

中图分类号:

UDC:

学校代码: 10055

密级: 公开

南开大学
硕士学位论文

基于深度霍夫变换的语义线检测

Deep Hough Network for Semantic Line Detection

论文作者	韩琦	指导教师	程明明教授
申请学位	硕士	培养单位	南开大学
学科专业	计算机技术	研究方向	计算机视觉
答辩委员会主席		评阅人	

南开大学研究生院

二〇二二年四月

南开大学学位论文使用授权书

本人完全了解《南开大学关于研究生学位论文收藏和利用管理办法》关于南开大学(简称“学校”)研究生学位论文收藏和利用的管理规定,同意向南开大学提交本人的学位论文电子版及相应的纸质本。

本人了解南开大学拥有在《中华人民共和国著作权法》规定范围内的学位论文使用权,同意在以下几方面向学校授权。即:

1. 学校将学位论文编入《南开大学博硕士学位论文全文数据库》,并作为资料在学校图书馆等场所提供阅览,在校园网上提供论文目录检索、文摘及前16页的浏览等信息服务;
2. 学校可以采用影印、缩印或其他复制手段保存学位论文;学校根据规定向教育部指定的收藏和存档单位提交学位论文;
3. 非公开学位论文在解密后的使用权同公开论文。

本人承诺:本人的学位论文是在南开大学学习期间创作完成的作品,并已通过论文答辩;提交的学位论文电子版与纸质本论文的内容一致,如因不同造成不良后果由本人自负。

本人签署本授权书一份(此授权书为论文中一页),交图书馆留存。

学位论文作者暨授权人(亲笔)签字: _____

20 年 月 日

南开大学研究生学位论文作者信息

论 文 题 目	基于深度霍夫变换的语义线检测				
姓 名	韩琦	学号	2120190459	答辩日期	
论 文 类 别	博士 <input type="checkbox"/> 学历硕士 <input type="checkbox"/> 专业学位硕士 <input checked="" type="checkbox"/> 同等学力硕士 <input type="checkbox"/> 划 <input checked="" type="checkbox"/> 选择				
学院(单位)	计算机学院		学科/专业(专业学位)名称		计算机技术
联系电话	18292813106		电子邮箱	qhan@mail.nankai.edu.cn	
通讯地址(邮编): 300381					
非公开论文编号				备注	

注:本授权书适用我校授予的所有博士、硕士的学位论文。如已批准为非公开学位论文,须向图书馆提供批准通过的《南开大学研究生申请非公开学位论文审批表》复印件和“非公开学位论文标注说明”页原件。

南开大学学位论文原创性声明

本人郑重声明：所提交的学位论文，是本人在导师指导下进行研究工作所取得的研究成果。除文中已经注明引用的内容外，本学位论文的研究成果不包含任何他人创作的、已公开发表或者没有公开发表的作品的内容。对本论文所涉及的研究工作做出贡献的其他个人和集体，均已在文中以明确方式标明。本学位论文原创性声明的法律责任由本人承担。

学位论文作者签名： _____ 年 月 日

非公开学位论文标注说明

(本页表中填写内容须打印)

根据南开大学有关规定，非公开学位论文须经指导教师同意、作者本人申请和相关部门批准方能标注。未经批准的均为公开学位论文，公开学位论文本说明为空白。

论文题目			
申请密级	<input type="checkbox"/> 限制 (≤2 年)	<input type="checkbox"/> 秘密 (≤10 年)	<input type="checkbox"/> 机密 (≤20 年)
保密期限	20 年 月 日至 20 年 月 日		
审批表编号		批准日期	20 年 月 日

南开大学学位评定委员会办公室盖章 (有效)

注：限制 ★2 年 (可少于 2 年); 秘密 ★10 年 (可少于 10 年); 机密 ★20 年 (可少于 20 年)

摘要

辅助构图技术作为辅助摄影、美学评估的重要组成部分，开始引起科研界的关注。本文主要针对辅助构图中重要的语义线结构展开研究，并探究、设计语义线检测的方法流程。现有的语义线检测方法将语义线检测看作是特殊的目标检测，并沿用目标检测的基本流程，忽略了直线的本质结构信息，导致检测结果欠佳，效率较低。本文将利用直线本质特性的霍夫变换与具有强大表征能力的深度神经网络相结合，构建了一个全新的端到端训练的语义线检测流程。针对语义线检测的数据集和评价指标加以分析和总结，提出了全新的数据集和评测指标。具体地，本文的主要贡献总结如下：

(1) 构建了全新的深度霍夫变换网络，利用深度霍夫变换将图像的深度特征表征转换为霍夫参数空间中的特征表达。因此，在图像中进行语义线检测的任务被转化为在参数空间检测点的任务，并将复杂耗时的非极大值抑制过程，转化为了在参数空间求连通区域质心的简单问题。此外，本文提出了一个边缘引导的线微调的后处理步骤，从而进一步修正检测结果。

(2) 指出了语义线检测任务中的基于交并比的直线相似性评价存在的不一致和歧义性问题，并分析了 EMD 距离、CD 距离等常用的距离度量指标在语义线检测任务中的适用性。然后，提出了一种全新的应用于语义线间相似度度量的评价指标，具有直观的可解释性且避免了歧义性的产生，并提出了基于二分图的最佳匹配过程，提升了评测流程的合理性。

(3) 分析了现有语义线检测数据集，指出了现有数据集存在的样本量少、场景单一的问题，提出了用于语义线检测任务的大规模、场景丰富的数据集，并对数据集进行了详细的分析与统计。

(4) 进行了大量的定量分析和定性分析实验，验证了提出的深度霍夫网络在语义线检测任务中的有效性与高效性，同时展示了在不同实现平台和 GPU 硬件设备上的推理速度。最后，呈现了深度霍夫网络在自然场景下的检测结果和实际应用中的效果，证明了该方法具有潜在的实际应用价值。

关键词： 语义线检测；深度霍夫网络；霍夫变换；神经网络

Abstract

Computer algorithm assisted photography gains widespread attention. As an important part of assisted photography and aesthetic evaluation, auxiliary composition technology has begun to attract investment in the scientific research community. This thesis mainly studies the important semantic line structure in auxiliary composition. Existing semantic line detection methods regard semantic line as a special object, and follow the basic pipeline of object detection, ignoring the essential structural information of lines, resulting in poor detection performance. This thesis incorporates the traditional hough transform, which exploits the essential properties of lines, with the powerful representation capabilities of deep neural network, constructing a new end-to-end training pipeline for semantic line detection. Besides, the evaluation metric and dataset for semantic line detection are analyzed. And new metric and dataset are proposed in this thesis. The main contributions are summarized as:

(1) Constructing a deep hough network which uses the deep hough transform to convert the deep representations of an image to the parameter space. Therefore, the task of semantic line detection in images is transformed to the task of detecting points in the parameter space, and the time-consuming non-maximum suppression is converted to a simple problem of finding the centroid of connected regions in the parameter space. At the same time, this thesis proposes an edge-guided refinement post-processing step, which can further modify the semantic line detection results.

(2) Analyzing the existing evaluation metric for semantic line detection task, pointing out its inconsistency and ambiguity, and then analyzes the applicability of commonly used distance measures for semantic line detection. Finally, this thesis proposes a new evaluation metric to assess the quality of semantic line detection and employs the bipartite graph matching to match ground-truth and detected lines during evaluation, removing redundant true positives so that each ground-truth line is associated with at most one detected line and vice versa.

(3) Analyzing the exists dataset for semantic line detection, and giving a new con-

structured large-scale dataset, which contains large number of samples and various scenes.

(4) Conducting a large number of quantitative and qualitative experiments, verifying the effectiveness and efficiency of the proposed deep hough network. The inference speed on different platforms and GPU devices are also analyzed. The detection results of the deep hough network in natural scenes and the application results are shown, illustrating the potential practical application value of this method.

Key Words: Semantic line detection; Deep hough transform; Hough transform; Deep neural network

目录

摘要	I
Abstract	II
第一章 绪论	1
第一节 研究背景	1
第二节 国内外研究现状	2
1.2.1 直线检测研究现状	2
1.2.2 语义线检测研究现状	3
第三节 本文的主要研究内容	5
第四节 章节安排	7
第二章 语义线检测的相关技术	9
第一节 边缘检测相关技术	9
第二节 霍夫变换与直线检测	10
第三节 目标检测相关技术	12
第四节 语义线检测相关方法	14
第五节 本章小结	14
第三章 基于深度霍夫变换的语义线检测方法	16
第一节 直线参数化和反参数化	16
第二节 霍夫变换基线方法	17
第三节 深度霍夫网络 DHN	18
3.3.1 深度霍夫变换 DHT	18
3.3.2 基于 FPN 的多尺度 DHT	19
3.3.3 参数空间中的线检测	20
3.3.4 损失函数	21
3.3.5 逆向映射	22
3.3.6 边缘引导的线微调	22
第四节 本章小结	24

第四章 语义线检测的数据集与评价指标	25
第一节 语义线检测的数据集	25
4.1.1 现有数据集总结与新数据集的提出	25
4.1.2 NKL 数据集与统计分析	26
第二节 语义线检测的评测指标	29
4.2.1 现有度量标准	30
4.2.2 本文提出的评价标准	32
第三节 语义线检测的评测过程	33
第四节 本章小结	36
第五章 实验与分析	37
第一节 实验环境与实现细节	37
5.1.1 实验环境	37
5.1.2 网络结构	37
5.1.3 超参数设置	37
5.1.4 数据集和数据增强	38
5.1.5 评价指标	38
第二节 对比实验	38
5.2.1 量化间隔的确定	38
5.2.2 定量对比实验	39
5.2.3 定性比较	41
5.2.4 消融实验	43
5.2.5 开放自然场景检测结果	46
第三节 方法的应用	47
第四节 本章小结	49
第六章 总结展望	50
第一节 文章总结	50
第二节 未来展望	51
参考文献	52
致谢	58
个人简历	59

第一章 绪论

第一节 研究背景

人类的信息获取途径多种多样，如视觉、听觉、嗅觉等，然而，绝大多数信息是通过视觉系统进行获取。早期人们主要依赖于视觉信息进行自然世界的理解、信息的传递，随着人们生活质量的不断提升，对视觉美学的追求也逐步增加，摄影作为一门艺术开始逐渐走进千家万户。然而，摄影作为一门学科，有着复杂的知识体系，要产生一个好的摄影作品，需要具备充分的理论知识与艺术积淀，同时需要花费大量的时间来学习构图等摄影知识。随着计算机视觉技术的发展和电子成像技术的进步，人们获取数字图像变得愈发便捷，通过简单的手机拍摄即可记录日常生活的场景。相应地，基于计算机视觉技术的图像处理技术飞速发展，为快节奏生活中的人们提供了便捷快速的图片美化工具，如专业的修图软件和便捷的美图应用程序。然而，这些成熟的技术往往用于拍摄后的照片美化，对于拍摄时的构图难以提供有效帮助。而大多数人没有时间和精力去专门学习摄影构图的方法，又希望能够拍摄出优美的照片。因此，辅助构图的技术应运而生，且受到了广泛关注^[1, 2]。

早期应用在艺术美学方面的计算机视觉技术，主要关注于图像质量的评估，包括成像质量的评估和美学质量的评估。后来随着需求的不断增加，研究方向逐渐演化为图像的增强和美学的优化。而美学优化的一个重要研究分支为辅助构图技术，辅助构图为用户在构图阶段提供指导，或将用户拍摄完的照片根据摄影学的规则和图像中的内容语义进行重构图，将摄影师和美学艺术家的研究理论和经验运用在建模计算中，利用人工智能、美学、心理学等技术手段共同实现。利用人工智能技术的辅助构图方法，一般基于图像中的边缘、直线、显著物体等结构进行图片分析，并结合摄影学技术进行辅助构图。近些年，摄影视图推荐和局部构图检测方法也被提出，其利用图像的语义信息和特殊结构与摄影构图中的三分法则、视觉平衡规则等为用户提供更好的构图方案^[2, 3]。本文将主要关注用于辅助构图的线结构，这种线结构是分割了图像中两个语义区域的直线，或具有明显边缘特性的线结构，被称为语义线。早期的直线检测方法基本依

依赖于边缘检测，通过对图像进行边缘检测，给出边缘结果图，然后再对边缘图后处理，进行霍夫变换提取图像中的直线。基于深度学习的语义线的提取近两年来受到了关注，最初由 Lee 等人在 2017 年被详细定义^[4]，并给出了基于目标检测网络的解决方案。基于深度学习的语义线检测方法，沿用了目标检测深度网络模型流程，通过两阶段的提案分析与分类回归完成直线的检测，这些做法忽略了直线的本质特性，检测结果不够准确，且采用迭代的方式进行语义线的检测，计算效率较低。而基于霍夫变换的传统直线检测方法不依赖于监督训练，容易受到噪声的干扰，检测效率低且效果差。此外，用于衡量语义线检测结果的评测体系与评价指标，存在不合理之处，对于检测结果的评测会产生歧义性。

基于深度学习的检测任务往往依赖于大量的训练样本，现有的语义直线检测数据集较少，且数据集样本量不充足，场景过于单一。现有的 SEL 数据集仅包含 1,715 张自然图像，对依赖大量样本的深度网络而言略显不足。本文在语义线检测的深度学习方法与数据集都较为匮乏的背景下，将传统的霍夫变换方法与深度神经网络相结合，提出了深度霍夫变换模块，将深度网络表征转化到霍夫参数空间，同时在参数空间设计直线检测器，完成语义线的检测，与多尺度神经网络相辅相成，共同构成了语义线检测的深度学习新范式。此外，本文针对现有数据集匮乏的问题，提出了全新的用于语义线检测的 NKL 数据集，包含大量的训练样本和丰富的场景。本文同样分析了语义线检测结果的评测体系与评价指标，指出了其中存在的问题，并设计了全新的用于图像中直线相似度的评价指标，并对现有评价体系进行了完善和统一。

第二节 国内外研究现状

本文的核心研究内容为语义线检测，而语义线是一种特殊的直线结构，在本节中，将介绍直线检测与语义线检测的国内外研究现状，并分析指出现有方法存在的问题。

1.2.1 直线检测研究现状

数字图像中的直线检测在计算机视觉中有着悠久的历史。线结构是将视觉信号转换为可解释、可理解的知识的过程中必不可少的中间概念，从很早就开始被研究^[5]。有非常多的研究工作被提出应用于图像中的直线检测，其中霍夫变换（Hough Transform, HT）是具有代表性的直线检测技术，初被提出来检测

气泡室照片中的直线^[6]，其关键思想是将置信度从图像域投影到参数域，然后在参数域中通过识别局部最大响应来检测参数域中的形状。在线检测中，图像域中的一条线可由其参数表示，即参数空间中的斜率和截距。霍夫变换在图片中沿直线收集置信度并在参数空间中累积置信度到单个点。因此，图像域中的线检测被转换为了检测参数空间中的峰值响应问题。由于其简单性和有效性，HT 被用来检测数字图像中的线^[7]，并被进一步扩展用于检测其他规则图形如圆和矩形^[8]。同时，自适应的 HT 被提出用于降低霍夫变换的计算复杂度^[9]。霍夫变换随后被大量推广与发展，形成了有代表性的三个分支：广义霍夫变换、概率霍夫变换、随机霍夫变换。广义霍夫变换主要用于求解非多项式参数化的形状检测问题，而概率霍夫变换和随机霍夫变换主要解决了霍夫变换中较高的计算复杂度和内存开销问题^[10, 11]。在直线检测当中，概率霍夫变换对该问题产生了深远影响，降低了计算复杂度的同时，让霍夫变换能够对图像中的线段进行检测，而不仅仅局限于直线。而后，大多数研究开始关注线段的检测，并将线段检测应用在了许多工业检测与定位问题中。比较有代表性的工作为 LSD 线段检测器^[12]，LSD 采用负样本抑制的思路，避免了大量的参数选择，具有很高的实用性与鲁棒性，快速占领了线段检测的统治地位。

然而，传统意义上的直线结构往往与边缘相关，一般通过边缘检测技术获取图像的边缘信息，然后再提取直线。如霍夫变换和 LSD 都依赖于边缘图的输入。这些传统直线检测方法，一般通过拉普拉斯算子、Canny^[13] 等快速计算的方法得到边缘图。然而，Canny 之类方法得到的边缘图往往过于杂乱，难以提取出显著的边缘结构，不能适用于语义线检测任务。

1.2.2 语义线检测研究现状

语义线作为一种显著的线结构，分割了图像中不同的语义区域并标示出显著的边缘结构。尽管很多技术被提出以检测显著物体^[14-18] 和显著区域^[19-21]，但是很少有关于检测突出的、显著的线结构的工作。最近的一项研究提出了检测显著的直线结构的问题，并将这些直线称为“语义线”^[4]。识别语义线对于计算机图形学和计算机视觉应用至关重要，例如摄影构图^[1, 22]、结构保留的图像处理^[23, 24]、美学评价^[25-28]、车道检测^[29] 和艺术创作^[30-33]，都依赖于显著的线结构信息。

如图1.1所示，Liu 等人提出根据黄金分割率使用“突出线”裁剪图片^[1]。检测这些语义线可以帮助拍出具有视觉上令人愉悦的摄影构图的图像，其中 (a)



图 1.1 基于语义线的摄影构图示例，图片来源于 Liu 等人的工作^[1]。(a): 任意角度拍摄的图像。(b): 符合黄金分割率原则^[30, 34]的照片，该照片根据 Liu 等人工作中^[1]描述的方法使用图片中所谓的“突出线”获得。(c): 本文提出的深度霍夫网络的检测结果是清晰的，仅包含对摄影构图有潜在帮助的少数有意义的线。(d): 使用经典线检测算法得到的线检测结果通常只集中在精细的细直线边缘上。

图为原始图像，(b) 图为利用“突出线”进行的重构图结果。而突出线就是本文阐述的语义线，难以通过传统的方法检测得到。图1.1(d)所示传统直线检测方法的结果通常都有噪声并且包含不相关的线^[35]，这些线是先通过传统的 Canny 边缘检测算子进行边缘检测，然后使用霍夫变换进行直线检测得到。因为没有对图像的语义信息加以分析，这些方法难以判断哪些线是突出的线，那些是噪声引起的杂乱线。图1.1(c)展示的是本文提出的语义线检测方法的检测结果，能够给出突出显著的直线结果，对构图等下游任务有重要意义。

卷积神经网络 (CNN) 在多种计算机视觉任务中均取得了显著成功。最近的部分研究^[4, 36]提出了基于 CNN 的方法进行线检测。具体来说，他们将线检测视为一种特殊的目标检测，并且使用现有的目标检测算法，如 Faster R-CNN^[37] 或 CornerNet^[38]，进行线检测。基于 Faster R-CNN 的直线检测方法，通过 ROI 池化进行特征汇聚，最后的检测也需要非最大值抑制来进行后处理。受限于对线的

ROI 池化和非最大抑制，这两种方法在运行效率方面都比较低效。此外，ROI 池化^[39] 仅沿着一条直线聚集特征，而许多最近的研究表明丰富的上下文信息对许多任务是至关重要的，如视频分类^[40] 和语义分割^[41] 等。因此仅汇聚一条线的特征将导致线与线之间的交互关系丢失，检测结果较差。

语义线检测的数据集和评测指标也都是近两年来被提出的^[4]，其中公开的数据集 SEL 仅包含 1,715 张图像的训练样本，对于大量数据依赖的深度学习而言略显不足，且仅包含了室外的风景场景，单一的场景容易导致训练得到的模型在自然场景中泛化能力较差。而应用在语义线检测的评测指标依赖于直线与直线的交并比来计算，这种计算方式与语义线出现在图像中的位置相关，且存在突变的情况，导致了评测不一致的歧义性。

第三节 本文的主要研究内容

本文聚焦于如何高效、准确地进行语义线检测。由于传统基于霍夫变换和边缘检测的方法会缺失对图像语义的分析，难以准确定位出显著线结构，而现有的基于深度学习的语义线检测方法依赖于目标检测的两阶段流程，忽略了直线与直线之间的关联性，检测效率低且不够准确。将神经网络强大的特征提取能力与霍夫变换对直线检测的能力相结合，将是一个语义线检测的新方向。

一个简单将 CNN 与霍夫变换组合的方式是使用基于 CNN 的边缘检测器进行边缘检测，例如 RCF 或 HED 方法^[42, 43]，然后在边缘图上应用标准霍夫变换来检测直线。但是，这两个组成部分有不同的优化目标，会导致结果欠优。且边缘检测网络一般专门为低层的图像信息处理与提取而设计，对于语义线这种依赖于高层级的语义信息的特殊结构的检测结果较差。

本文提出了一种将霍夫变换以端到端的形式整合到 CNN 中的方法，从而使得本文所设计的方法中的每个部分都有相同的优化目标，并称之为深度霍夫网络。深度霍夫网络首先使用基于 CNN 的编码器提取像素级别特征表示，然后在深度特征表示上，提出了用于将特征表示从图像空间中转换到参数空间中的深度霍夫变换。然后语义线检测问题就转换为了在变换后的特征中简单地逐点检测的问题，使得问题变得更加简单。在深度霍夫变换的参数空间中，每一个参数空间的点，都表达了原始图像空间中一条直线的语义特征，在参数空间进行特征提取操作对应了多条直线间的语义信息交互，更好地利用了直线的本质特性进行了特征增强。因此，最终的决策不仅取决于一条线，也取决于周围线的信

息，有利于最终的检测问题的优化。同时，复杂耗时的非最大抑制过程可被简单地替换为在参数空间中计算相邻区域的质心，这使得深度霍夫网络具有非常高效的执行效率，可以实时地进行语义线检测。在深度霍夫网络的基础上，本文又提出了全新的“边缘引导的线微调”后处理模块来调整线的位置，为深度霍夫网络提供了可选择的超精细语义线结果修正的后处理方法，使其在精细的边缘信息的协助下能够获得更好的线检测结果。

尽管 Lee 等人提出了使用直线交叉面积与总面积的比值来衡量一对线的相似性的评价指标^[4]，但是这一基于交并比的度量指标会因为两直线出现在图像中的位置导致歧义性，或因为直线的角度处于垂直状态产生模棱两可的结果。为了更好地评估语义线检测方法，本文分析了现有的用于三维图形相似度量度的 CD (Chamfer Distance, CD) 距离和 EMD (Earth Mover's Distance, EMD) 距离，并将其用于直线相似性度量，但发现其评测复杂，需要对直线栅格化然后逐像素地去求解。为了给出一种简单有效的评测两图像中直线相似度的方法，本文引入了基于参数化直线的角度偏差和距离偏差的 EA Score，并将其用于语义线的检测结果评测，给出了符合现有 CD 距离和 EMD 距离度量的结果，同时具有直观的可解释性，计算更简洁，速度更快。同时在整个评测流程中，采用二分图最大匹配来匹配真实标注和检测到的线。该过程剔除了冗余的真阳性样本的匹配，以便于使每一条真实的线最多与一条检测到的线相关联，使评测流程更合理，不存在歧义性。

最后，本文收集了由 6,500 张精心标注的图片构成的一个大规模数据集用于语义线检测。该数据集被称为 NKL，包括多种多样的场景，并且在图片数量和标注的线方面比现有的 SEL^[4] 数据集规模大很多。该数据集对于语义线检测任务的发展提供了数据支撑，提升了基于监督训练的模型准确率和泛化能力。

综合以上，本文的贡献可总结为以下几点：

(1) 本文对现有的直线检测和语义线检测方法及其相关研究进行了充分的总结与阐述，分析了现有语义线检测方法存在的不足。

(2) 本文提出了一个全新的端到端训练框架，将神经网络的特征学习能力与利用直线本质特性的霍夫变换相结合，构成了深度霍夫网络 DHN。该网络以深度霍夫变换为核心，高效地在霍夫参数空间中进行特征交互，得到了高效的实时语义线检测方案。同时可以增加边缘引导的微调 (ER) 模块，使得检测结果更精细。

(3) 为了促进语义线检测有关的研究，本文收集、使用了 6,500 张图片构建了一个新的数据集 NKL，其规模和多样性都优于现有的 SEL 数据集^[4]。

(4) 本文提出了合理的评价标准 EA Score 用以度量两条线的相似性。相比于原来的基于直线交并比的评价标准，本文设计的评价标准有直观的解释并且易于实现。

(5) 公开数据集上的评估结果表明，本文提出的深度霍夫网络方法在准确率和计算效率上都大幅度超过现有的方法。

图1.2展示了本文的主要研究内容，包括现有方法存在的缺陷与不足以及语义线检测任务面临的问题与挑战、本文的核心理论创新点、实验结果与方法应用等几个部分。

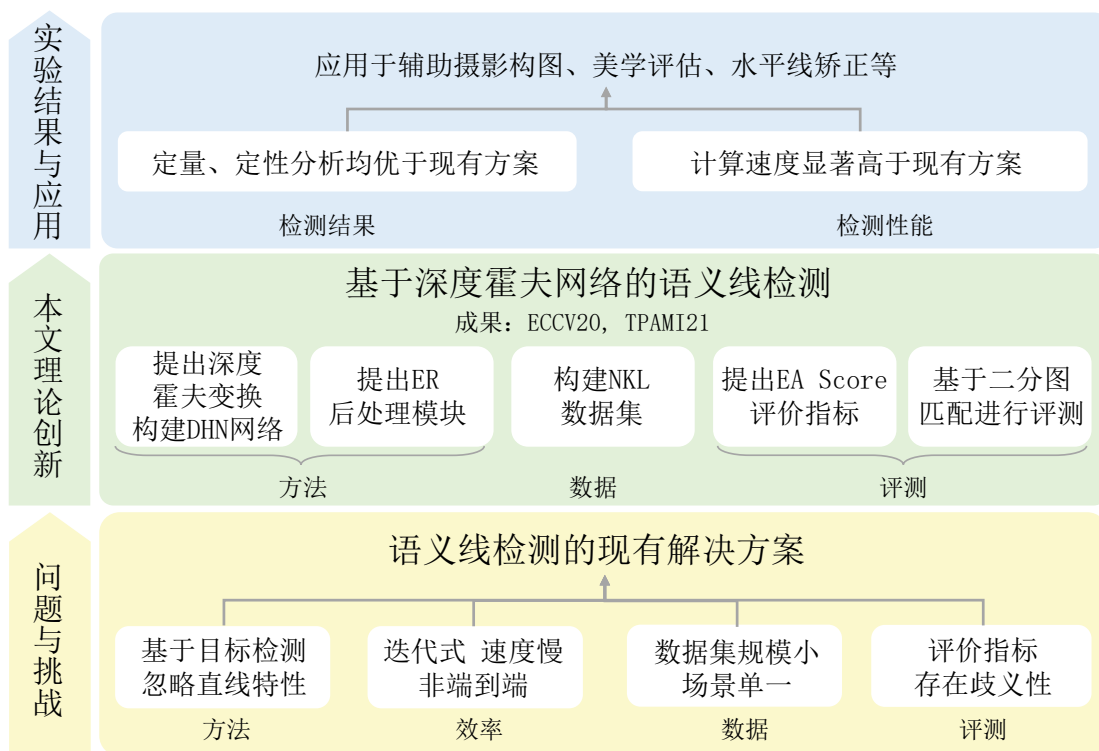


图 1.2 文章的主要研究内容

第四节 章节安排

第一章：首先介绍了研究背景和意义，概括了语义线检测任务的来源与意义，分析了国内外在直线检测和语义线检测领域的研究现状，并指出了其中的不足。概括了本文的主要研究内容与贡献，介绍了文章整体的章节安排。

第二章：分类介绍了语义线检测任务的相关研究技术，包括霍夫变换及其变种、边缘检测技术、目标检测技术以及语义线检测技术。为文章方法的研究与设计提供了理论基础。

第三章：详细阐述了本文提出的深度霍夫网络新流程，介绍了其中包含的直线参数化与反参数化过程、霍夫变换和本文提出的深度霍夫网络的整体流程。对深度霍夫网络中的深度霍夫变换、多尺度霍夫变换、参数空间检测器、网络的损失函数与优化、结果的逆向映射以及边缘引导的线微调模块进行了详细介绍。

第四章：介绍了现有的语义线检测数据集，并对其加以分析，指出其中存在的不足。提出了本文的 NKL 语义线检测数据集，并进行了详细地数据分析与对比。阐述了语义线检测的现有评测方法和评测流程，提出了用于衡量图像中两直线相似度的 EA Score，并分析了其合理性，对现有的评价体系做出了修正，提出了基于二分图匹配的检测结果评价流程。

第五章：介绍了实验环境与实现细节，进行了详细的对比试验与分析，通过定量实验和定性分析验证了本文提出的深度霍夫网络的有效性与高效性并对提出的方法进行了充分的消融实验，验证了深度霍夫网络每个组成部分的有效性。最后对基于本文语义线检测方法的实际检测结果与应用进行了简单分析。

第六章：对本文的工作内容进行总结，分析了文章提出的语义线检测新流程存在的不足，并提出未来可进一步改进的方向。

第二章 语义线检测的相关技术

早期数字图像中线结构的研究主要以直线检测和线段检测为主。而霍夫变换占据了直线检测的主导地位，基于霍夫变换的直线检测主要分为边缘图的检测和霍夫变换两部分。因此，本章中将介绍边缘检测的相关研究技术和霍夫变换的基本方法及其变种，为后续文章中的方法和实验提供了理论基础并构建了基线方法。而现有的语义线检测方法都是依赖于深度学习目标检测流程的方法，本章也将介绍目标检测的相关技术研究，并将详细介绍语义线检测的现有工作方法。

第一节 边缘检测相关技术

边缘检测是一项用于定位出自然图像中具有明显的强度或色彩变换的像素点，并构成连续的曲线的图像处理技术。边缘作为图像中的基本元素，很早就被研究者广泛研究，并广泛应用于物体检测和分割领域。边缘检测技术可以分为三类：传统的边缘检测方法、基于机器学习的边缘检测方法、基于深度学习的边缘检测方法。

传统的边缘检测方法一般通过判断图像中像素的灰度值和颜色、纹理等的突变性来确定边缘。早期一些工作基于滤波和微分的方法来求取图像的边缘，如基于一阶导数的 Sobel 算子，通过计算图像的微分来寻找边缘^[44]，也有基于二阶导数的方法，如 Laplacian 算子和 Canny 算子^[13]。Torre 等人对各种算子进行了充分的分析与论证^[45]。Canny 边缘检测方法首先对图像进行平滑滤波，然后使用八个方向的梯度算子计算图像灰度图的梯度强度和方向，最后对得到的边缘进行非极大值抑制和基于双阈值的连接算法对断开的边缘进行连接，得到比较清晰干净的边缘结果图。由于其出色的检测效果，成为了广泛使用的传统边缘检测方法。

而以上这些方法都是基于图像强度的计算方法，在特定的干扰下表现效果不足。随着机器学习方法的兴起，Martin 等人提出了使用图像亮度、颜色和纹理的图像局部评估方法^[46]，使用基于样本训练的线性分类器，结合亮度、颜色和纹理三个维度的信息来判别该位置是否为边缘，最终得到图像的边缘图。另外

一个基于学习的有代表性的方法，是 Dollar 等人提出的基于局部结构特征的随机决策树方法^[47]，使用样本训练离散的决策树表达，并快速准确地进行边缘检测，取得了当时在 BSDS500 边缘检测数据集上的最优结果。

随着深度学习的发展，大量基于深度学习的边缘检测研究工作被提出。Bertasius 等人率先利用一个五层的卷积神经网络和全连接网络进行图像边缘的检测并取得了突出的结果^[48]。随着全卷积神经网络（FCN）与短连接技术的发展，Xie 等人提出了在边缘检测领域具有划时代地位的 HED 方法^[43]，它以 FCN 为基础，提出了用于边缘检测的全卷积神经网络，并将中间层特征分别用于边缘图的端到端预测，同时添加了短连接以保证网络的优化。HED 的简洁与端到端的检测特性，使得 HED 的计算效率高，得益于多尺度 CNN 特征，其检测结果也较为突出。此后，Liu 等人在 HED 的基础上提出了富尺度神经网络的特征融合方法，首次在 BSDS500 数据集上超过人类的准确率^[42]。Xu 等人又引入了门函数条件随机场来进一步融合多尺度特征^[49]。

传统基于霍夫变换的直线检测一般先通过 Canny 算子来进行边缘检测，但 Canny 算子检测的边缘缺少语义信息，难以满足语义线的需求。因此，在本文后续的实验中，采用了基于深度学习方法的 HED 网络作为边缘检测器，配合后文详细介绍的霍夫变换构建了方法基线。

第二节 霍夫变换与直线检测

霍夫变换最早由 Hough 提出^[6]，用于直线的检测，后来被 Duda 等人广泛推广用在自然图像处理中^[7]。其核心思想是一条图像中的直线，可以唯一地被两个参数所表示。如直线方程 $y - ax - b = 0$ ，其中斜率 a 和截距 b 可以唯一刻画一条直线，然而由于斜率 a 的取值范围为无限区间，Duda 等人利用极坐标来表示一条直线，通过辐角和距离来唯一刻画一条直线，并将两个参数规范的有限区间内。因此，直线检测问题可以被转化为点根据参数方程的参数来向参数空间投票，并在参数空间中寻找极值的问题。相同的思路也被用于圆和椭圆等容易参数化的形状检测^[7]。然而，霍夫变换也存在着一些固有缺点，如计算和显存的开销太大，对于复杂形状的检测，其计算量难以负担，且复杂形状难以被有效参数化。高精度的检测结果依赖于细致的参数空间量化划分，然而精细的划分会导致现存的开销急剧增大和计算时间的急剧增加。

为了解决物体参数化的问题，广义霍夫变换（GHT）也被 Ballard 等人提出

用于检测任意形状^[8]，尤其是针对非参数化的形状，利用学习的手段计算物体形状映射表，并通过边缘点来投票解决。为进一步推广到任意物体检测 and 更精细的检测结果，提出了动态 GHT 和两阶段 GHT 来弥补 GHT 中遗留的计算精度不高的问题^[50, 51]。

为了解决霍夫变换计算速度上的不足，一些基于概率和随机方法的霍夫变换（PHT 和 RHT）被提出，来加速霍夫变换过程。这些方法随机或者按照某一概率函数选择一部分的点来进行霍夫投票，大大加速了计算过程，而准确率只有少量的损失。对于经典的霍夫变换直线检测，假设图像像素点个数为 M ，而直线角度的量化数为 N_θ ，则霍夫投票过程的计算复杂度为 $O(MN_\theta)$ 。其中，一般情况下 M 会远远大于 N_θ ，因此 PHT 方法通过缩小 M 来进行计算的加速。Kiryati 等人提出了以概率的形式选择 M 的子集进行投票，并证明了其对结果影响较小，此时计算复杂度为 $O(mN_\theta)$ ，其中 m 为小于 M 的子集。同时，概率霍夫变换的子集选择也使得霍夫变换能够从直线检测转换为线段检测，更为通用。由于选择的子集不同，在参数空间内极值点的阈值也会相应变化，渐进概率霍夫变换（PPHT）被提出，根据子集的大小自适应地选择接受阈值^[52]。也有方法将梯度信息融合进去来进一步控制投票过程，来加强方法的鲁棒性^[53]。随机霍夫变换 RHT 的系列方法，将霍夫投票的一对多映射过程转换为多对一映射，通过随机过程不断地选择符合要求的子集来快速筛除不合理的投票点，加速了投票过程。Kiryati 等人分析了 RHT 的过程，并指出 RHT 的计算速度在高质量图像上要快于 PHT 的方法，然而其鲁棒性较低，更容易受到噪声的影响。同时，一些工作采用了基于核方法的霍夫变换来进行霍夫投票，在共线像素上使用椭圆高斯核来加速原始霍夫变换^[54, 55]。John 等人将输入图片分为层次化的图像块，然后在这些块中独立地进行霍夫变换^[56, 57]。Illingworth 等人^[9] 使用由粗到细的累积和搜索策略在霍夫参数空间中识别主要峰值。有效缓解了霍夫变换计算复杂度高，内存开销大的问题，同时提高了检测的精度。

霍夫变换的思想，直至今日也在各个领域深刻影响着研究者们，最近两个工作采用了霍夫变换的核心思想，分别是使用截距和方向参数化 CNN 的输出来预测表面网格^[58] 和对三维形状进行凸分解^[59]。霍夫投票方案在许多其他任务中也被采用，如检测点云中三维形状的重心^[60] 和识别图像对应点^[61]。

而将霍夫变换用于直线检测，是直线检测领域的标准化做法。由于研究者们一般假设直线一定出现在图像的边缘区域，直线检测首先由图像边缘检测前

处理构成，通过边缘检测得到的边缘图，进行霍夫变换进行直线检测。然而在线段检测领域，尽管霍夫变换鲁棒性强且可并行化，但不能直接应用于线段检测，因为其不能识别线段的端点。PHT 在票决方案中使用随机采样^[10]，并通过定位样品点位置重建线段，但该方法仍偏向于检测长直线。在线段检测中，除去霍夫变换，还有许多其他研究用于检测线段。Burns 等人使用边缘方向作为引导提取线段^[5]。主要优点是梯度方向可以帮助识别低对比度的线条和端点。Etemadi 等人^[62]在给定的边缘图中建立链，通过遍历这些链来提取线段。Chan 等人^[63]使用量化的边缘方向来搜索和合并短的线段。Gioi 等人^[12]在不调节参数的前提下提出了线性时间复杂度的线段检测器 (LSD)，并被许多后续研究^[35, 64, 65]所采用。

第三节 目标检测相关技术

现有的基于深度学习的语义线检测和直线、线段检测的方法大多沿用了目标检测的方法思路，本节将介绍目标检测的相关研究工作，尤其是以深度学习为基础的目标检测方法。早期的目标检测方法大多依赖于手工设计的特征，如梯度直方图 HOG 等，在通过分类器判别每一个可能的区域是否为要检测的物体。这种依赖于手工设计的特征容易受到光照、变形的影响，鲁棒性较差。基于深度学习的目标检测方法在近十年来被广泛研究，主要可以被分为：两阶段的目标检测检测器、基于锚点的单阶段目标检测器、不基于锚点的单阶段目标检测器。

两阶段的目标检测器是近些年来比较受关注的目标检测方法，这些方法的检测流程分为两个阶段，第一个阶段通过一个目标提案器产生诸多候选框，避免了滑动窗口搜索的过大时间复杂度，第二个阶段通过基于深度网络的分类器对目标框内的特征进行提取、聚合和分类，判断是否为要检测的物体，并确定其类别。RCNN 是在深度学习领域最早取得成功的两阶段目标检测方法^[66]，RCNN 首先通过选择性搜索 (Selective Search) 进行候选区域分析，启发式地产生候选框。然后利用 CNN 的表征能力，对每一个候选框内的图片裁剪后使用 CNN 进行特征提取，最后通过 SVM 进行分类和回归。RCNN 的流程为两阶段目标检测器奠定了基础，后续的工作大多仍沿用这个框架进行。由于 RCNN 的计算复杂度高，Girshick 在 RCNN 的基础上共享卷积层，并提出了有效的特征聚合 RoI 池化操作，极大的加快了计算速度，同时提升了特征的准确性，增强了检测精度^[39]。同时期的 SPPNet 也利用池化的思想加快了计算效率^[67]。Faster RCNN 在

这些工作的基础上，将预选框的生成也用全卷积网络来实现，将所有操作统一到 GPU 计算上，形成了更快速的解决方案^[37]。而近几年的方法大多沿用着一框架，在其之上进行特征网络的修改，检测头的增强和多尺度特征的增强，成为了目标检测领域的一次又一次最领先解决方案。

单阶段目标检测器将两阶段中的目标提案器去除，通过端到端的整体训练直接得出目标检测的结果。这种做法牺牲了一个前处理过程，因此速度有了很大的提升，但检测效果略微欠佳。最早有代表性的网络是 YOLO^[68]，该网络去除了目标提案步骤，通过网格划分，在每个网格内去预测中心落在该网格内的物体，将分类任务转换成了回归任务来处理，且网络结构没有复杂的池化操作，仅通过全卷积来实现，具有极高的计算效率。此后，YOLOv2 在其基础上引入了锚点的概念，通过预定义一系列的锚点框，然后预测回归目标与锚点之间的偏差，简化了网络的优化过程，取得了更高的检测准确度^[69]。基于锚点的方法后续被广泛研究，大量工作基于锚点这一概念进行网络结构设计，如 SSD^[70]、YOLOv3^[71] 等研究工作。

虽然基于锚点的网络结构取得了较高的检测准确度和网络推理速度，但是锚点需要预先精心设计，例如在 YOLOv3 等工作中采用了基于聚类的方法来设置锚点的大小和数量。此外，锚点这一概念仍然导致网络有额外的计算开销和先验不准确的问题。DenseBox 做了最早的去掉锚点的尝试，通过在密集像素点上通过全卷积网络直接回归预测每个物体的位置^[72]。Tian 等人提出了一种不需要提案的且不需要锚点的全卷积网络 FCOS，并引入了中心置信度在不同尺度的特征上进行目标检测^[73]。而 ConerNet 将回归的目标转换为检测框的左上端点和右下端点位置，让网络更容易优化^[38]。近两年来，Transformer 技术引起了广泛关注，DERT 是第一个将 Transformer 相关技术应用在目标检测领域的研究工作^[74]，它将特征提取分为编码和解码两部分，在解码阶段利用注意力机制在不同尺度上聚合信息，并通过随机初始化的查询点来求得最终检测的结果，该网络不仅是单阶段不依赖于锚点的端到端训练方法，且通过注意力机制去除了复杂的 NMS 后处理流程，极大简化了目标检测的流程。

以上目标检测方法作为通用目标检测技术，被广泛借鉴应用于人脸检测、行人检测、缺陷检测、线段检测、直线检测和语义线检测等领域，对检测任务产生了深远影响。

第四节 语义线检测相关方法

受到目标检测方法的影响，很多研究者利用目标检测的流程进行线结构的检测。最早被应用在了线段检测任务中，该任务同样需要语义信息来进行判别，完成室内场景的线框解析^[36, 75-77]。Zhang 等人^[36]采用了目标检测中的 CornerNet^[38]的概念，在室内场景中通过提取成对的关键点来提取线段，通过检测线段的两个端点，然后进行配对来完成线框解析。Huang 等人^[75]提出了两路分支网络来预测线和节点以进行线框解析，通过添加一个线提取子网络来扩展此方法^[76]，将像素级的线检测和端点回归任务分离，用两个子网络来分别处理，更好的优化了检测目标。Zhou 等人^[76]提出了端到端的结构在线框解析任务中进行详细的线段检测。

随着深度学习技术在线段检测领域中逐步获得成功，研究者开始将目标聚焦于有助于摄影构图和美学评估的语义线检测任务。而语义线是一种根据语义信息判断得到的具有特殊属性的直线，最早在 Lee 等人的工作中被提出^[4]。Lee 等人同时给出了一种简单的检测方案 SLNet，遵循 Faster-RCNN^[37]的两分支流程，一条分支验证线的存在，另一条分支进一步回归线的详细位置。与传统的 Faster-RCNN 不同，SLNet 提出了一种针对于直线的 LOI 池化方式，利用双线性插值在一条直线上进行特征汇聚，形成固定长度的特征向量，并由一个分类器和回归器进行验证与偏移量回归。然而，这种直接使用 LOI 对直线进行池化的操作，难以捕捉直线与直线之间的交互关系。

卷积网络通过局部扩散感受野的形式进行特征提取，同时自然物体大多是紧凑的从局部到整体的规则形状，因而基于 CNN 的目标检测能够取得较为理想的检测结果。然而，直线的检测与自然物体的检测略有不同，直线这种细长的、贯穿整个图像的特殊结构，使得 CNN 这种局部特征聚合的方式难以捕捉到非局部特征，难以对整个直线进行充分的语义分析。因此，现有的语义线检测亟需一种能够同时综合了 CNN 强大的局部特征提取能力和考虑直线本质特性的特征汇聚的方法的解决方案。

第五节 本章小结

本章介绍了本文研究的语义线检测任务的相关技术，详细阐述了传统方法中直线检测的基本流程，包括边缘检测和霍夫变换。同时，霍夫变换作为本文的重点研究对象和参考方法，详细介绍了其相关工作及发展路线，指出了边缘

检测加霍夫变换方案用于语义线检测存在的问题。然后介绍了基于深度学习的目标检测的发展历程和经典方法，以及受目标检测方法启发而产生的基于深度学习的线段检测和语义线检测方法。最后本章着重分析了现有语义线检测任务的研究现状与存在的问题。

第三章 基于深度霍夫变换的语义线检测方法

在本章中，针对现有语义线检测方法存在的问题，阐述介绍了一种全新的应用于语义线检测任务的方法流程。该方法中的重要模块为基于霍夫变换的深度神经网络模块，因此，将本方法命名为深度霍夫网络（Deep Hough Network, DHN）。本节将详细介绍提出的用于语义线检测的深度霍夫网络，以及整个深度神经网络架构和语义线检测的完整流程。本文提出的方法主要包括五部分：（1）一个基于 CNN 的编码器用于提取像素级深度特征表示；（2）几个深度霍夫变换模块用于将不同尺度的深度特征表示从空间空间转换到参数空间；（3）一个基于内容的线检测器用来负责在参数空间检测语义线；（4）一个逆向霍夫变换将检测到的线映射回图像空间；（5）一个可选的边缘对齐修复模块进一步修正直线的位置。所有的组成部分整合形成了一个端到端的全新框架，可以在单个步骤中实现前向推理和反向传播训练。

第一节 直线参数化和反参数化

首先将介绍直线的参数化和反参数化以及量化过程，为后续的霍夫变换和深度霍夫变换提供基础。直线参数化将图像中的一条直线通过距离和角度进行参数式表达，将欧式空间中的直线，转换成参数空间的点来表示。如图3.1所示，给定一张二维图像 $I_{H \times W} \in \mathbb{R}^{H \times W}$ ，设定原点为图像中心，在二维平面上，一条图像内的直线 l 可由两个参数进行参数化：一个方向参数 $\theta_l \in [0, \pi)$ 代表 l 与 x 轴正方向的夹角角度，以及一个距离参数 r_l 表示 l 和原点的距离。这两个参数可以通过简单地求取直线的参数方程，然后计算点到直线的距离得到。同时，对于图像 I 中的任意一个直线 l ，有 $r_l \in [-\sqrt{W^2 + H^2}/2, \sqrt{W^2 + H^2}/2]$ ，其中距离为负表示原点到直线的垂足落在三四象限，反之为一二象限。本文定义线到参数的映射过程和逆向映射过程为：

$$\begin{aligned} r_l, \theta_l &\leftarrow P(l) \\ l &\leftarrow P^{-1}(r_l, \theta_l) \end{aligned} \quad (3.1)$$

显然, P 和 P^{-1} 为双射映射。给定图像 I 上的任意直线 l , 可以用以上方程中 P 进行参数化, 也可以将任意有效的 (r, θ) 通过 P^{-1} 逆向映射为一条直线。

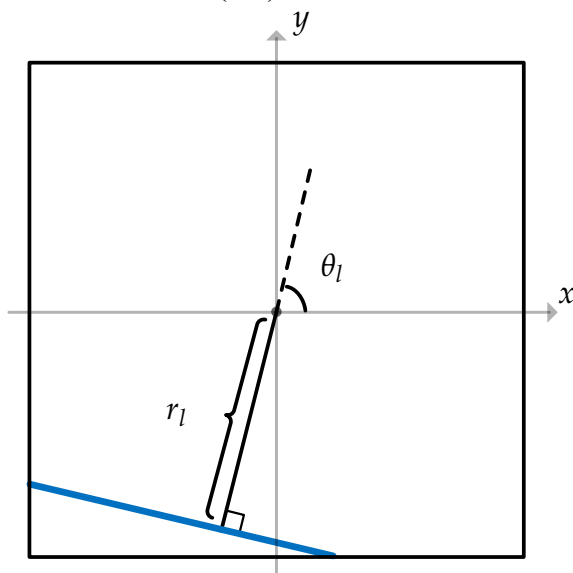


图 3.1 直线参数化示例。一条直线可由截距 r_l 和斜率 θ_l 参数化表示。

实际情况中, 需要将 r 和 θ 量化至离散的整数中以供计算机程序处理。而这一参数的量化过程决定了处理的精度, 较小的量化间隔带来精细的处理程度, 同样也会导致后续的计算开销变大。设 r 和 θ 的量化间隔分别为 Δr 和 $\Delta \theta$ 。则量化过程可由以下公式表示:

$$\hat{r}_l = \left\lfloor \frac{r_l}{\Delta r} \right\rfloor, \hat{\theta}_l = \left\lfloor \frac{\theta_l}{\Delta \theta} \right\rfloor \quad (3.2)$$

其中 \hat{r}_l 和 $\hat{\theta}_l$ 为量化后的整数值。

用 Θ 和 R 表示的量化度的总个数, 即距离和角度两个维度得到的量化总数为:

$$\Theta = \frac{\pi}{\Delta \theta'}, R = \frac{\sqrt{W^2 + H^2}}{\Delta r} \quad (3.3)$$

以上直线的参数化和反参数化过程, 以及量化过程, 为霍夫变换提供了理论基础。

第二节 霍夫变换基线方法

霍夫变换是一种应用于传统图像领域的方法, 最初用于检测图像中的直线结构, 而后被推广到检测任意形状。本节对霍夫变换在直线检测上的工作原理

简单介绍，同时依据此方法与基于深度学习的边缘检测方法共同构成了本文的基线方法，用于与提出的深度霍夫网络进行对比分析。

如公式 (3.1) 所述，图像中的直线可以一一映射成两个参数表示，即 r_l 和 θ_l 。因为这两个参数都是在有限区间内的值，因此可以被表示为霍夫参数空间中的一个点。一簇相似的直线，经过霍夫变换后会被映射到参数空间中距离比较近的点。霍夫变换根据这样的对应关系，把所有线的映射问题转换为原始空间的投票问题。首先对图像进行边缘检测，然后在图像空间中的每一个被检测为边缘的像素点都为其支撑的直线进行投票，而支撑的线为过该点任意方向的直线。经过投票，可以直接在参数空间中寻找极值点，进而确定检测出的直线。

使用霍夫变换进行直线检测，依赖于边缘图的输入。在本文中，构建了一个基线方法，即使用基于深度神经网络的边缘检测方法 HED，并在语义线检测数据集中使用语义线的稠密像素标注重新训练，得到一个检测语义线边缘的 HED 网络。然后对图像进行语义线边缘检测，再对检测的结果使用霍夫变换，得到矢量化的直线参数，实现直线检测。图 3.2 展示了基线方法的流程示意图。该基线方法与传统的直线检测方法相比，使用了基于 CNN 的 HED 方法，能够利用 CNN 的强大表征能力对语义信息进行学习与理解。在实验中，本文除了对专门进行语义线检测的网络对比外，也对该基线方法进行对比，验证后续方法的有效性。

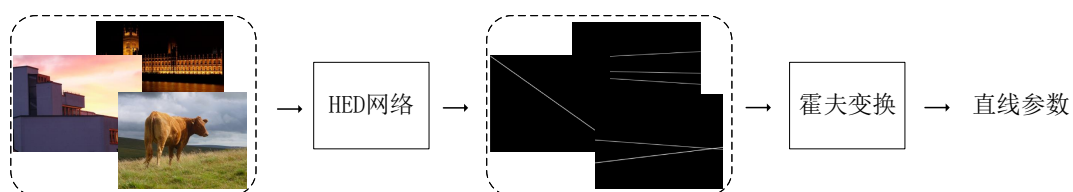


图 3.2 HED 与霍夫变换结合的基线方法

第三节 深度霍夫网络 DHN

3.3.1 深度霍夫变换 DHT

深度霍夫变换 (Deep HT, DHT) 以霍夫变换为基础，将欧几里得空间的图像的深度网络表征转换为参数空间的参数表征。给定输入图像 I ，本方法首先使用编码器网络提取深层 CNN 特征 $\mathbf{X} \in \mathbb{R}^{C \times H \times W}$ ，其中 C 表示特征通道数， H 和 W 为空间尺寸。DHT 以 \mathbf{X} 为输入并计算霍夫变换后的特征 $\mathbf{Y} \in \mathbb{R}^{C \times \Theta \times R}$ 。变换

后特征的尺寸 Θ, R 由量化级数决定, 如公式 (3.3) 所述。如图3.3所示, 给定图像中任意直线 l , 沿着 l 聚集所有像素的特征到参数空间 Y 中的点 $(\hat{\theta}_l, \hat{r}_l)$ 上:

$$Y(\hat{\theta}_l, \hat{r}_l) = \sum_{i \in l} X(i) \quad (3.4)$$

其中 i 为位置索引。 $\hat{\theta}_l$ 和 \hat{r}_l 由直线 l 的参数根据公式 (3.1) 决定, 然后根据公式 (3.2) 量化为离散值。公式 (3.4) 描述了将欧式空间的表征, 转换到参数空间的表征的过程。对于每一条可能的直线, 都通过参数化的形式, 将原始的特征向量累加到参数空间中指定的位置, 进而完成对整个特征的变换操作。其中给定量化级数 Θ 和 R , 一共可得到 $\Theta \cdot R$ 个独立的候选线。与传统霍夫变换 HT 相比, DHT 将 C 维特征的张量转换到了参数空间张量中, 而 HT 转换的一般是单通道的边缘图或者响应值。

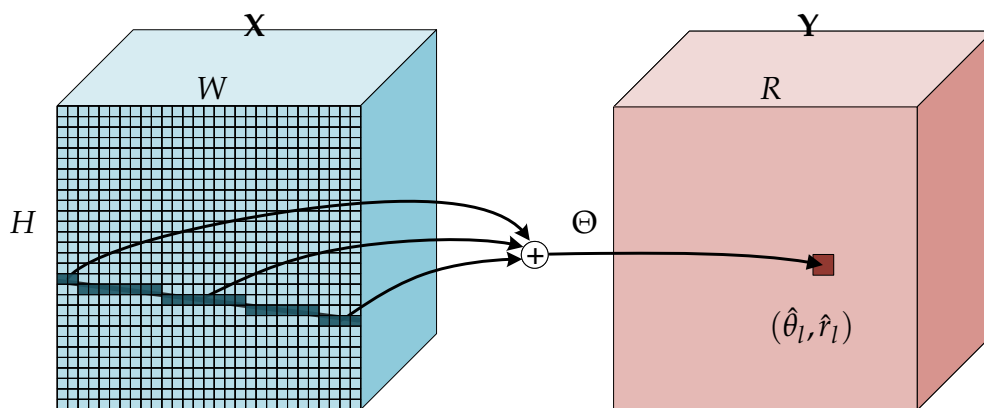


图 3.3 图像空间 (蓝色, 左边) 中沿直线的特征可累积至参数空间 (红色, 右边) 中的点 $(\hat{r}_l, \hat{\theta}_l)$ 。

在具体实现中, 采用霍夫投票的方式, 通过遍历欧氏空间中的每个像素点, 然后遍历量化级数 Θ 并计算对应的距离 r , 可以得到经过该点的所有直线, 然后依次将其累加到参数空间中。由于整个深度霍夫变换过程是每条线独立的, 或者根据实现方式是每个点独立的, 使得该操作可以高度并行化。得益于高度并行的 GPU 运算, 该操作引入的时间开销微乎其微。

3.3.2 基于 FPN 的多尺度 DHT

由于语义线的检测需要丰富的语义信息来判断直线的属性, 因此提取多尺度的特征描述是非常有必要的。而特征金字塔网络 (FPN)^[78] 是一个常用于目标检测领域的多尺度特征结构, 它用于不同尺度间的特征的融合, 并构建出不同

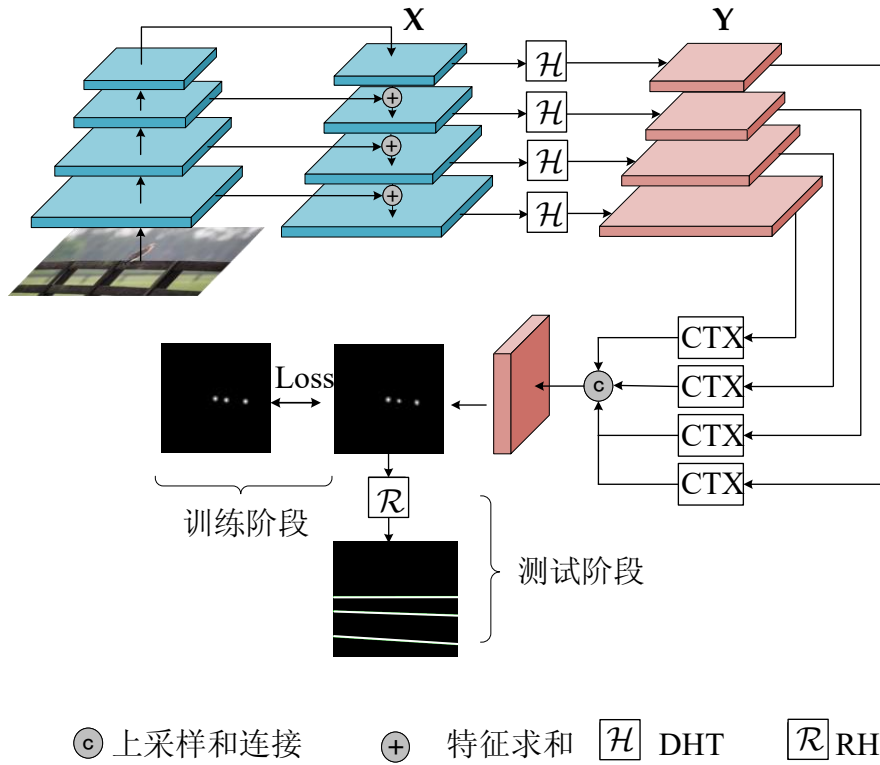


图 3.4 本文提出的深度霍夫网络流程图。DHT 为深度霍夫变换的缩写，RHT 代表逆向霍夫变换。CTX 为内容感知的线检测器，包含多层卷积层。

尺度的特征表达，有助于提取多尺度的、有丰富语义的特征。通常，在 ResNet 的四个层级中，逐渐通过对深层级进行额外的卷积处理与上采样，然后与浅层级特征融合，得到不同尺度的表达。本文提出的深度霍夫变换 DHT 可以轻易地应用于任意尺度特征，本节中将其与 FPN 结合，构成多尺度 DHT。

在图像特征编码的 CNN 后，增加了 FPN 作为多尺度特征描述结构，FPN 输出在四个尺度的特征图 X_1 、 X_2 、 X_3 、 X_4 ，为多尺度融合后的特征表征，其对应的分辨率为图像输入分辨率的 $1/4$ 、 $1/8$ 、 $1/16$ 、 $1/16$ 。然后每个特征图分别由一个 DHT 模块进行变换，如图 3.4 所示，共四个深度霍夫变换结构。本方法在不同的 DHT 间使用相同的量化间隔，以避免在后续的特征融合过程中存在歧义。由于这些特征图有不同的分辨率，变换后的特征 Y_1 、 Y_2 、 Y_3 、 Y_4 也有不同的尺寸，例如，角度量化总数相同，而距离量化总数因特征分辨率不同而不同。

3.3.3 参数空间中的线检测

经过多尺度深度霍夫变换后，特征被转换至参数空间，其中网格位置 (θ, r) 对应特征空间中沿整条直线 $l = P^{-1}(\theta, r)$ 的特征。将特征转换入参数空间的一

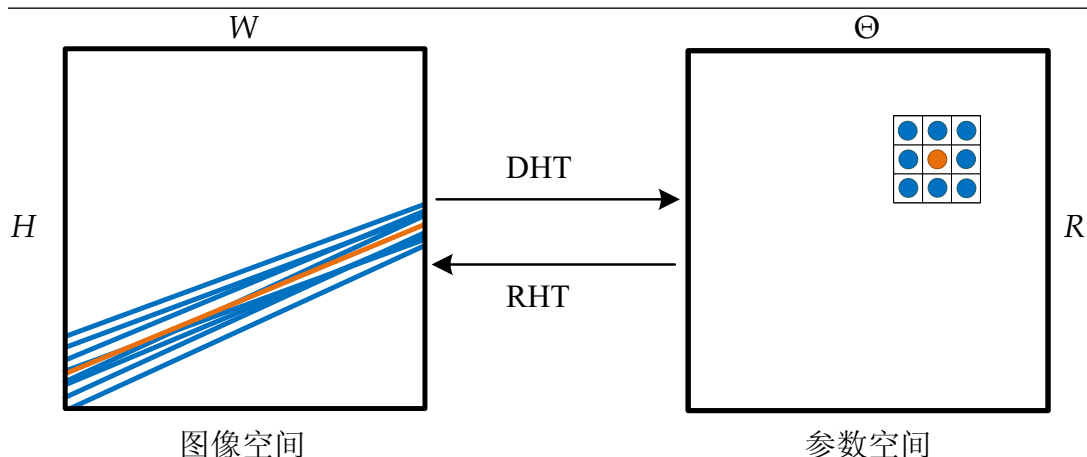


图 3.5 所提出的内容感知的线检测器的图示。特征空间 (左边) 中附近线的特征转换至参数空间 (右边) 中的相邻点。参数空间中, 简单的 3×3 卷积操作可以轻易地获得中心线 (橙色) 周围线的内容信息。

一个重要原因是线结构可以更紧凑地表示。如图3.5所示, 线 l 附近的其他线也被转换为 (θ_l, r_l) 周围的点。将原本图像空间中非局部紧凑的直线簇特征向量结构, 转换成了参数空间中局部紧凑表示的点特征向量结构。有了这种紧凑表示后, 周围线的特征可以通过在参数空间中使用卷积层进行有效的聚集。在 FPN 每一层, 本文使用两个 3×3 卷积层聚集上下文相关线特征。在参数空间中进行 3×3 卷积, 其聚合九个参数点间的信息, 对应了聚合原始图像空间中九条比较相近的直线的特征表示。通过一个这种 DHT 转换, 就可以将非局部的结构, 通过卷积在参数空间中进行高效局部信息聚合。这种聚合能够感知不同直线的语义内容, 并学习到合适的参数进行自适应的线特征聚合, 被称为内容感知的参数空间线检测器 (CTX)。在不同尺度的特征上使用 CTX, 对应地可以在原始图像中检测出精细粒度不同的直线, 如清晰明确的边缘, 在高分辨率的特征变换后经过 CTX 被检测出来, 而抽象的语义划分线, 一般通过低分辨率的特征变换后经过 CTX 进行检测。分别在每个尺度进行局部信息聚合, 将处理后的 Y_2 、 Y_3 、 Y_4 特征图双线性插值到 Y_1 的尺寸, 以匹配不同层的特征图的分辨率然后将其在通道维度连接在一起, 得到融合后的多尺度参数空间表征, 如图3.4所示, 最后, 在连接后的特征图上应用一个 1×1 卷积层生成像素级别的预测。

3.3.4 损失函数

由于预测直接在参数空间中通过 1×1 卷积生成, 相应的损失函数也在参数空间中进行计算, 并完成监督学习过程。对于一张训练图像 I , 标注的线首先通

过标准霍夫变换按照确定的量化度，转换入参数空间中，即求出该直线量化后的角度和距离参数，得到一个由点构成的参数空间通道图。然后，为了加速收敛，本文使用高斯滤波对得到的参数空间通道图数据进行平滑操作，该操作将点拓展到了一个小的范围，有利于神经网络优化，增加了正样本量。在其他许多任务中也使用了相似的技巧，如人群计数^[79, 80]和道路分割^[81]。

设 \mathbf{G} 为通过霍夫变换将标注直线变换后的参数空间单通道数据图， $\mathbf{G}_{i,j} = 1$ 表示参数空间中索引 i, j 处存在一条线。高斯滤波扩展后的数据图为：

$$\hat{\mathbf{G}} = \mathbf{G} \circledast K \quad (3.5)$$

其中 K 为 5×5 高斯卷积核， \circledast 代表卷积操作。如图3.4中的训练 Loss 部分所示为一对平滑后的真实数据和预测图的示例图。最后，在参数空间中计算平滑后真实数据和预测图的交叉熵函数表示为：

$$L = - \sum_i \left\{ \hat{\mathbf{G}}_i \cdot \log(\mathbf{P}_i) + (1 - \hat{\mathbf{G}}_i) \cdot \log(1 - \mathbf{P}_i) \right\} \quad (3.6)$$

并将该交叉熵作为损失函数，优化使其最小来监督训练整个网络。

3.3.5 逆向映射

内容感知检测器 CTX 在参数空间中产生预测，表示存在线的概率。得到概率图后，使用给定的阈值进行二值化，得到二值化的置信图。由于监督标签是经过高斯滤波平滑拓展的，预测图也呈现成团分布的状态。首先，本方法对二值化后的置信图的每个真值区域，计算其连通分量，并计算得到联通分离的质心，这些质心代表了参数空间中检测到的线。通过这种联通分量求质心的操作，避免了大量时间开销的非极大值抑制（NMS）过程，将杂糅到一起的线直接融合为了一条最终表达的线。同时，求质心操作不受到量化操作的限制，具有更高的表达精度。最后，所有参数空间的点坐标根据 $P^{-1}(\cdot)$ 映射回图像空间，按照公式 (3.1) 计算得到。本文将“映射回”的步骤称为逆向霍夫变换 (RHT)，如图3.4所示。深度霍夫网络的输出结果，经过 RHT 后，即得到了图像中检测到的语义线的直线方程，为最终的检测结果。

3.3.6 边缘引导的线微调

然而，受限于量化尺度的约束，检测出的语义线有时并不十分精确。同时，语义线通常为场景中分割不同区域的突出的结构。因此，图像的边缘信息能为

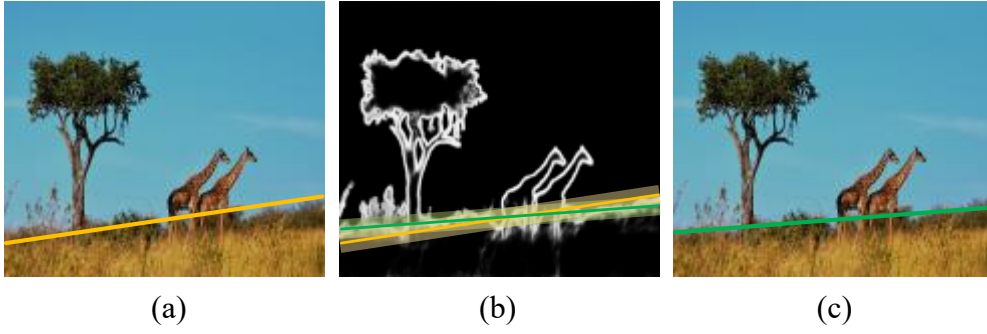


图 3.6 所提出的边缘引导的线微调模块。(a) 原始图像中检测的语义线结果。(b) 基于边缘的线微调模块产生的局部搜索邻域。(c) 根据邻域内可行线的边缘密度，求解的最佳微调结果。

语义线的精准定位提供指引作用。本文提出使用边缘信息对齐线的方法 (Edge based Refinement, ER) 来优化检测结果的方法，并将其作为深度霍夫网络的一个可选项，进一步增强检测精度。首先，使用在 BSDS500 边缘检测数据集上预训练的 HED^[43] 网络来计算图像的边缘图 E 。然后，给定检测到的线 l ， l 的边缘密度定义为沿 l 的平均边缘响应：

$$\rho(l) = \frac{\sum_{i \in l} E_i}{|l|} \quad (3.7)$$

其中 $|l|$ 为 l 贯穿图像的像素数。为了稳定性考虑，本文计算公式 (3.7) 时在 l 两边各扩展一个像素 (总宽度为 3)。设 \mathcal{L} 为 l 周围线的集合。这些线通过将 l 的两个端点顺时针和逆时针移动 δ_r 个像素获得。由于每条线有两个端点，并且每个端点有 $\delta_r + 1$ 个可能的位置，集合中元素个数为 $|\mathcal{L}| = (\delta_r + 1)^2$ 。然后最优位置可通过查找 \mathcal{L} 中有最高边缘密度的最优线 l^* 获得：

$$l^* = \arg \max_{l \in \mathcal{L}} \rho(l) \quad (3.8)$$

图3.6展示了边缘引导的线微调模块的整体流程，对于检测到的每个语义线，根据 δ_r 确定邻域范围，并对邻域内每个线的密度进行计算，如图3.6所示。最后选择边缘密度最大的线作为最终微调的结果。在后续章节的消融实验部分，本文记录了使用不同 δ_r 时“边缘引导的线微调”的性能，并充分验证了 ER 模块的有效性。

第四节 本章小结

本章主要介绍了本文提出的全新的应用于语义线检测任务的深度霍夫网络方法。该方法构成了端到端训练的全新框架，可以在单个步骤中实现前向推理和反向传播训练。本章对方法中的每个步骤依次进行了详细介绍，首先介绍了直线参数化和反参数化以及参数的量化过程，为霍夫变换和深度霍夫变换提供了理论基础。然后介绍了由霍夫变换和边缘检测网络构成的基线方法，并提出了深度霍夫网络。对深度霍夫网络中用于特征变换的深度霍夫变换、多尺度深度霍夫变换、参数空间直线检测器，以及损失函数和霍夫逆向映射进行了详细阐述。最后介绍了深度霍夫网络检测结果的可选后处理模块：边缘引导的线微调，共同构成了本文提出的全新的用于语义线检测的新框架。

第四章 语义线检测的数据集与评价指标

基于深度学习的检测方法依赖于大量数据进行监督训练，数据集的规模和数据集的场景内容的多样性直接影响了算法的检测效果和泛化性。此外，检测结果的合理评价是度量方法好坏的重要步骤，一个合理的评价指标和评价体系能够客观地度量不同方法检测结果的好坏。因此，本章主要就以上两个方面来展开介绍，通过三部分阐述现有的数据集与评价指标存在的问题，并提出新的解决方案。第一部分将介绍语义线检测数据集，首先介绍现有的数据集 SEL，然后在其基础上，介绍了本文收集、标注、制作的用于语义线检测的规模更大、场景更丰富的语义线检测数据集，并对提出的数据集与现有的数据集对比，加以详细的分析和阐述。第二部分将介绍语义线任务的评测指标，介绍现有的评测方式，分析了其存在的不足之处。然后，提出更适用于图像中语义线检测结果评测的指标 EA Score，并对其进行详细阐述和分析。第三部分将介绍语义线检测的评测流程，分析其中存在的问题并加以修正，为后续语义线检测的实验提供了基础。

第一节 语义线检测的数据集

本节总结了现有语义线检测数据集 SEL 的基本信息，并在其基础上，介绍了本文提出的 NKL 数据集。通过对数据集构建的详细介绍以及对数据集数据的统计与分析，验证了 NKL 在样本数量、样本多样性等多方面对语义线检测任务的贡献。

4.1.1 现有数据集总结与新数据集的提出

语义线检测任务是在自然图像中检测出能够划分两个语义区域的贯穿图像的直线，这种直线往往通过数据集来进行刻画，并用于一些基于监督学习方法的训练和语义线检测结果的评测。一个好的数据集的数据应符合问题的定义，同时场景尽可能丰富，以代表自然条件下多变的真实场景。同时，良好的数据集的样本数应尽可能多，以更好地为现有的基于深度监督学习的方法提供数据基础并构建评测基线。

在进行充分调研后，发现目前语义线检测领域，仅有一个公开使用的数据集 SEL^[4]。该数据集是 Lee 等人在 2017 年提出的，包括 1,715 张图片，被划分成了训练集和测试集两部分，其中测试集包括 175 张图片，其他用于训练集。然而，基于监督学习的深度模型一般拥有上百万的网络模型参数量，这些方法往往要依赖于较大规模的数据集才能取得较好的训练结果。SEL 数据集的 1,715 张图像相比之下略显不足。

为了弥补基于深度神经网络的大规模模型和现有数据集规模的差距，本文收集、标注、制作、公开了用于语义线检测的新数据集。新的数据集，命名为 NKL，包括 6,500 张图片，是现有的 SEL 数据集中图片数量的四倍多，且在场景和线数量方面有更丰富的多样性。

4.1.2 NKL 数据集与统计分析

4.1.2.1 数据集收集与标注

在制作 NKL 数据集时，本文首先从互联网上获取免费、公开的且授权科研用途使用的自然图像，利用海洋、草地、街道、城市、天空等特定关键词爬虫获得。经过版权检查后，仔细地筛选出至少包括一条以上语义线的图像，然后进行数据的标注。对 NKL 中每张图片都进行三次标注，以保证标注的质量。首先对每张图片的场景加以理解，根据语义线的定义判断图中语义线的数量和位置，然后将其标注出来。每张图片分别标注三次，并将三次的标注结果加以分析，将多次标注不一致的图片舍弃，仅保留每次标注结果均一致的图片。其中，当且仅当语义线的标注数目一致、且三次标注之间的 EA Score 评测差异均小于 0.05 时，被视作是标注结果均一致。最终，将三次标注一致的结果求均值，得到最终的直线标注。因此，对于每条直线，至少经过了三次标注，形成了最终的语义线检测数据集 NKL。然后类似于 SEL，对 NKL 进行训练集和测试集的划分。

4.1.2.2 数据集的统计数据

本文对标注好的数据集进行统计与分析。NKL 中共有 6,500 张图像，其中包含 13,148 条语义线，而 SEL^[4] 数据集中共有 1,715 张图像，2,791 条语义线。表 4.1 分别统计了两个数据集中的图像数和语义线数以及其在训练集和测试集上的划分结果。从该表中可以看出，相比于 SEL 数据集，在图像样本和语义线样本上，NKL 都显著高于 SEL 数据集，具有更广泛的应用价值。图 4.1 展示了 NKL

数据集和 SEL 数据集中每张图片语义线数量的柱状图。NKL 数据集中超过半数的图片 (67%, 4,356/6,500) 中的语义线的数量大于一, 而 SEL 的比例仅为 45.5%。NKL 数据集中, 包含两条语义线的图片数目最多, 而 SEL 数据集中包含一条语义线的图片数目最多。因此, 本文提出的 NKL 数据集在线的数量上更具有多样性, 更多的语义线数量更符合自然场景, 为语义线检测任务带来了更多挑战性。

表 4.1 SEL^[4] 数据集和 NKL 数据集中图像和线的数量。

数据集	总数 图片数, 直线数	训练集 图片数, 直线数	测试集 图片数, 直线数
SEL ^[4]	1,715, 2,791	1,541, 2,493	175, 298
NKL (本文提出)	6,500, 13,148	5,200, 10,498	1,300, 2,650

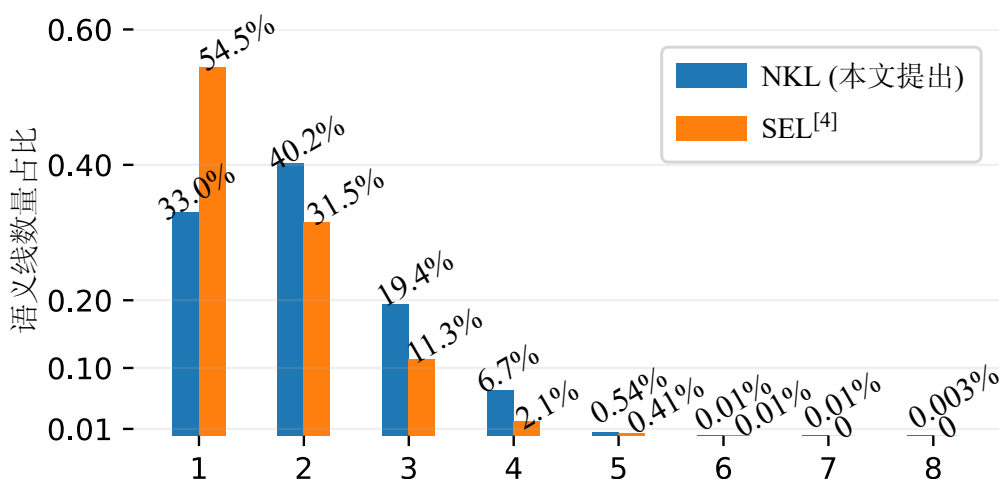


图 4.1 SEL 和 NKL 数据集中线数量分布的柱状图表。NKL 数据集中的线比 SEL 数据集中线的分布更均匀。

4.1.2.3 数据集的多样性分析

数据多样性同样是数据集的一个重要考虑因素。为了分析 SEL 和 NKL 数据集的多样性, 本节使用公开的场景分类模型对数据集内的场景进行分类统计。使用包含 365 类自然场景的场景分类数据集 Place365^[82] 和图像分类的常用卷积神经网络 ResNet50^[83], 将所有图片输入在 Place365 上预训练的 ResNet50 网络, 然后收集输出的类别作为分类标签, 对分类标签加以统计。Place365 数据集共包括 365 种类别, 通过统计分析, 在 SEL 数据集种共得到 167 种类别而在 NKL 中有 327 类。在图 4.2 中展示了统计结果。图 4.2(a) 展示了 SEL 的场景类别分布, 其中天空、田野和沙漠类别占比较大, 呈现明显的长尾分布。从图 4.2(b) 中的结

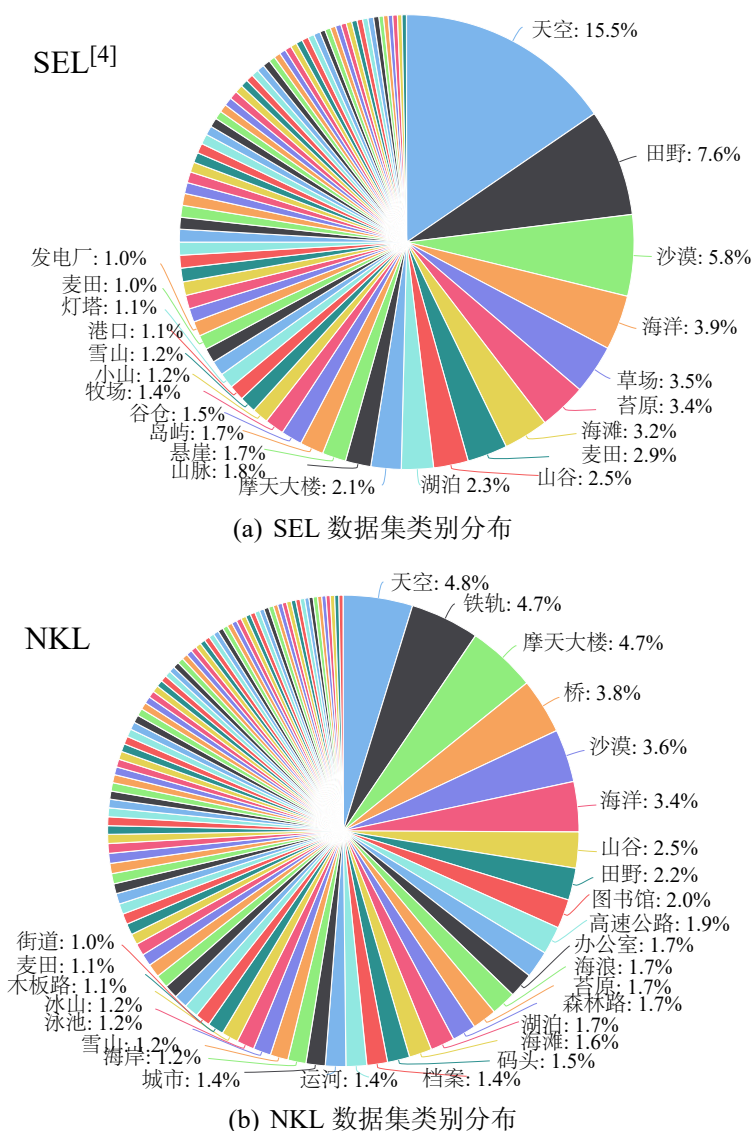


图 4.2 数据集类别分布示意图。类别标签通过 Places365 预训练模型获得。NKL 数据集中包括 327(共 365) 种场景标签，不同于 SEL 数据集的 167 种。

果可以看出，NKL 的类别更多，且分布更为均匀，在 NKL 数据集中，前三个主要类别少于全部的 20%，而 SEL 数据集达到了 25% 以上。所以，NKL 中的数据场景的多样性更丰富，更接近真实世界。

最后，本节对文章给出的 NKL 数据集进行可视化展示，图4.3展示了 NKL 中的部分数据图片以及标注结果，图中黄色的线为标注线，代表图像中的语义线。可以看出，展示的图片场景包含了超市、会议室、泳池等室内场景，也包括沙滩、城市夜景、海边等室外场景，具有多样性。而 SEL 数据集中仅包含室外

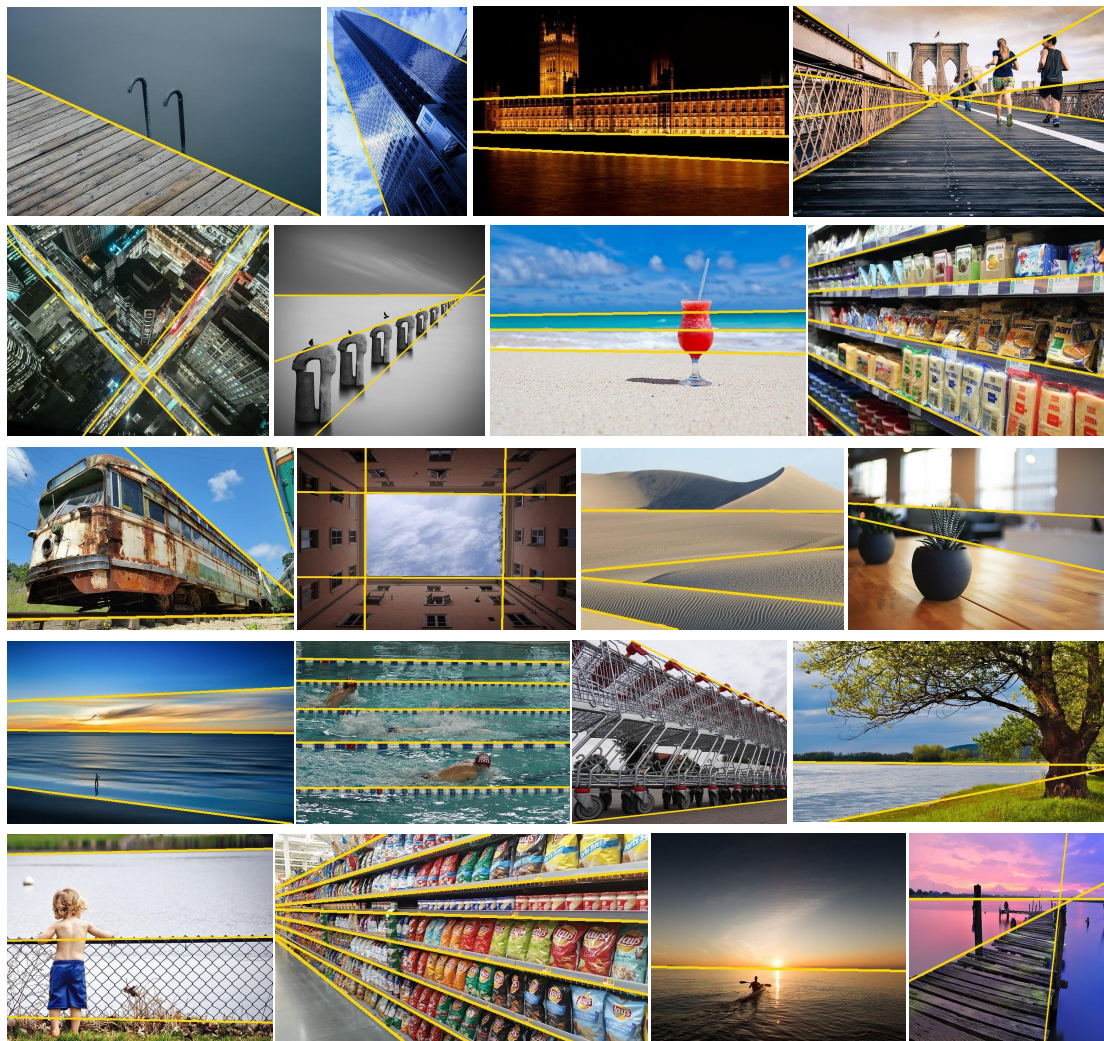


图 4.3 NKL 中的示例图片和标注 (图中黄线)。NKL 中的图片包含多样的场景和丰富的语义线标注。

风景场景，且场景内容较为单一。

第二节 语义线检测的评测指标

在本节中，首先介绍了之前用于语义线检测领域的评价指标，解释为什么现有评价标准不适用于语义线任务的评测。然后介绍在计算机视觉与图形学中常用的度量物体相近程度的指标，并分析其在语义线检测领域的可用性。最后，综合以上，本文提出了一个衡量直线之间相似度的评价标准，衡量贯穿图像的两条直线的相近程度，该评测方法通过欧氏空间的相似度与角度空间相似度两部分来衡量两条直线的相似度，具有简洁、高效的特性，并给出可视化的指标

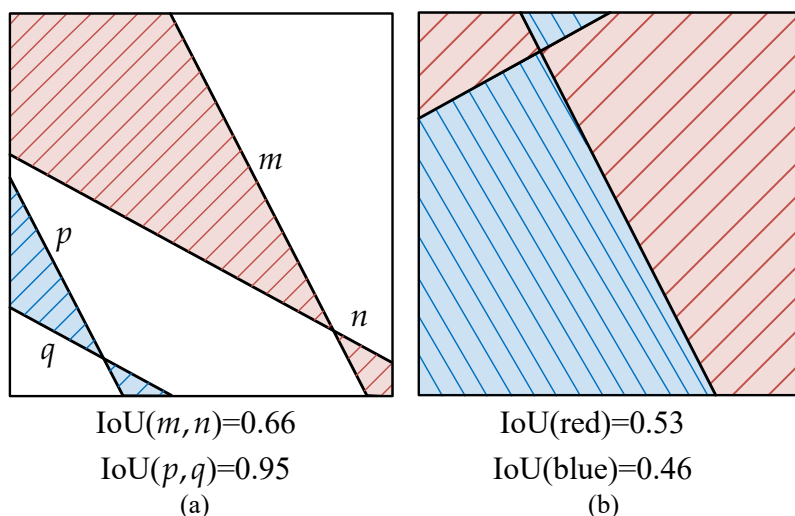


图 4.4 (a): 两对有相似相对位置的线可能有非常不同的 IoU 分数。(b): IoU 度量的交叉区域存在歧义性。

数值和计算过程。

4.2.1 现有度量标准

直线作为基本的图形元素，在数学上很早就有针对于直线相似度的研究工作。然而，基于数学度量的方式通过直线的解析方程，在无限欧几里得空间进行度量，不适用于自然图像中两条直线的相似度量。最直观的方式是度量两条直线重叠的像素，通过衡量像素重叠数量来表示直线的相似程度。然而，这种度量方式的约束太强，依赖于图像的分辨率对直线进行像素栅格化，且对于有偏差的两条直线，其相似度基本为零，没有区分度。同时，这种通过像素化的手段来度量矢量图像，会因为图像尺寸的变换产生不一致的问题。

交并比 (IoU) 是一种广泛应用于目标检测、语义分割等许多任务的一种评价指标，它通过计算两个图形之间的交集面积与并集面积之比，来衡量两个图形的相似度。例如，目标检测中两个检测框之间相似性的度量等。Lee 等人^[4] 利用 IoU 的思想，提出了针对直线的 IoU 计算方式，并用于语义线检测任务的评测。具体而言，它将相交面积定义为两条在图像中交叉时所形成锐角对应的两块面积之和，若两条直线在图像中不相交，则为两条直线所夹区域的面积。而将两个直线的并集面积定义为整个图形面积。如图 4.4(a) 为例，阴影面积即为两条直线的相交面积。则两条直线 m 和 n 的 IoU 被定义为：

$$\text{IoU}(m,n) = \text{area}(\text{相交阴影}) / \text{area}(\text{总图像}) \quad (4.1)$$

其中 $area(\cdot)$ 为计算面积的操作。

然而，基于 IoU 的评价标准并不能完美地用于评测两条直线的相似度，在某些场景下可能会导致不合理或模棱两可的结果。如图4.4(a)所示，两对具有相似相对位置的线 (m, n) 和 (p, q) 会有差异较大的 IoU 计算分数。这是由于两条直线出现在图像中不同的位置引起的。然而，直观的思路是希望两个相对误差相似的线拥有相似的得分，而不受到位置因素的影响。此外，图4.4(b)所示的垂直线样例，即使人也无法准确确定哪个区域应该被用作线 IoU 的相交区域，这种情况是由于两条线极度不相似，接近于垂直的角度出现。这种条件下，轻微的扰动就会导致两直线的交集面积发生突变，进而导致 IoU 值发生突变，而作为一个评价指标，其平滑性和单调性在这种情况下不成立。因此，本文认为基于直线 IoU 的计算方式存在上述两个较不合理的地方，并因此引入新的图形相似度计算方法，以及设计全新的专门用于贯穿图像的直线相似性度量方法。

常用于计算机图形学中的一些其他度量标准，例如 EMD (Earth Mover's Distance, EMD 距离)^[84] 和 CD (Chamfer distance, CD 距离)^[85] 也是良好的形状相似度度量函数。EMD 和 CD 一般用于点云等结构的相似度的度量，本文首先介绍如何将 EMD 和 CD 用于直线的相似度量中。使用 EMD 在进行直线相似度度量时，首先需要将直线转换为离散的点来表示，然后计算将第一条线中每个点，搬运到第二条线中每个点位置的代价，并求这个搬运方案的最小代价作为 EMD 的计算值。而使用 CD 来度量两个直线时，也需要将直线转换为离散的点，然后计算第一个线中的每个点，到第二条线中点的最小距离的平均值作为最终得分，即：

$$d_{CD}(l_1, l_2) = \frac{1}{|l_1|} \sum_{p \in l_1} \min_{q \in l_2} \|p - q\|_2^2 \quad (4.2)$$

其中， l_1, l_2 表示两条直线， $\|\cdot\|_2^2$ 表示欧氏空间二范数距离。经过 EMD 和 CD 计算的距离，使用图像上最大得分为基数进行归一化，即可表达为区间 $[0,1]$ 的相似度度量得分。然而，这些度量标准需要将线栅格化为像素，然后计算像素级别距离，该操作效率较低，且计算结果受到了图像分辨率的影响。为了弥补这些不足之处，本文提出了简单而高效的度量标准来度量相似度，直接对图像中的矢量直线进行度量，其计算效率要远高于现有的 EMD 和 CD 指标。同时，本文中的实验对比表明提出的评价标准可得到与 EMD 和 CD 非常相似的结果，间接证明了本文提出的指标的度量准确性。

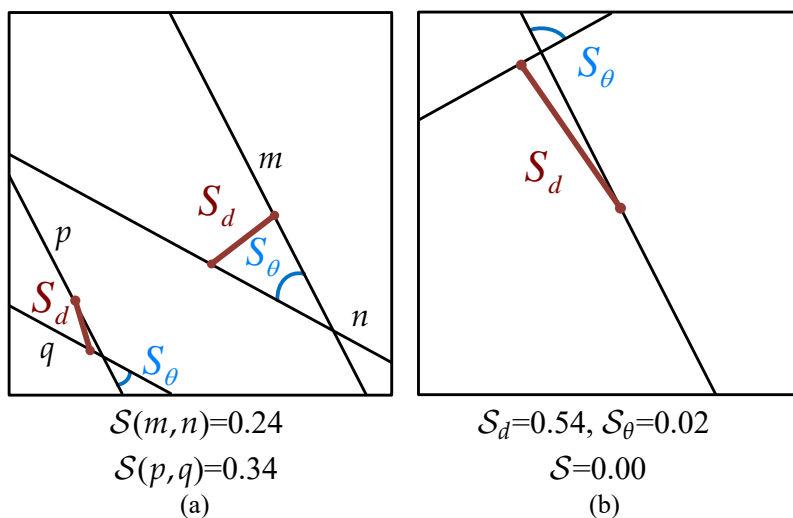


图 4.5 本文提出的评价标准 EA Score 的评测示意图。该指标考虑了一对线的欧氏距离和角度距离，得到了一致且合理的分数。

4.2.2 本文提出的评价标准

针对上述评测指标存在的问题，本文提出了一个全新的、专门用于贯穿图像的两条直线间的相似度度量指标。而语义线恰好为贯穿图像的直线，因此本文提出的评价指标能够直接用于语义线检测的评测。本文提出的评价指标，称作 EA Score，该指标同时计算了一对线的欧氏距离 (Euclidean distance) 和角度距离 (Angular distance)，用首字母简写为 EA。

设 l_i 和 l_j 为一对待度量的线，本文计算其对应的角度距离 S_θ 为：

$$S_\theta = 1 - \frac{\theta(l_i, l_j)}{\pi/2} \quad (4.3)$$

其中 $\theta(l_i, l_j)$ 为 l_i 和 l_j 间的角度，可以通过计算直线的斜率，然后通过反三角函数计算得到。

而欧氏距离衡量了欧式空间两条直线的距离，欧式距离相似度被定义为：

$$S_d = 1 - D(l_i, l_j) \quad (4.4)$$

其中 $D(l_i, l_j)$ 为 l_i 和 l_j 两端点的中点间的欧氏距离，即 $\|\bar{l}_1, \bar{l}_2\|_2^2$ ， \bar{l} 表示直线的中点坐标。本文在计算 $D(l_i, l_j)$ 前将图像正则化为单位尺寸，以避免不同尺度的图像对评测指标带来的影响。最后，本文提出的 EA Score 表示为：

$$S = (S_\theta \cdot S_d)^2 \quad (4.5)$$

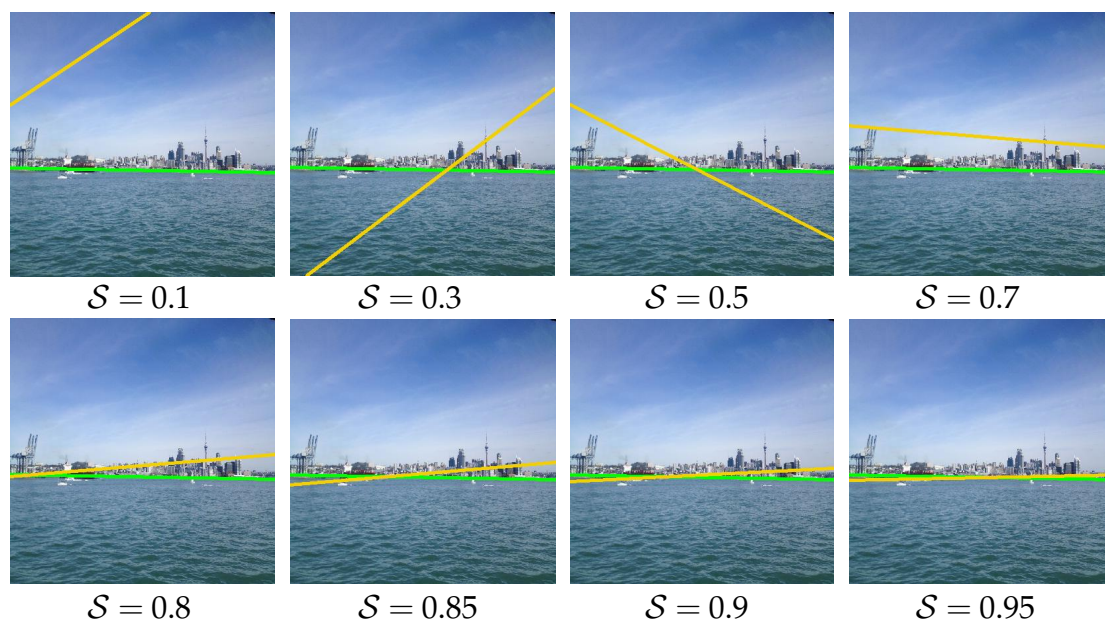


图 4.6 EA Score 在不同样本中的得分示例。EA Score 越大两条线越相似。

因为两个维度的指标都是 0-1 区间内归一化的表示，因此将两个维度的指标相乘作为最终的评价指标。同时，公式 (4.5) 使用二次方以便使得数值过大时更敏感且更有区分度，使得两条直线越接近，其相似度指标的变化速率越大。

在图4.5(a)和图4.5(b)中，展示了欧式空间相似度 \mathcal{S}_d 和角度空间相似度 \mathcal{S}_θ ，以及最终的计算结果 \mathcal{S} 的计算示例与计算值。相比于图4.4(a)和图4.4(b)展示的直线 IoU 计算结果，这种度量能够得到合理一致的分数，且计算清晰明了，可解释性强。在图4.6中给出了不同样式的两条直线对应地计算得到地 EA Score 指标值，可以看出，随着两条直线地接近，其相似度指标越来越大，并在十分接近地范围内，也具有一定地区分度，验证了提出的指标的有效性。

第三节 语义线检测的评测过程

准确率、召回率和 F-measure 是检测任务常用的评测指标，Lee 等人^[4] 也通过计算这三个指标来度量最终的检测结果。本文同样使用准确率、召回率和 F-measure 来评价检测到的线的质量。然而，其他方法采用的计算过程存在不合理之处，本节将介绍本文使用的评测流程及计算过程，并说明其与其他方法的区别，并分析阐述现有做法的合理性。

在计算准确率、召回率和 F-measure 时，第一步是匹配检测到的线与数据集

中标注的线。设 \mathcal{P} 和 \mathcal{G} 分别为预测的线和数据集中标注的线的集合。 p_i 和 g_j 表示单个预测的线和标注的线。之前的方法对每一个预测的直线 p_i ，计算标注线 \mathcal{G} 中与其最相似的线 \hat{g} 。其中相似度通过基于 IoU 的直线相似度量，或使用基于 CD、EMD、EA Score 的度量方法。根据最佳的匹配结果 \hat{g} ，使用一个相似度阈值，来确定哪些匹配是正确的匹配，哪些是错误的匹配，进而计算得到评价结果。然而，对每个检测到的直线 p_i ，从整个标注集中寻找最接近的线是不合理的。例如，当检测出的线数量为 2，而真实标注只有 1 条语义线时，两个检测样本均和同一个标注样本进行了匹配，产生了多对一匹配关系，且对数量出错的情况没有从根本上反映在最后的指标中。而合理的检测匹配是一一对一匹配，即任意一条标注线 $g \in \mathcal{G}$ ，只能与最多一条检测线 $p \in \mathcal{P}$ 匹配。

针对以上问题，本文使用二分图匹配算法来匹配 \mathcal{P} 和 \mathcal{G} 中的线。设 $G = \{V, E\}$ 为一个二分图。顶点集 V 可被分为两个不相交的、独立的集合，在本文的实验中，两个不相交集为 \mathcal{P} 和 \mathcal{G} ：

$$\begin{aligned} V &= \mathcal{P} \cup \mathcal{G} \\ \mathcal{P} \cap \mathcal{G} &= \emptyset \end{aligned} \tag{4.6}$$

E 中每条边代表在某一相似性度量下一对线间的相似性。在本文的实验中，将相似性度量设置为本文提出的 EA Score，同时，为了更有效的对比，也将第二节所述的 EMD^[84] 和 CD^[85] 评价指标作为相似度量来进行实验。在使用 EMD 和 CD 时，根据 EMD 和 CD 可能的最大值将两者正则化到 $[0, 1]$ 范围内（当两条直线在对角线上且垂直时达到最大值）。

给定图 $G = \{V, E\}$ ，二分图匹配结果是保证任意两条边之间没有公共点的前提下得到的边集，而计算二分图中这样一个边集权重的最大值，为二分图的最大匹配，可使用经典的 Hungarian 方法^[86] 在多项式时间复杂度内解决。在本文的语义线检测任务中，给定预测的线的集合 \mathcal{P} 和标注的线的集合 \mathcal{G} ，通过二分图匹配寻找合适的匹配，以保证每条真实线 g_i 对应于不超过一条检测到的线 p_j ，反过来亦是如此。同时，二分图的权重根据 \mathcal{P} 和 \mathcal{G} 中每两两之间的相似度和相似度阈值来决定，当两个线间的相似度大于相似度阈值时，二分图中两样本之间的权重为 1，否则为 0。具体地，对于每一条预测的语义线 $p \in \mathcal{P}$ ，计算其与 \mathcal{G} 中每一条线的相似度，然后根据阈值 τ 设置为为 0 或 1 权重的边长，构成一个预测先 \mathcal{P} 与标注线 \mathcal{G} 之间的二分图，然后求解二分图的最大匹配，最为预测结果与标注线的匹配关系。

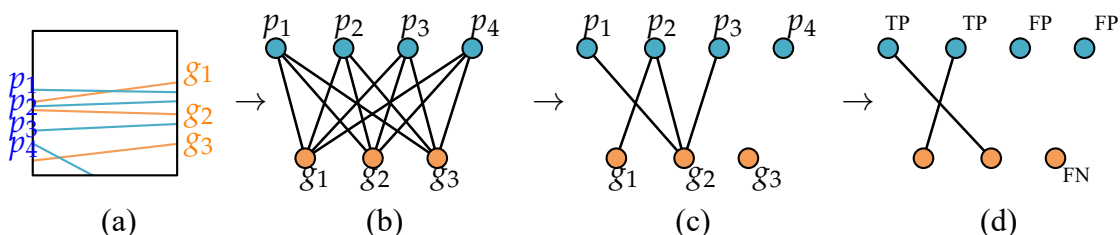


图 4.7 基于二分图匹配的评测流程图解。(a) 包含 3 个真实的线 (g_1, g_2, g_3) 和 4 个预测 (p_1, p_2, p_3, p_4) 示例图像。(b) 对应的二分图。一对节点间的边缘代表两条线间的相似度。(c) 经过阈值筛查后, 只有部分线可以进行匹配 (d) 通过二分图最大匹配, 每个子图中的节点与另一个子图中的节点有不超过 1 条连线。

求解最大匹配后, 根据 \mathcal{P} 和 \mathcal{G} 之间的匹配关系, 计算真阳性样本数 (TP), 假阳性样本数 (FP), 和假阴性样本数 (FN)。其中 TP 为预测线为成功匹配到标注线的数目, FP 为预测线没有匹配到标注线的数目, FN 为没有匹配到预测线的标注线数目。

图 4.7 展示了整个匹配和计算过程, 其中有三个预测的语义线和三个标注的语义线, 首先计算每一个预测线和每一个标注线间的相似度, 然后根据给定的阈值 τ , 构建边权重为 0 或者 1 的二分图, 最后求解最大匹配, 得到二分图的最大匹配关系。最终预测的直线对 (p_1, p_2) 与真实直线对 (g_2, g_1) 匹配时被视为 TP。预测的线 (p_3) 与真实直线间不存在匹配时被视为 FP, 真实直线 (g_3) 不存在对应的预测的线时被认为是 FN。

最后, 准确率 (Precision)、召回率 (Recall) 和 F-measure 计算为:

$$P = \frac{TP}{TP + FP}, R = \frac{TP}{TP + FN}, F = \frac{2PR}{P + R} \quad (4.7)$$

本文对预测数据和真实数据对采用一系列相似度阈值 $\tau = 0.01, 0.02, \dots, 0.99$ 。对应地, 得到了一系列准确率、召回率和 F-measure。最后, 使用平均准确率、召回率和 F-measure 评价性能。由于每个图像中都存在若干条标注线和预测线, 本文在计算 TP, FP 和 FN 时, 通过计算整个数据集中所有的标注线和样本线在每个图像中的匹配关系来得到, 最后计算整个数据集上在不同阈值下的平均准确率、召回率和 F-measure 值。此外, 根据不同相似度阈值下的准确率和召回率值, 可以构成 PR 曲线来衡量检测结果的好坏。本文按照上述准确率、召回率和 F-measure 计算流程, 分别使用 EMD^[84], CD^[85] 和本文提出的 EA Score 评价标准用作量化比较。而在消融实验中, 为了简便仅用提出的 EA 值作为评价标准。

第四节 本章小结

本章针对本文的研究内容，对语义线检测中的数据集与评价体系进行分析，并提出了本文收集、公开的数据集和本文设计的语义线评价指标与评测流程。首先对语义线检测数据集进行了总结和分析，说明了现有数据集 SEL 的不足以及语义线检测领域数据集匮乏的情况，提出了 NKL 数据集，并对 NKL 数据集和 SEL 数据集进行了充分的分析与比较，包括数量、分布、多样性等方面的详细阐述，为语义线检测领域的发展提供了数据基础。然后对语义线检测领域的评价指标进行了分析，指出了现有的基于直线交并比的评价指标中存在的合理和歧义性。同时阐述了用于度量图形相似度的 EMD 距离和 CD 距离指标，并分析了其在语义线检测任务中的适用性，提出了本文设计的语义线评价指标 EA Score。针对 EA Score，进行了详细的阐述和展示，分析了其合理性。最后，对语义线检测领域的评测流程进行分析，提出了基于二分图匹配的评价过程，相比于直接匹配，二分图匹配能够更准确、合理地衡量检测结果。

第五章 实验与分析

本章对本文提出的深度霍夫网络用于语义线检测的方法加以对比验证与分析。首先介绍实验环境与实现细节，包括如何使用本文提出的数据集以及评测方法，以及实验的硬件环境。然后进行大量对比实验，包括定量对比与可视化对比来验证方法的有效性，并与现有方法进行对比。对方法的模块进行消融实验，分别验证了每个步骤的贡献度，以及可选的基于边缘的线微调模块的有效性。最后，使用基于本文提出的语义线检测方法进行应用并展示结果。

第一节 实验环境与实现细节

5.1.1 实验环境

在本文的实验中，主要以深度学习框架 PyTorch^[87] 为主来开展实验，同时为了进一步验证本文提出算法的高效性以及在不同设备上的表现结果，也使用了 Jittor^[88] 开源框架进行实验。由于本文提出的深度霍夫变换 (DHT) 高度并行化，使用原生 CUDA 编程实现了 DHT，而所有其他部分基于普通的 Python API 实现。在对比实验中，本文用单个 RTX 2080 Ti GPU 上进行实验，而为了对比不同设备的影响与 Jittor 框架的加速效果时，本文采用了更多的 GPU 型号并加以说明。

5.1.2 网络结构

本文提出的方法首先使用一个特征编码的 CNN，这里使用两个有代表性的网络结构 ResNet50^[83] 和 VGGNet16^[89] 作为骨干网络，并使用 FPN^[78] 提取多尺度深度特征表示。对于 ResNet 网络，按照现有工作^[90-92] 的常用做法，在最后一层使用空洞卷积^[93]，将空洞率设置为 2，来提高特征图的分辨率为原来的二倍。

5.1.3 超参数设置

本节将介绍方法中一些超参数的设定。首先是对标注样本经过霍夫变换后得到的单通道数据图，对其进行高斯滤波拓展，3.3.4节中使用的高斯核尺寸定义为 5×5 。在对深度霍夫网络的输出进行阈值化处理时，将阈值设置为 0.01。

然后对于所有数据集中的图像，在训练和测试时，本文将其尺寸调整为 (400, 400)，以统一神经网络处理的图像尺度。经过检测后，将检测到的直线再调整尺寸到原始图像大小上，作为最后的输出。在训练时，使用批样本数量为 8。

在网络优化中，使用不包含权重衰减的 Adam 优化器^[94] 对所有模型训练 30 轮，即权重衰减指数设置为 0，而学习率和权重更新动量分别设为 2×10^{-4} 和 0.9。公式 (5.1) 中的量化间隔 $\Delta\theta$ ， Δr 将在 5.2.1 节中进行详细地分析。

5.1.4 数据集和数据增强

本文的实验在 SEL^[4] 数据集和本文提出的 NKL 数据集上进行。两数据集的统计数据细节详见第四章。参考 SEL 数据集文章^[4] 的设置，本文在所有实验中仅使用左右翻转进行数据增强，以保持方法间的公平对比。此外，为了更强的性能，未来可以增加更强的数据增强方式，如顺时针旋转和逆时针旋转。

5.1.5 评价指标

本章节的实验主要依赖于第四章中介绍的评测流程，在参数的选择实验中，通过对比平均 F-measure 来进行衡量，而在定量对比实验中，采用基于 CD、EMD 和 EA-Score 的平均准确率、召回率和 F-measure 来衡量。并在整个对比中，对所有方法采用第四章中描述的基于二分图匹配的评测流程，以结果衡量的准确性，且进行公平对比。

第二节 对比实验

5.2.1 量化间隔的确定

公式 (3.2) 中的量化间隔 $\Delta\theta$ 和 Δr 是影响性能和运行效率的重要因素。量化间隔 Θ 和 R 越大量化级数越小，模型也越快。量化间隔越小，量化级数越多，总体计算代价越大。本文在 SEL^[4] 数据集上使用坐标下降法寻找合适的量化间隔，以使得计算高效且功能上高效。为了简便，使用 EA-score 度量线的相似性。

本文使用坐标下降法来确定量化的间隔，在第一阶段，先固定角度量化间隔为 $\Delta\theta = \pi/100$ ，寻找 Δr ，结果如 5.1(a) 所示。根据图 5.1(a) 所展示的结果，随着 Δr 的下降模型性能首先缓慢上升随后下降，转折点在 $\Delta r = \sqrt{2}$ 附近。在第二阶段，固定 $\Delta r = \sqrt{2}$ 然后使用不同 $\Delta\theta$ 训练。与图 5.1(a) 的结论相似，图 5.1(b) 结果表明随着 $\Delta\theta$ 下降模型性能首先平滑地上升，随后快速振荡下降。因此，转

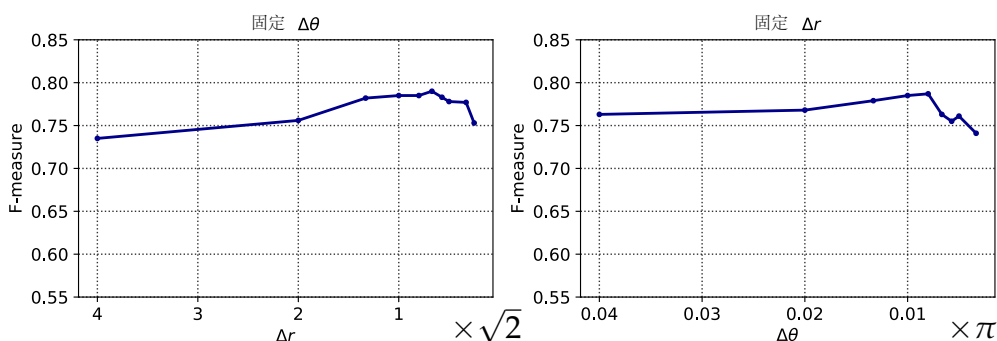


图 5.1 左图：固定角度量化间隔 $\Delta\theta = \pi/100$ ，使用不同距离量化间隔 Δr 时的性能。量化间隔 Δr 越大代表量化级数 R 越少。右图：固定距离量化间隔 $\Delta r = \sqrt{2}$ ，使用不同角度量化间隔 $\Delta\theta$ 时的性能。

折点 $\Delta\theta = \pi/100$ 为角度量化的合适的选择。因此，本文使用 $\Delta\theta = \pi/100$ 和 $\Delta r = \sqrt{2}$ 进行量化，对应的量化级数为：

$$\Theta = 100, R = \sqrt{\frac{W^2 + H^2}{2}} \quad (5.1)$$

其中， H, W 为 DHT 中需要转换的特征图的长宽尺寸。

5.2.2 定量对比实验

在本节中，将本文设计提出的深度霍夫网络方法（DHN）与基于传统霍夫变换和使用 HED^[43] 作为边缘检测器的基线方法（HED+HT），以及现有的工作语义线网络 SLNet^[4] 进行对比实验。在基线方法中，对 HED 网络使用 SEL 数据集进行重新训练，并把语义线作为像素级的边缘标注，以提升任务的适应性，进行更公平的对比。

5.2.2.1 在 SEL 数据集上的对比

首先，本文在 SEL 数据集上进行实验，对比了公开数据集上的表现结果。表 5.1 总结了在 SEL 数据集^[4] 上不同方法的表现性能，其中准确率、正确率和召回率，均为阈值在 0.01 到 0.99 下 99 个划分的平均结果。从表格中的结果可以看出，使用 VGG16 或 ResNet50 作为骨干网络，本文的方法均持续大幅度超过 SLNet 和 HT+HED 基线方法。同时，使用不同的测量标准，如 EMD、CD 和本文提出的 EA，均取得相同的规律，侧面印证了 EA 指标的有效性和合理性。除了表 5.1 的数值对比结果，本节还对比了 F-measure-阈值和准确率-召回率曲线。图 5.2 中可以看出，在 F-measure 曲线中，本文的 DHN 方法在不同阈值下效果均

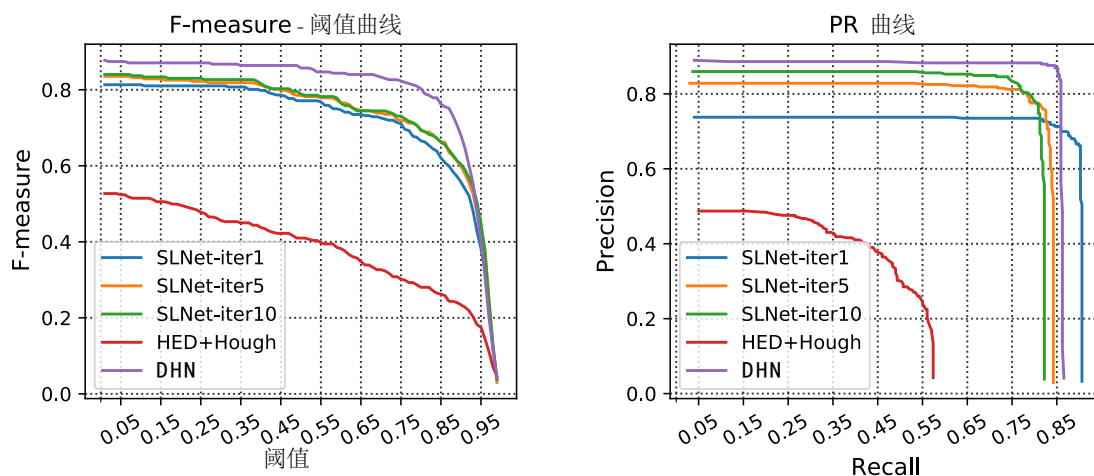


图 5.2 左图：不同阈值下的 F-measure。右图：准确率-召回率曲线。本文提出的方法大幅度超过 SLNet^[4] 和经典霍夫变换^[7]。此外，即使经过 10 轮的位置细化，SLNet 仍展示出更差的性能。

表 5.1 在 SEL^[4] 上，使用 CD、EMD、EA 不同评价标准的定量比较，其中 P、R、F 代表平均准确率、召回率和 F-measure 值。深度霍夫网络 (不使用 ER) 在平均 F-measure 大幅度超过其他方法。

方法	CD			EMD			EA		
	P	R	F	P	R	F	P	R	F
HED ^[43] + HT ^[7]	0.491	0.578	0.531	0.461	0.543	0.498	0.356	0.420	0.385
SLNet-iter1 ^[4]	0.740	0.905	0.812	0.723	0.888	0.797	0.654	0.803	0.721
SLNet-iter5 ^[4]	0.826	0.841	0.834	0.810	0.824	0.817	0.735	0.747	0.741
SLNet-iter10 ^[4]	0.858	0.821	0.839	0.840	0.804	0.822	0.762	0.729	0.745
DHN (VGG16)	0.841	0.835	0.838	0.830	0.824	0.827	0.756	0.774	0.765
DHN (ResNet50)	0.886	0.815	0.849	0.878	0.807	0.841	0.819	0.755	0.786

好于其他方法。在 PR 曲线中，本文的 DHN 方法在图中所围成的面积更大。综合以上结果，本文提出的深度霍夫网络方法具有较好的表现效果。

5.2.2.2 NKL 数据集上的结果

此外，本文展示了在新构建的 NKL 数据集上的结果。由于 SLNet^[4] 并未开源训练代码，仅与 HED 边缘检测器 + 霍夫变换的基线方法作为对比对象来比较本文提出的方法，其中 HED 边缘检测器和 DNH 均使用 NKL 数据集重新训练。并给出在 NKL 数据集上使用 EMD、CD 和 EA 不同评测方法下得到的评测值，包含在不同阈值下的平均准确率、平均召回率和平均 F-measure。如表 5.2 所示，本文提出的 DHN 方法明显领先于基线方法 (HED 边缘检测器 + 霍夫变换)，且

当 DHN 与 ResNet50 骨干网络结合时达到了最好的表现效果。

表 5.2 NKL 数据集上, 使用 CD、EMD、EA 不同评价标准的定量比较, 其中 P、R、F 代表平均准确率、召回率和 F-measure 值。

方法	CD			EMD			EA		
	P	R	F	P	R	F	P	R	F
HED ^[43] + HT ^[7]	0.301	0.878	0.448	-	-	-	0.213	0.622	0.318
DHN (VGG16)	0.750	0.864	0.803	0.726	0.837	0.778	0.659	0.759	0.706
DHN (ResNet50)	0.766	0.864	0.812	0.743	0.839	0.789	0.679	0.766	0.719

5.2.2.3 运行效率对比

在本节中, 对不同方法进行运行效率基准测试, 包括不同迭代次数的 SLNet^[4]、经典霍夫变换与边缘检测基线和本文提出的深度霍夫网络方法。其中, SLNet^[4] 和 HT 均需要使用 HED^[43] 边缘检测器作为预处理步骤。SLNet 中的非最大抑制需要边缘图引导, 而经典霍夫变换需要以边缘图为输入。此外, SLNet 使用微调网络迭代地增强结果, 因此, 推理速度也与迭代次数相关。然而, 本文提出的方法仅需要一次前向传播便可得到结果, 而且 NMS 可简化为在参数空间中计算相连区域的重心。

在对比时间开销时, 由于语义线检测的时间开销容易受图像中语义线个数的影响, 本文利用 SEL 数据集中的所有图片, 对所有图片进行语义线检测, 并计算平均时间开销。5.3 展示了不同方法在不同阶段的时间开销, 结果表明, 本文提出的方法在速度上大幅度超过其他对比方法。本文设计的方法不需要任何额外处理, 如果不借助 ER 模块进行修复, 则不需要边缘检测。速度达到了 49FPS 的速度, 能够实时地进行图像语义线检测。其中, 即使仅使用一次迭代, SLNet 仍然慢于深度霍夫网络方法。说明本文提出的深度霍夫网络方法具有高效的运行速度, 同时拥有最好的表现性能。

5.2.3 定性比较

除了在数值上的定量对比外, 本文进行了检测结果的可视化, 通过可视化结果定性分析不同方法的表现。图 5.3 展示了 HED+HT、SLNet 迭代 1 和 10 次和 DHN 在 SEL 数据集上训练和测试的检测结果, 并与数据集标注 GT 对比。从检测结果可以看出, HED 结合霍夫变换的方法 (HED+HT) 容易产生漏检, 而 SLNet 的检测结果在第一次迭代后较乱, 随着不断迭代, 结果变好, 但不够精

确。本文提出的深度霍夫网络方法能够给出最合理的检测结果与真实标注 GT 有最一致的表现。

表 5.3 不同方法的检测速度比较。本文的深度霍夫网络方法 (不使用 ER 后处理模块) 在前向传播时远快于其他两个对比方法。

方法	网络推理	NMS	边缘检测	总时间
SLNet-iter1 ^[4]	0.354 s	0.079 s	0.014 s	0.447 s
SLNet-iter3 ^[4]	0.437 s	0.071 s	0.014 s	0.522 s
SLNet-iter10 ^[4]	0.827 s	0.068 s	0.014 s	0.909 s
HED ^[43] + HT ^[7]	0.014 s	0.117 s	0.024 s	0.155 s
DHN (VGG16)	0.030 s	0.003 s	0	0.033 s
DHN (ResNet50)	0.017 s	0.003 s	0	0.020 s

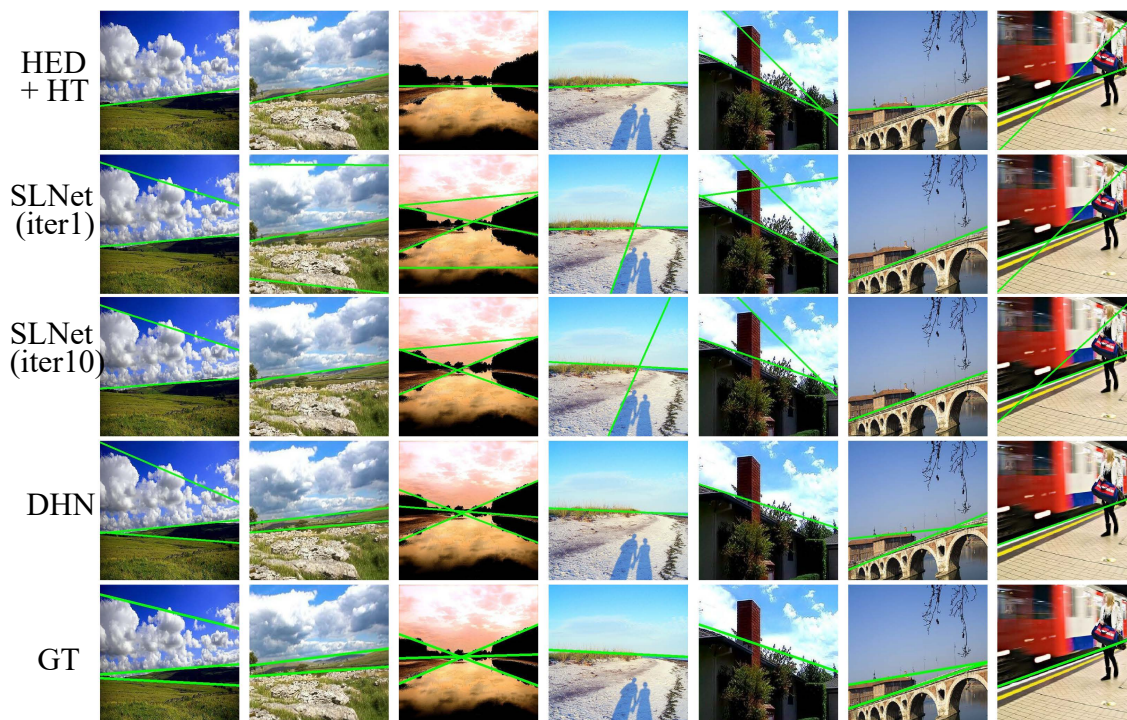


图 5.3 SEL 数据集上的不同方法检测结果示例。对比于 SLNet^[4] 和经典霍夫变换^[7], 本文的结果与真实数据更一致。

在本文提出的 NKL 数据集上, 给出了本文的深度霍夫网络方法在不同图片上的检测结果。图5.4深度霍夫网络方法在 NKL 上的检测结果。可以看出, 在不同的场景下, 如室外泳池、街道、建筑物、草原、河流、桥、室内货架等各式各样的条件下, DHN 均能给出合理的语义线检测结果。

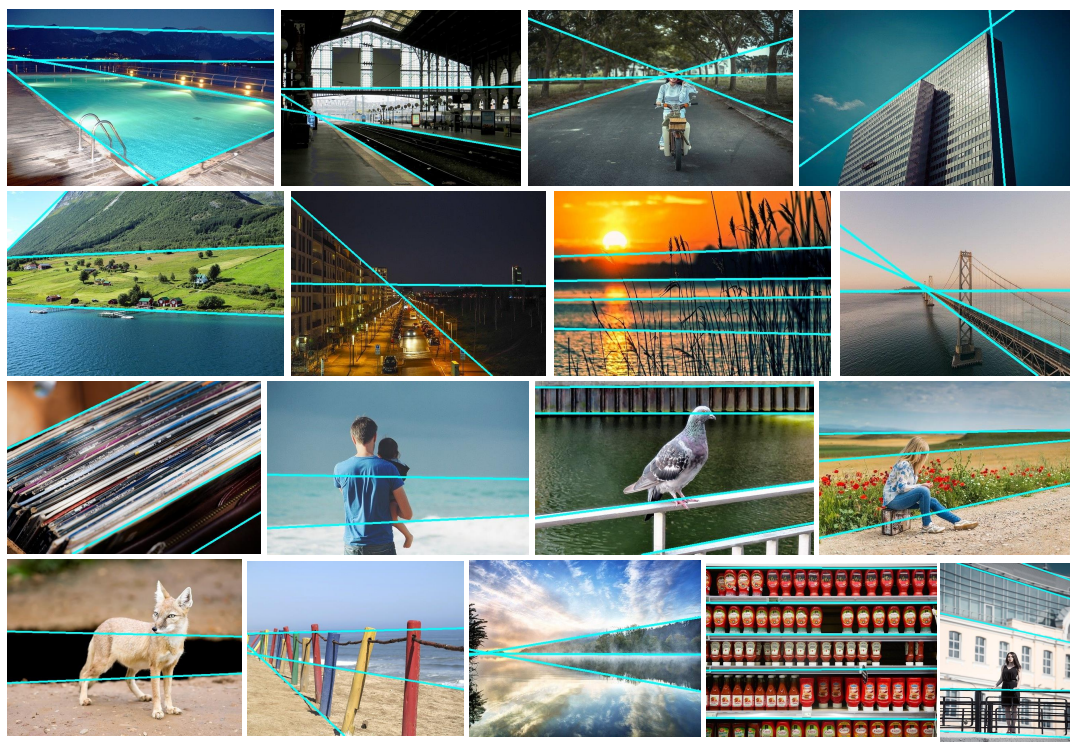


图 5.4 在NKL数据集上的检测结果。本文的深度霍夫网络方法可获得视觉上符合人类感知的结果。

同时，通过可视化结果可以看出，语义检测的结果刻画了图像中具有显著结构的线，这种结构具有语义划分功能，能够在辅助摄影构图、美学评估等的结构分析中发挥重要作用。

5.2.4 消融实验

在本节中，进行了消融实验来验证本文提出的 DHN 方法的每个组成成分的有效性，进行自我对比，包括 DHN 当中的不同模块间的消融实验和边缘引导微调模块的消融实验。

5.2.4.1 深度霍夫网络的组成部分

本节首先研究深度霍夫网络的组成部分。根据方法的不同阶段，将其分为：用于特征空间变换的深度霍夫变换（DHT）模块、用于构建多尺度特征的 FPN 结合的多尺度 DHT 结构（MS）、用于参数空间直线检测的内容感知线检测器（CTX）。

由于本文提出的 DHT 模块构建了全新的语义线检测方法流程，DHT 在该

流程中必不可少，因此本节首先使用朴素 ResNet50 和 DHT 模块构建基线模型，同样通过在参数空间中进行监督训练并使用霍夫反变换来得到输出结果。然后分别验证多尺度 DHT 策略（用 MS 表示）的效果和内容感知线检测器 CTX 的效果。分别将多尺度 DHT 和 CTX 加在基线模型后，分别验证其性能。最后展示了所有步骤结合的总体效果。表5.4展示了不同步骤在 SEL 数据集上训练得到的独立效果，结果表明 MS 和 CTX 可以提升 DHT 基线模型的性能。当使用所有组件组合成完整方法时，模型取得最好的性能。本节中的实验结果清晰地表明，每个组成部分对于本文提出的方法都有帮助。

表 5.4 对每个组成部分的消融实验。DHT 代表深度霍夫变换与 ResNet50 构成的基线，MS 代表使用多尺度特征 DHT，CTX 代表感知线检测器。

DHT	MS	CTX	F-measure
✓			0.664
✓	✓		0.758
✓		✓	0.771
✓	✓	✓	0.786

5.2.4.2 边缘引导的线微调

在前面的对比实验中，边缘引导的线微调（ER）并没有被使用，因为 ER 是作为方法的一个扩充选择，仅用于对位置精度有超高要求时使用。本节验证了 ER 作用的有效性，并给出了本文提出的深度霍夫网络配合 ER 使用时的数值结果。首先，测试了使用不同 δ_r 时 ER 模块的性能。其中参数 δ_r 控制 ER 的搜索空间（公式 (3.8) 中的 \mathcal{L} ）。实验结果通过采用 ResNet50 作为骨干的网络在 SEL 数据集上进行展示。

表5.5中结果表明随着 δ_r 提升，性能首先提升随后饱和。而随着 δ_r 的提升，搜索空间变大，计算的开销与计算时间随之变大，因此本文选择刚出现饱和状态时 δ_r 的值作为默认值。即峰值性能在 $\delta_r = 5$ 时出现，设 $\delta_r = 5$ 为最优选择。将 δ_r 设置为 5 后，测试了深度霍夫网络方法使用不同骨干网络和数据集时，ER 后处理的表现性能。

表5.6展示了使用不同骨干网络在 NKL 和 SEL 数据集上的表现结果，由于高精度对比容易在高阈值下体现，本表格除了给出平均的准确率、召回率和 F-measure 值外，还给出了在相似度阈值 0.95 下的 F-measure 值，记作 F@0.95，

表 5.5 使用不同 δ_r 时 DHN+ER 的性能。模型以 ResNet50 为骨干网络在 SEL 数据集上训练/测试。 $\delta_r = 0$ 代表不使用 ER 的原始 DHN 方法。

δ_r	Precision	Recall	F-measure
0	0.8190	0.7530	0.7861
1	0.8199	0.7561	0.7866
3	0.8208	0.7569	0.7874
5	0.8214	0.7574	0.7880
7	0.8213	0.7573	0.7878
9	0.8212	0.7571	0.7877

表 5.6 使用不同骨干网络和数据集时，模型在使用 ($\delta_r = 5$) 和不使用 ER 时的性能。

Dataset	Arch	ER	P	R	F	F@0.95
SEL ^[4]	VGG16		0.756	0.774	0.765	0.380
	VGG16	✓	0.758	0.777	0.770	0.439
	Resnet50		0.819	0.753	0.786	0.420
	Resnet50	✓	0.821	0.757	0.788	0.461
NKL	VGG16		0.659	0.759	0.706	0.434
	VGG16	✓	0.664	0.765	0.711	0.472
	Resnet50		0.679	0.766	0.719	0.459
	Resnet50	✓	0.684	0.771	0.725	0.486

该值越高，说明高精度的检测结果越准确。表格中的结果清晰地表明，边缘引导的微调模块可以在不同骨干模型和数据集下有效地提升检测结果。

5.2.4.3 其他框架和设备的结果

本节主要对比深度霍夫网络在 Jittor 深度学习框架上的运行效率，以及在不同的 GPU 平台执行的效率。Jittor 是清华大学自主研发的深度学习框架，可用于神经网络的训练和推理，得益于“元算子融合”和“统一计算图”技术，神经网络在 Jittor 框架下能够发挥更极致的速度。为了进一步加速本文提出的深度霍夫网络方法，使用 Jittor 框架进行 DHN 的实现，并对比其相比朴素 PyTorch 实现版本的加速情况。

表5.7给出了深度霍夫网络在不同设备和平台的运行效率，以 FPS 为计量单位。分别使用了 TITAN XP、Tesla P100、RTX 2080Ti、Tesla V100 (16G PCI-E)、Tesla V100 和 RTX TITAN 这几个不同型号的 GPU 作为计算硬件来进行对比。图中 bs 表示网络推理时使用的批处理图像数量。在进行效率对比时，只计算了深度霍夫网络的推理开销，而忽略了后处理过程，例如二值化与求连通分量

表 5.7 使用不同 GPU 设备在不同平台下的推理速度对比，使用 FPS 作为衡量单位且不考虑连通区域求解与 ER 后处理过程。

	TITAN XP			Tesla P100			RTX 2080Ti		
	bs=1	bs=4	bs=8	bs=1	bs=4	bs=8	bs=1	bs=4	bs=8
Jittor	44	54	56	42	49	52	82	98	100
PyTorch	39	48	49	35	44	44	64	71	71
加速比	1.13	1.13	1.14	1.20	1.11	1.18	1.28	1.38	1.41

	Tesla V100 (16G PCI-E)			Tesla V100			RTX TITAN		
	bs=1	bs=4	bs=8	bs=1	bs=4	bs=8	bs=1	bs=4	bs=8
Jittor	89	115	120	88	108	113	27	74	106
PyTorch	38	75	82	10	34	53	9	15	34
加速比	2.34	1.53	1.46	8.80	3.18	2.13	3.00	4.93	3.12

质心的过程没有进行时间开销的计算，且没有使用边缘引导的线微调 ER 模块。可以看到，在 TITAN XP 和 Tesla P100 上，使用 Jittor 平均有 1.15 左右的加速，而在 Tesla V100、RTX TITAN 等新款的显卡上，针对不同的 bs，几乎都有两三倍的加速，其中 RTX TITAN 在 bs 为 8 时，能达到 3.12 倍的加速。得益于 Jittor 的算子融合技术，DNH 方法在该实现下能够进行高速实视检测，能够轻松应用于视频等复杂的大规模数据。

5.2.5 开放自然场景检测结果

最后，在本小节中，将本文提出的深度霍夫网络应用在自然开放场景，并测试检测结果。首先，在互联网上检索了无版权的公开图片，且内容为自然生活中的常见场景，模拟用户的真实使用条件。同时，也手工地在自然生活中地不同场景采集了不同的图片，用于方法的检验。

图5.5展示了在互联网检索的自然生活场景上的检测结果，可以看出，在丰富的自然场景中，检测结果都自然合理，且足够准确。同时，图5.6展示了手工采集的自然图像上的语义线检测表现，从结果可以看出，本文提出的深度霍夫网络方法在不同的真实场景下具有一定的检测鲁棒性，对真实世界的摄影构图等具有应用价值。

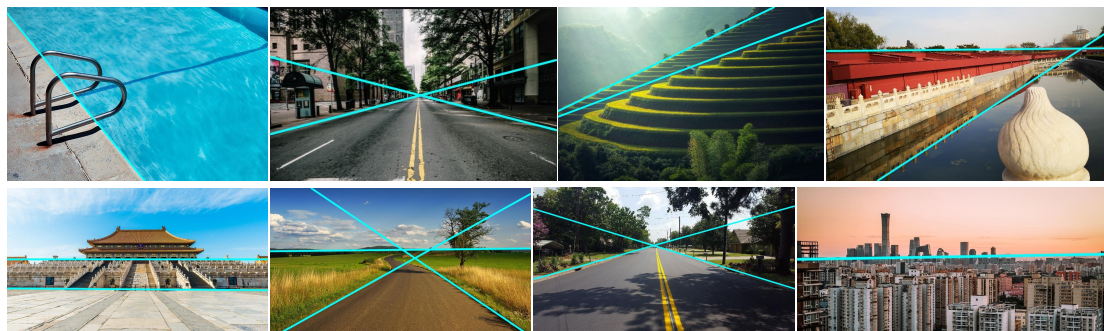


图 5.5 深度霍夫网络在检索的自然生活场景中的检测结果

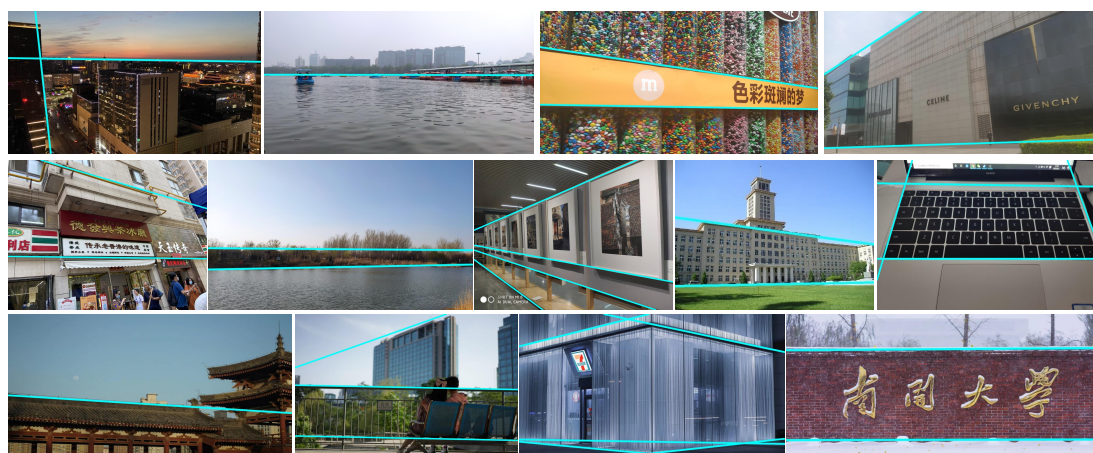


图 5.6 深度霍夫网络在手工采集的自然生活场景中的检测结果

第三节 方法的应用

语义线的检测可以应用在许多领域，如水平线检测与校正^[95]、图像抽象化^[4]和摄影构图^[1]。本节依靠本文提出的深度霍夫网络方法，对自然图像进行语义线检测，并将其应用在这三个领域，展示应用的结果。

首先是水平线的检测与矫正应用。水平线是自然图像中水平面所处的线，常见于海天交界线，城市的天际线等。在应用时，通过对水平线的检测，将图像依据参考水平线对齐，得到旋转矫正后的图像，产生更美观的场景。图5.7展示了在自然图像上水平线的检测结果，与水平对齐的旋转矫正后的图像。其中，由于水平线也是一种特殊的语义线，因此，直接使用语义线检测方法即可检测出水平线。然后将图像按照水平线的角度旋转到水平，并裁剪到规则矩形。而语义线检测到多条时，以深度霍夫网络输出置信度图中值较大的为准。从图中呈现的结果可以看出，经过水平线旋转矫正的图像能够产生更优美的照片。

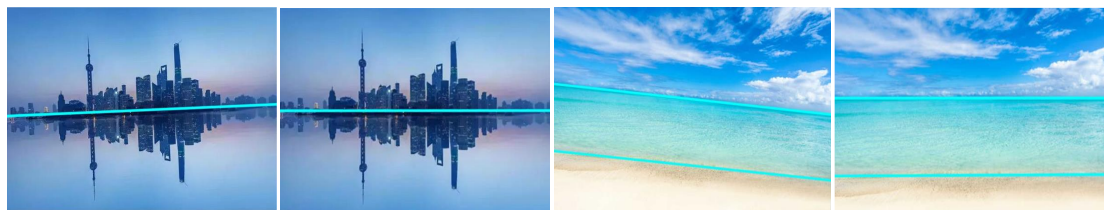


图 5.7 基于语义线检测的水平线矫正

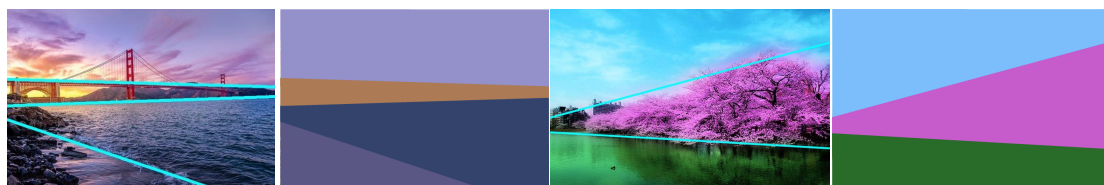


图 5.8 基于语义线检测的图像抽象化

然后是图像抽象化应用。图像抽象化通过线对图像进行区域分割，然后通过计算每个区域的颜色均值，对图像进行抽象。通过图像抽象，图像可以被表示成几个离散的块区域，对图像理解任务具有潜在价值。图5.8展示了图像语义线检测结果，以及通过抽象化的表现结果，其中直线通过本文提出的深度霍夫网络进行语义线检测得到。在进行抽象化时，根据检测到的语义线将图像划分成若干区域，然后求每个子区域内所有像素的均值颜色作为抽象的颜色。可以看出，经过图像抽象化，图像被简化表达为了几个离散的区域。

最后是摄影构图上的应用。常见的摄影构图手段包括三分构图法、黄金分割比构图法、对角线构图法等，这些构图方法往往依赖于图像中具有主导性的线结构与图像中被拍摄的主导物体。而图像中的主导线结构，可以通过语义线检测实现。本节通过深度霍夫网络进行语义线检测，作为图像中的主导线，同时使用显著性物体检测方法^[96]来进行显著性物体的检测。通过语义线检测结果与显著性检测结果，通过常见的摄影手段进行图像裁剪与旋转的重构图。图5.9展示了三分构图法的调整结果，其中左侧为原始图像，中间为显著物体检测和语义



图 5.9 基于语义线检测的图像重构图

线检测的结果，右侧为依据检测结果进行的依靠三分构图法的重裁剪结果。从图中可以看出，经过重构图后，图像的美观性大大增强，为摄影后期提供了有价值的参考性。

第四节 本章小结

本章主要对本文提出的用于语义线检测的深度霍夫网络方法进行了验证。首先对实验的环境，包括参数的设置和硬件的情况加以说明，阐述了具体的实现与优化细节。然后进行了对比实验，分析了方法的敏感参数，并在 SEL 和 NKL 数据集上进行了和其他方法的定量对比和定性可视化对比，实验结果表明，本文提出的深度霍夫网络方法在效率和效果上均远超现有的语义线检测方法和基线方法。在与其他方法对比后，进行了消融实验与分析，验证了深度霍夫网络中每个模块的有效性，并对边缘引导的线微调模块进行了充分分析与验证，结果表明了每个模块都是必不可少的，且加入边缘引导的微调模块后，可以让深度霍夫网络的输出得到进一步修正，取得更高的表现结果。然后利用 Jittor 平台，在不同的 GPU 设备上测试了方法的推理性能。最后，本节展示了文章提出的深度霍夫网络方法在真实的网络图片和自然采集的图片上的检测结果，验证了方法的鲁棒性，并给出了语义线检测结果在图像水平矫正、图像抽象化，图像重构图上的应用，说明了本文提出的深度霍夫网络方法的实用价值。

第六章 总结展望

第一节 文章总结

本文针对语义线检测任务进行了分析和研究，该问题对于视觉技术辅助的摄影构图、美学评估等具有重要意义。现有的直线检测和语义线检测方法一部分依赖于边缘检测和霍夫变换，这部分方法难以有效提取图像语义信息，另一部分基于深度学习的语义线检测方法忽略了直线的本质特性，难以有效地进行直线间的信息交互。本文针对以上问题，总结了现有的直线检测和语义线检测方法，并在这基础上提出了一种基于霍夫变换和深度学习结合的深度霍夫网络方法，该方法能够有效地处理图像语义信息，同时进行多尺度直线特征交互。此外，本文对语义线检测任务的数据集和现有的评价指标进行了总结，指出了其中数据集较为匮乏和评价指标存在不一致和歧义性的问题，提出了更大规模的语义线检测数据集以及更适用于语义线检测结果度量的评价指标和评价流程。本文的主要研究内容和贡献点可以总结如下：

(1) 本文总结了现有的直线检测和语义线检测方法，以及相关的研究技术，分析了这些方法在语义线检测中存在的问题。

(2) 本文提出了一个全新的用于直线检测的端到端训练的深度霍夫网络 DHN，该网络通过多尺度卷积网络进行富尺度特征的提取，并使用了一种全新的深度霍夫变换模块，将图像空间的表征变化到参数空间，在参数空间中使用内容感知的直线检测器进行直线间的信息交互和直线检测，将图像空间中直线检测的任务转换为了在参数空间中点检测的任务，并通过霍夫反变换与简单的连通分量质心计算得到直线的参数方程。同时，提出了一个用于深度霍夫网络输出结果修正的边缘指导的线微调后处理模块，根据图像的边缘信息进一步精细修正了检测结果。

(3) 本文对语义线检测的数据集进行了分析，指出了现有的语义线检测数据集 SEL 中样本量少、场景单一的问题，提出了规模更大、场景更丰富的语义线检测数据集 NKL，并进行了详细的统计分析，该数据集对语义线检测领域的发展提供了数据支撑。

(4) 本文对语义线检测的评价指标进行了分析, 分析了现有基于直线交并比的评价指标存在不一致性和歧义性的问题, 同时分析了其他评测指标如 CD 距离和 EMD 距离在语义线检测评价中的适用性, 提出了应用于衡量图像中两个直线相似度的全新评价指标 EA Score。该指标具有直观的可解释性, 且不受图像分辨率影响, 计算便捷。同时, 阐述了基于二分图的匹配检测结果评价流程, 使评价过程更直观合理。

(5) 本文进行了充分的实验验证, 通过定量实验和定性对比, 将提出的深度霍夫网络和现有方法进行了对比, 实验结果表明深度霍夫网络在准确度和计算效率上均远超其他方法。通过消融实验分析, 证明了深度霍夫网络每一个部分的有效性, 对可选的边缘指导的线微调后处理模块的有效性进行了充分分析。在不同的 GPU 硬件设备和实现平台上进行了推理速度的对比分析。最后, 将深度霍夫网络应用在自然场景上以验证方法的鲁棒性, 并展示了一些基于语义线检测的应用效果。

第二节 未来展望

本文设计的深度霍夫网络为语义线检测提供了一个新的方案, 并取得了突出的结果。然而, 深度霍夫网络作为一个新范式, 在设计上力求简洁而非极致的表现效果, 仍然存在一些可以继续提升的部分。本文在最后将讨论未来如何对深度霍夫网络进一步提升来增强语义线检测的性能。具体地, 未来该工作的改进可以主要关注以下两个方面:

(1) 深度霍夫网络构建了多尺度特征表达, 在每个尺度分别进行了深度霍夫变换与特征交互, 缺少了跨尺度间的信息交互。未来可以在参数空间设计特殊的跨尺度信息交互模块, 以进一步增强特征的表征能力, 提升检测效果。

(2) 深度霍夫变换作为网络的核心模块, 将深度特征表征转换到直线的参数空间中, 这一模块配合霍夫反变换, 可以为现有的线段检测、边缘检测、直线检测网络提供特征增强能力。未来可以加以细微修改, 以泛化推广成为一种即插即用的线特征增强模块, 扩大方法的适用领域。

参考文献

- [1] LIU L, CHEN R, WOLF L, et al. Optimizing Photo Composition. [J]. Computer Graphics Forum, 2010, 29 (2): 469–478.
- [2] 王耀艇. 基于构图规则的图像美学优化方法研究. [D]. 硕士论文. 天津工业大学, 2022.
- [3] 王伟凝, 刘剑聪, 徐向民, 等. 基于构图规则的图像美学优化. [J]. 华南理工大学学报: 自然科学版, 2015 (5): 51–58.
- [4] LEE J.-T, KIM H.-U, LEE C, et al. Semantic line detection and its applications. [C] // International Conference on Computer Vision. IEEE, 2017: 3229–3237.
- [5] BURNS J B, HANSON A R, RISEMAN E M. Extracting straight lines. [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1986, 8 (4): 425–455.
- [6] HOUGH P V. Method and means for recognizing complex patterns. US Patent 3,069,654. 1962.
- [7] DUDA R O, HART P E. Use of the Hough transformation to detect lines and curves in pictures. [R]. Sri International Menlo Park Ca Artificial Intelligence Center, 1971.
- [8] BALLARD D. Generalizing the Hough Transform to detect arbitrary shapes. [J]. Pattern Recognition, 1981, 13 (2).
- [9] ILLINGWORTH J, KITTLER J. The adaptive Hough transform. [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1987, 9 (5): 690–698.
- [10] KIRYATI N, ELDAR Y, BRUCKSTEIN A M. A probabilistic Hough transform. [J]. Pattern Recognition, 1991, 24 (4): 303–316.
- [11] XU L, OJA E, KULTANEN P. A new curve detection method: randomized Hough transform (RHT). [J]. Pattern Recognition Letters, 1990, 11 (5): 331–338.
- [12] VON GIOI R G, JAKUBOWICZ J, MOREL J.-M, et al. LSD: A fast line segment detector with a false detection control. [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2008, 32 (4): 722–732.
- [13] CANNY J. A computational approach to edge detection. [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1986, 8 (6): 679–698.
- [14] ZHAO K, GAO S, WANG W, et al. Optimizing the F-measure for Threshold-free Salient Object Detection. [C] // International Conference on Computer Vision. IEEE, 2019: 8849–8857.
- [15] HOU Q, CHENG M.-M, HU X, et al. Deeply supervised salient object detection with short connections. [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019, 41 (4): 815–828.
- [16] GAO S.-H, TAN Y.-Q, CHENG M.-M, et al. Highly Efficient Salient Object Detection with 100K Parameters. [C] // European Conference on Computer Vision. Springer, 2020.

-
- [17] BORJI A, CHENG M.-M, HOU Q, et al. Salient object detection: A survey. [J]. *Computational Visual Media*, 2019, 5 (2): 117–150. ISSN: 2096-0662.
- [18] WANG W, SHEN J, XIE J, et al. Revisiting Video Saliency Prediction in the Deep Learning Era. [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021, 43 (1): 220–237.
- [19] CHENG M.-M, MITRA N J, HUANG X, et al. Global Contrast based Salient Region Detection. [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, 37 (3): 569–582.
- [20] ZHU W, LIANG S, WEI Y, et al. Saliency optimization from robust background detection. [C] // *IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2014: 2814–2821.
- [21] FAN R, CHENG M.-M, HOU Q, et al. S4Net: Single Stage Salient-Instance Segmentation. [J]. *Computational Visual Media*, 2020, 6 (2): 191–204.
- [22] FREEMAN M. *The photographer's eye: composition and design for better digital photos*. [M]. CRC Press, 2007.
- [23] CHENG M.-M, LIU X.-C, WANG J, et al. Structure-Preserving Neural Style Transfer. [J]. *IEEE Transactions on Image Processing*, 2020, 29: 909–920.
- [24] HU S.-M, ZHANG F.-L, WANG M, et al. PatchNet: a patch-based image representation for interactive library-driven image editing. [J]. *ACM Transactions on Graphics*, 2013, 32 (6): 1–12.
- [25] KO K, LEE J.-T, KIM C.-S. PAC-Net: pairwise aesthetic comparison network for image aesthetic assessment. [C] // *IEEE International Conference on Image Processing*. IEEE, 2018: 2491–2495.
- [26] LEE J.-T, LEE C, KIM C.-S. Property-Specific Aesthetic Assessment With Unsupervised Aesthetic Property Discovery. [J]. *IEEE Access*, 2019, 7: 114349–114362.
- [27] KONG S, SHEN X, LIN Z, et al. Photo aesthetics ranking network with attributes and content adaptation. [C] // *European Conference on Computer Vision*. Springer, 2016: 662–679.
- [28] MAI L, JIN H, LIU F. Composition-preserving deep photo aesthetics assessment. [C] // *IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2016: 497–506.
- [29] FAN R, WANG X, HOU Q, et al. SpinNet: Spinning convolutional network for lane boundary detection. [J]. *Computational Visual Media*, 2019, 5 (4): 417–428.
- [30] KRAGES B. *Photography: the art of composition*. [M]. Simon, Schuster, 2012.
- [31] HU S.-M, XU K, MA L.-Q, et al. Inverse image editing: Recovering a semantic editing history from a before-and-after image pair. [J]. *ACM Transactions on Graphics*, 2013, 32 (6): 1–11.
- [32] CHEN T, CHENG M.-M, TAN P, et al. Sketch2photo: Internet image montage. [J]. *ACM Transactions on Graphics*, 2009, 28 (5): 1–10.
- [33] ZHANG S.-H, ZHOU Z.-P, LIU B, et al. What and where: A context-based recommendation system for object insertion. [J]. *Computational Visual Media*, 2020, 6: 79–93.

-
- [34] CAPLIN S. Art and Design in Photoshop. [M]. Elsevier/Focal, 2008. ISBN: 9780240811093.
- [35] AKINLAR C, TOPAL C. EDLines: A real-time line segment detector with a false detection control. [J]. Pattern Recognition, 2011, 32 (13): 1633–1642.
- [36] ZHANG Z, LI Z, BI N, et al. PPGNet: Learning Point-Pair Graph for Line Segment Detection. [C] // IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2019.
- [37] REN S, HE K, GIRSHICK R, et al. Faster r-cnn: Towards real-time object detection with region proposal networks. [C] // Annual Conference on Neural Information Processing Systems. 2015: 91–99.
- [38] LAW H, DENG J. Cornernet: Detecting objects as paired keypoints. [C] // European Conference on Computer Vision. Springer, 2018: 734–750.
- [39] GIRSHICK R. Fast r-cnn. [C] // International Conference on Computer Vision. IEEE, 2015: 1440–1448.
- [40] WANG X, GIRSHICK R, GUPTA A, et al. Non-local neural networks. [C] // IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2018: 7794–7803.
- [41] HUANG Z, WANG X, HUANG L, et al. Ccnet: Criss-cross attention for semantic segmentation. [C] // International Conference on Computer Vision. IEEE, 2019: 603–612.
- [42] LIU Y, CHENG M.-M, HU X, et al. Richer Convolutional Features for Edge Detection. [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019, 41 (8): 1939–1946.
- [43] XIE S, TU Z. Holistically-nested edge detection. [C] // International Conference on Computer Vision. IEEE, 2015: 1395–1403.
- [44] KITTLER J. On the accuracy of the Sobel edge detector. [J]. Image and Vision Computing, 1983, 1 (1): 37–42.
- [45] TORRE V, POGGIO T A. On edge detection. [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1986 (2): 147–163.
- [46] MARTIN D R, FOWLKES C C, MALIK J. Learning to detect natural image boundaries using local brightness, color, and texture cues. [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2004, 26 (5): 530–549.
- [47] DOLLÁR P, ZITNICK C L. Fast edge detection using structured forests. [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2014, 37 (8): 1558–1570.
- [48] BERTASIUS G, SHI J, TORRESANI L. Deepedge: A multi-scale bifurcated deep network for top-down contour detection. [C] // IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2015: 4380–4389.
- [49] XU D, OUYANG W, ALAMEDA-PINEDA X, et al. Learning deep structured multi-scale features using attention-gated crfs for contour prediction. [J]. Annual Conference on Neural Information Processing Systems, 2017, 30.
- [50] LEAVERS V F. Active intelligent vision using the dynamic generalized Hough Transform. [C] // The British Machine Vision Conference. Citeseer, 1990: 1–6.

-
- [51] TSAI D.-M. An improved generalized Hough transform for the recognition of overlapping objects. [J]. *Image and Vision computing*, 1997, 15 (12): 877–888.
- [52] GALAMHOS C, MATAS J, KITTLER J. Progressive probabilistic Hough transform for line detection. [C] // *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. Vol. 1. IEEE, 1999: 554–560.
- [53] GALAMBOS C, KITTLER J, MATAS J. Gradient based progressive probabilistic Hough transform. [J]. *IEE Proceedings Vision, Image and Signal Processing*, 2001, 148 (3): 158–165.
- [54] FERNANDES L A, OLIVEIRA M M. Real-time line detection through an improved Hough transform voting scheme. [J]. *Pattern Recognition*, 2008, 41 (1): 299–314.
- [55] LIMBERGER F A, OLIVEIRA M M. Real-time detection of planar regions in unorganized point clouds. [J]. *Pattern Recognition*, 2015, 48 (6): 2043–2053.
- [56] PRINCEN J, ILLINGWORTH J, KITTLER J. A hierarchical approach to line extraction based on the Hough transform. [J]. *Computer vision, graphics, and image processing*, 1990, 52 (1): 57–77.
- [57] YACOUB S B, JOLION J.-M. Hierarchical line extraction. [J]. *IEE Proceedings Vision, Image and Signal Processing*, 1995, 142 (1): 7–14.
- [58] CHEN Z, TAGLIASACCHI A, ZHANG H. Bsp-net: Generating compact meshes via binary space partitioning. [C] // *IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2020: 45–54.
- [59] DENG B, GENOVA K, YAZDANI S, et al. Cvxnet: Learnable convex decomposition. [C] // *IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2020: 31–44.
- [60] QI C R, LITANY O, HE K, et al. Deep hough voting for 3d object detection in point clouds. [C] // *International Conference on Computer Vision*. IEEE, 2019: 9277–9286.
- [61] MIN J, LEE J, PONCE J, et al. Hyperpixel flow: Semantic correspondence with multi-layer neural features. [C] // *International Conference on Computer Vision*. IEEE, 2019: 3395–3404.
- [62] ETEMADI A. Robust segmentation of edge data. [C] // *International Conference on Image Processing and its Applications*. IET, 1992: 311–314.
- [63] CHAN T, YIP R K. Line detection algorithm. [C] // *International Conference on Pattern Recognition*. Vol. 2. IEEE, 1996: 126–130.
- [64] AKINLAR C, TOPAL C. EDCircles: A real-time circle detector with a false detection control. [J]. *Pattern Recognition Letters*, 2013, 46 (3): 725–740.
- [65] FENG H, JIANG Z, XIE F, et al. Automatic fastener classification and defect detection in vision-based railway inspection systems. [J]. *IEEE Transactions on Instrumentation and Measurement*, 2013, 63 (4): 877–888.
- [66] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation. [C] // *IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2014: 580–587.

- [67] 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): 1904–1916.
- [68] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection. [C] // IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2016: 779–788.
- [69] REDMON J, FARHADI A. YOLO9000: better, faster, stronger. [C] // IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2017: 7263–7271.
- [70] LIU W, ANGUELOV D, ERHAN D, et al. Ssd: Single shot multibox detector. [C] // European Conference on Computer Vision. Springer, 2016: 21–37.
- [71] REDMON J, FARHADI A. Yolov3: An incremental improvement. [J]. ArXiv preprint arXiv:1804.02767, 2018.
- [72] HUANG L, YANG Y, DENG Y, et al. Densebox: Unifying landmark localization with end to end object detection. [J]. ArXiv preprint arXiv:1509.04874, 2015.
- [73] TIAN Z, SHEN C, CHEN H, et al. Fcos: Fully convolutional one-stage object detection. [C] // International Conference on Computer Vision. IEEE, 2019: 9627–9636.
- [74] CARION N, MASSA F, SYNNAEVE G, et al. End-to-end object detection with transformers. [C] // European Conference on Computer Vision. Springer, 2020: 213–229.
- [75] HUANG K, WANG Y, ZHOU Z, et al. Learning to parse wireframes in images of man-made environments. [C] // IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2018: 626–635.
- [76] ZHOU Y, QI H, MA Y. End-to-end wireframe parsing. [C] // International Conference on Computer Vision. IEEE, 2019: 962–971.
- [77] XUE N, WU T, BAI S, et al. Holistically-attracted wireframe parsing. [C] // IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2020: 2788–2797.
- [78] LIN T.-Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection. [C] // IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2017: 2117–2125.
- [79] LIU W, SALZMANN M, FUA P. Context-aware crowd counting. [C] // IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2019: 5099–5108.
- [80] CHENG Z.-Q, LI J.-X, DAI Q, et al. Learning spatial awareness to improve crowd counting. [C] // International Conference on Computer Vision. IEEE, 2019: 6152–6161.
- [81] TAN Y.-Q, GAO S, LI X.-Y, et al. VecRoad: Point-based Iterative Graph Exploration for Road Graphs Extraction. [C] // IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2020.
- [82] ZHOU B, LAPEDRIZA A, KHOSLA A, et al. Places: A 10 million Image Database for Scene Recognition. [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017.
- [83] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition. [C] // IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2016: 770–778.

-
- [84] RUBNER Y, TOMASI C, GUIBAS L J. The earth mover's distance as a metric for image retrieval. [J]. *International Journal of Computer Vision*, 2000, 40 (2): 99–121.
- [85] BORGEFORS G. Distance transformations in digital images. [J]. *Computer vision, graphics, and image processing*, 1986, 34 (3): 344–371.
- [86] KUHN H W. The Hungarian method for the assignment problem. [J]. *Naval research logistics quarterly*, 1955, 2 (1-2): 83–97.
- [87] PASZKE A, GROSS S, MASSA F, et al. PyTorch: An imperative style, high-performance deep learning library. [C] // *Annual Conference on Neural Information Processing Systems*. 2019: 8024–8035.
- [88] HU S.-M, LIANG D, YANG G.-Y, et al. Jittor: a novel deep learning framework with meta-operators and unified graph execution. [J]. *Science China Information Science*, 2020, 63 (222103): 1–21.
- [89] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition. [C] // *International Conference on Learning Representations*. 2015.
- [90] ZHAO H, SHI J, QI X, et al. Pyramid scene parsing network. [C] // *IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2017: 2881–2890.
- [91] CHEN L.-C, ZHU Y, PAPANDREOU G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation. [C] // *European Conference on Computer Vision*. Springer, 2018: 801–818.
- [92] GAO S.-H, CHENG M.-M, ZHAO K, et al. Res2Net: A New Multi-scale Backbone Architecture. [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021, 43 (2): 652–662.
- [93] YU F, KOLTUN V. Multi-scale context aggregation by dilated convolutions. [C] // *International Conference on Learning Representations*. 2016.
- [94] KINGMA D P, BA J. Adam: A method for stochastic optimization. [C] // *International Conference on Learning Representations*. 2015.
- [95] WORKMAN S, ZHAI M, JACOBS N. Horizon lines in the wild. [J]. *The British Machine Vision Conference*, 2016.
- [96] LIU J.-J, HOU Q, CHENG M.-M. Dynamic feature integration for simultaneous detection of salient object, edge, and skeleton. [J]. *IEEE Transactions on Image Processing*, 2020, 29: 8652–8667.

致谢

随着时间的流逝，硕士研究生学习生涯即将结束，三年求学路，虽忙碌但也十分充实。在毕业论文即将完成之际，回望三年的科研学习路途，自己收获了很多，也成长了很多，这些都离不开师长朋友和家人的支持与帮助，谨此向曾经给予我帮助和支持的人们表以衷心的感谢。

首先，我要感谢我的导师程明明教授。自 2019 年加入南开大学媒体计算实验室以后，程老师可谓是亦师亦友，无时无刻地关注着我的学习科研与生活。在科研学习上，程老师是我的启蒙者：从科研选题，到方法探究，再到实验与写作，程老师都不厌其烦地对我们进行教导与帮助，对于每一个方法细节，每一句文字表达，都帮助我反复地推敲、修改，帮助我养成了良好的科研习惯。同时程老师积极教育我们科研诚信，让我形成了正确的科研价值观。在生活上，程老师是我的陪伴者：程老师在生活中更像是我的朋友和兄长，会对我进行嘘寒问暖，关心日常生活，与我一起进行体育锻炼，陪伴我的成长。在此，衷心感谢程老师对我人生产长与科研上的帮助与陪伴。此外，我要感谢王井东老师和戴琦老师，感谢你们在我科研道路上对我的帮助，你们对我的指导让我终生受益。

时光荏苒，三年转瞬即逝，这期间我结识了许多科研伙伴与知心朋友，感谢他们对我科研学习上的帮助，以及生活上一起度过的快乐时光。感谢实验室的赵凯师兄和高尚华师兄，你们对我的科研成长帮助甚多。感谢实验室的小伙伴们、同学朋友们、师兄师弟们，一起度过了快乐的三年时光，一起品尝了科研道路上的酸甜苦辣和人生道路上的喜怒哀乐。此外，感谢我的室友三年来的陪伴，感谢班级每位同学对我的帮助。

最后，我要感谢我的家人，你们对我的全心支持和关心让我能够顺利走过求学的路程。感谢我的爱人对我的支持与陪伴。在你们的关心下，我能够安心求学、顺利毕业。

个人简历

韩琦，男，河北廊坊人，出生于1997年。在2019年毕业于西安电子科技大学计算机科学与技术专业并获得学士学位。于2019年至今在南开大学就读计算机技术研究生。

研究生期间发表论文：

- Han Q, Zhao K, Xu J, Cheng M M. Deep Hough Transform for Semantic Line Detection[C]//European Conference on Computer Vision (ECCV). Springer, Cham, 2020: 249-265. 第一作者
- Jin W D, Xu J, Han Q, Zhang Y, Cheng M M. CDNet: Complementary depth network for RGB-D salient object detection[J]. IEEE Transactions on Image Processing (TIP), 2021, 30: 3376-3390. 第三作者
- Gao S H, Han Q, Li D, Cheng M M. Representative batch normalization with feature calibration[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021: 8669-8679. 第二作者
- Gao S H, Han Q, Li Z Y, Peng P, Wang L, Cheng M M. Global2local: Efficient structure search for video action segmentation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021: 16805-16814. 共同第一作者
- Zhang C B, Jiang P T, Hou Q, Wei Y, Han Q, Li Z, Cheng M M. Delving deep into label smoothing[J]. IEEE Transactions on Image Processing (TIP), 2021, 30: 5984-5996. 第五作者
- Han Q, Fan Z, Dai Q, Sun L, Liu J, Cheng M M, Wang J. On the Connection between Local Attention and Dynamic Depth-wise Convolution[C]//International Conference on Learning Representations (ICLR), 2022. 第一作者
- Zhao K, Han Q, Zhang C B, Xu J, Cheng M M. Deep hough transform for semantic line detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 2021. 共同第一作者