, LI W, VAN GOOL L. Road: Reality oriented adaptation for semantic segmentation of urban scenes; proceedings of the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, F, 2018 [C].
[67] GEBRU T, HOFFMAN J, FEI-FEI L. Fine-grained recognition in the wild: A multi-task domain adaptation approach; proceedings of the Proceedings of the IEEE international conference on computer vision, F, 2017 [C].
[68] RAJ A, NAMBOODIRI V P, TUYTELAARS T. Subspace alignment based domain adaptation for rcnn detector [J]. arXiv preprint arXiv:150705578, 2015.
[69] TANG K, RAMANATHAN V, FEI-FEI L, et al. Shifting weights: Adapting object detectors from image to video [J]. Advances in Neural Information Processing Systems, 2012, 25.
[70] PENG X, SUN B, ALI K, et al. Learning deep object detectors from 3d models; proceedings of the Proceedings of the IEEE international conference on computer vision, F, 2015 [C].
[71] HATTORI H, NARESH BODDETI V, KITANI K M, et al. Learning scene-specific pedestrian detectors without real data; proceedings of the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, F, 2015 [C].
[72] SZEGEDY C, LIU W, JIA Y, et al. Going deeper with convolutions; proceedings of the Proceedings of the IEEE conference on computer vision and pattern recognition, F, 2015 [C].
[73] JOHNSON J, ALAHI A, FEI-FEI L. Perceptual losses for real-time style transfer and super-resolution; proceedings of the European conference on computer vision, F, 2016 [C]. Springer.

[74] CORDTS M, OMRAN M, RAMOS S, et al. The cityscapes dataset for semantic urban scene understanding; proceedings of the Proceedings of the IEEE conference on computer vision and pattern recognition, F, 2016 [C].
[75] SAKARIDIS C, DAI D, VAN GOOL L. Semantic Foggy Scene Understanding with Synthetic Data [J]. International Journal of Computer Vision, 2017.
[76] EVERINGHAM M, GOOL L V, WILLIAMS C, et al. The Pascal Visual Object Classes (VOC) Challenge [J]. International Journal of Computer Vision, 2010, 88(2): 303-38.
[77] NARASIMHAN S G, NAYAR S K. Vision and the Atmosphere [J]. International Journal of Computer Vision, 2002, 48(3): 233-54.
[78] HOIEM D, CHODPATHUMWAN Y, DAI Q. Diagnosing Error in Object Detectors; proceedings of the Proceedings of the 12th European conference on Computer Vision - Volume Part III, F, 2012 [C].
[79] KIM T, JEONG M, KIM S, et al. Diversify and match: A domain adaptive representation learning paradigm for object detection; proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, F, 2019 [C].
[80] HSU H-K, YAO C-H, TSAI Y-H, et al. Progressive domain adaptation for object detection; proceedings of the Proceedings of the IEEE/CVF winter conference on applications of computer vision, F, 2020 [C].
[81] ZHAO Z, GUO Y, SHEN H, et al. Adaptive object detection with dual multi-label prediction; proceedings of the European Conference on Computer Vision, F, 2020 [C]. Springer.
[82] SINDAGI V A, OZA P, YASARLA R, et al. Prior-based domain adaptive object detection for hazy and rainy conditions; proceedings of the European Conference on Computer Vision, F, 2020 [C]. Springer.
[83] CHEN C, ZHENG Z, DING X, et al. Harmonizing transferability and discriminability for adapting object detectors; proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, F, 2020 [C].

致　谢

当我长久地审视这篇也许是在天津大学的最后一篇论文时，我也在长久地审视着自己的这四年大学时光。都说人有两重黑暗性，即对自己内心黑暗性的认知匮乏，以及对这个世界黑暗性的了解缺失。正是这四年的经历，天津大学和它里面这些可爱、可敬的人们让我对这两重缺失有了重新地认识。他们不仅仅以授业为己任，更注重传道解惑。故而，内心涌动的感激让我不自觉且必要地写下这些感恩的文字。

关于天大，走入天大的时候，我带着巨大的朦胧与困顿，如今即将离开它，我带走的是从容和自信。在这里，我学会了如何与自己的情绪和解，感受了专业中的乐与美，发现了生活的诸多意义.....于是，走出卫津路的大门，我的身上从此铭刻着“天大人”的这块历久弥新的铭牌。

关于我的老师们，周圆老师是我数字信号处理的任课老师和毕设的指导老师，从老师那里收获最多的不只是模拟、数字的理论推演，更有对于个人职业道路上的灼见和学术中的平和心态。每一次与老师的谈话都成为自我思考的开端。以及本科道路上遇到的每一位“指引者”，在我继续前行的征途中也蕴藏着他们教导的影子。于是，终将消散的我们带着不朽的智慧行于世间，终其一生。

关于我的师兄师姐，在遇到这群学长学姐前，我逃避错误；在遇到他们之后，我喜欢上了犯错，这种促使我快速进步的“惩罚”。闫志宇学长稳定的知识输出、尚小纯学姐细心的建议、张启源学长无比耐心地修改，这些美好的时光让我时常阴天的毕设进程中总有拨云见日的结局。希冀他们也都能拥有黎明包围的日子，充满着温暖的色彩。

关于我的同学们，是他们一次次从阴影将我中拥入怀中，让我重获信心与活力。他们同时也是最直白的批判者，规劝着我的一时一刻。铁人紫琪老师、老好人梦雅、雨晴、李浩宇，或许我们都即将各奔前程、难再聚首，但那些一起度过的日子令我们永恒的友谊不证自明。也同样期冀着多年后再相聚时，各有各的皎洁。

关于我的亲人，我的父母、我的姐姐还有我五位可爱的室友，像春日里拂过的风，时刻陪伴着我、默默地包容着我、无条件地支持着我。如同他们对我深沉的关爱，我也同样交关偏爱于他们，私心期许神明也偏爱他们，百事从欢。

在天大的四年至此业已结束，但每一次的结束同时也意味着新的开始。竚念于此，欢欣和伤感同怀我心，或许前路茫茫，坎坷依旧，慢慢来，不要急，满载敏感的爱和温柔的智慧慢慢走，时间会给出答案。