

# 基于多尺度特征融合的直线检测算法

张亮 王震\*

西南民族大学 计算机与人工智能学院, 四川成都, 610225;

**摘要:** 直线检测是计算机视觉领域中的一个重要任务, 广泛应用于室内场景分析和建模等领域。现有的直线检测算法如 Holistically-Attracted Wireframe Parsing (HAWP) 算法在复杂的环境下仍具有一定的局限性, 由于部分环境结构复杂且部分区域光线黯淡、纹理较弱, 因此检测效果不佳。为了更加准确地检测出复杂环境下的直线, 针对 HAWP 算法进行研究改进, 通过引入 HRnet 网络作为 Backbone 网络融合图像多尺度的特征信息。实验表明改进的 HAWP 算法相比原算法在复杂环境下的直线检测任务的平均精度由 65.6% 提升到了 68.4%。

**关键词:** 直线检测; HAWP 算法; 复杂环境

**DOI:** 10.69979/3041-0673.25.01.080

## 引言

随着计算机视觉技术的发展, 复杂环境场景分析和建模成为研究热点。直线作为常见几何结构, 直线检测对智能导航、三维重建等应用意义重大。当前直线检测算法分传统与深度学习算法两类。

传统算法以霍夫变换和 LSD<sup>[1]</sup> 为主。霍夫变换将直线映射为参数空间的点来检测, 计算复杂度高且丢失长度信息; LSD 使用梯度方法, 先计算像素点水平线角度, 划分支持域再提取线段, 但对直线断裂鲁棒性差。改进的 Light-LSD<sup>[2]</sup> 在弱纹理场景可提高效率, 不过仍难处理直线间断, 影响准确率。

复杂场景的噪声、遮挡和光线变化等问题, 降低传统直线检测准确性。深度学习技术的成熟, 催生了基于深度学习的直线检测算法, 如 Wireframe<sup>[3]</sup>、L-CNN<sup>[4]</sup> 和 HAWP<sup>[5]</sup> 等, 这些方法可弥补传统方法对参数设置和噪声敏感的不足。Wireframe 网络分两路, 分别用来检测端点和分割直线像素, 但 Wireframe 不是端到端的可训练模型且结构复杂、计算量大; L-CNN 先通过 Backbone 网络提取特征图, 再依次预测连接点、线段提议并精细化, 但线段提议依赖连接点提议; HAWP 类似 L-CNN, 分别预测连接点和线段再匹配, 精度优于 L-CNN, 但 HAWP 采用的 Hourglass<sup>[6]</sup> 网络在复杂场景中不同层级特征图直线预测效果有差异。

本研究通过引入 HRnet<sup>[7]</sup> 网络作为新 Backbone 网络来融合图像多尺度特征信息, 提升线段预测的准确性和鲁棒性, 实验表明该方法在实际应用优势明显。

## 1 HAWP 算法模型

HAWP 是一个端到端的、Two-Stage 的直线检测算法。

HAWP 的结构可以分为三部分, 1) 提案初始化: 线段预测和连接点预测。给定一幅图像, 它首先经过一个共享的 Hourglass 网络。然后采用与 L-CNN 中相同的基于热图的头回归器, 从中选择 Top-K 个连接点作为初始连接点提议。2) 提议细化: 线段与连接点的匹配。HAWP 规定如果一个线段提议的两个端点有两个连接点提议支持, 则保留该线段提议。如果一个连接点提议没有找到任何支撑线段提议, 则将其删除。3) 提议验证: HAWP 使用与 L-CNN 中相同的兴趣线池化 (Line-of-Interest, LOI) 来计算线段的特征。图 1 是 HAWP 的网络结构示意图。

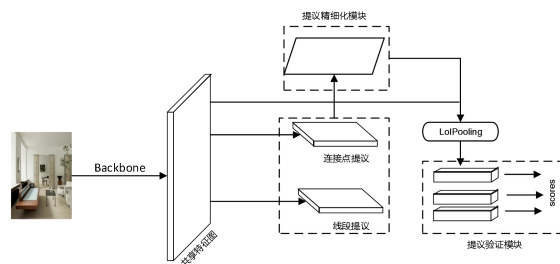


图 1 HAWP 的网络结构示意图

HAWP 以沙漏神经网络作为 Backbone, 沙漏神经网络可以融合图像多个尺度的信息并且能够不改变输入图像的尺寸。

## 2 改进的 HAWP 算法模型

多尺度特征融合能够充分利用图像的信息, 提高直线检测的准确率。因此我们希望生成一个能够融合多个尺度特征信息的高分辨率特征图。这就与一般的网络结构比如 VGG 的要求不同, 因为 VGG 网络最终得到的特征

图的分辨率很低,损失了大部分的空间结构。HAWP 算法在实现过程中选取了 Hourglass 网络作为 Backbone, Hourglass 网络的结构示意图如图 2 所示。Hourglass 网络获取高分辨率的方式采用的是先将分辨率降低,然后再提高分辨率。但是 Hourglass 网络结构本质上是不同分辨率的特征图之间进行了串联。

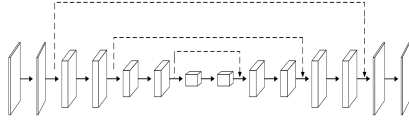


图 2 Hourglass 网络的结构

我们引入了当前在人体关键点检测任务中表现优秀的 HRnet 网络。HRnet 网络在不同分辨率特征图进行并联的基础上,添加了不同分辨率特征图之间的交互,如图 3 所示。

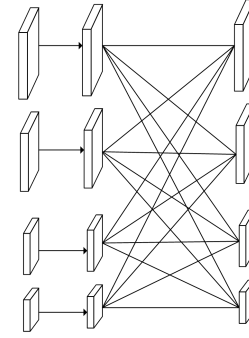


图 3 不同分辨率之间的交互

在 HRnet 中我们首先将 Layer1 模块替换为简单的残差模块。在此基础上为了尽可能减少网络参数,我们将 Stage3 模块与 Stage4 模块的堆叠次数设置为 2, Stage2 模块的堆叠次数保持不变。这样设计的 Backbone 网络可以更好地融合图像多个尺度的特征信息并降低计算量。改进后的 Backbone 结构如图 4 所示。

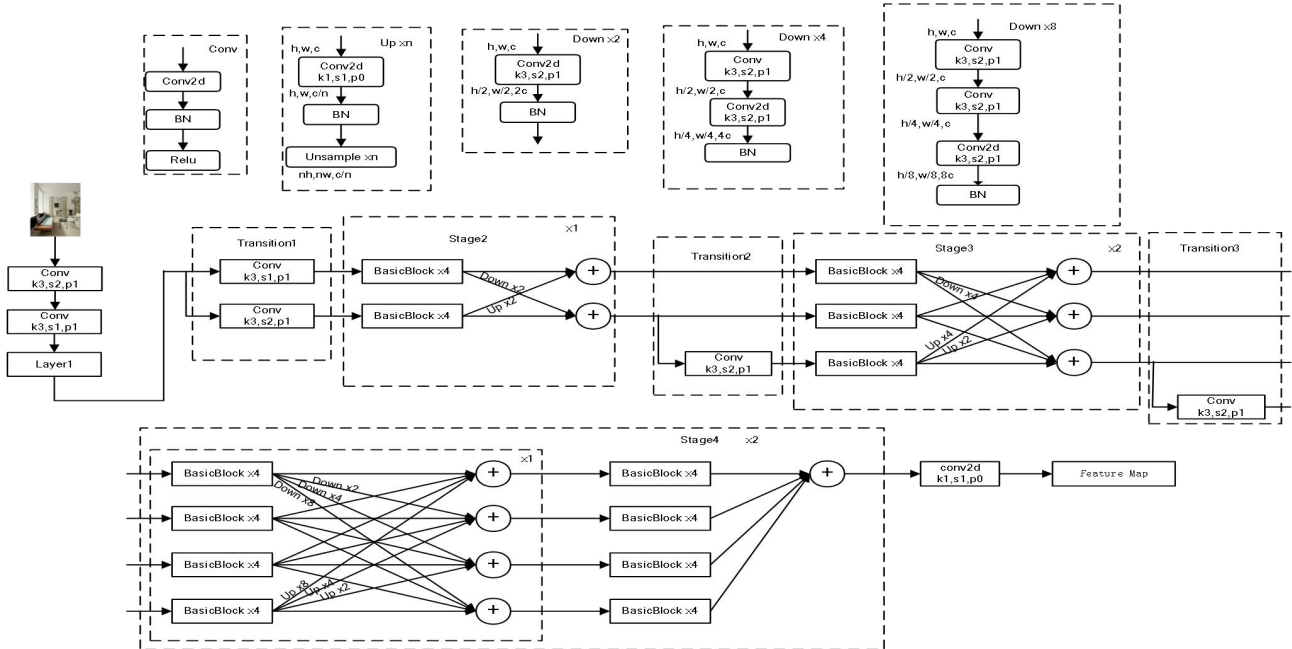


图 4 改进后的 Backbone 网络

### 3 实验结果与分析

#### 3.1 评价标准

我们使用线段结构平均精度 (Structural Average Precision of Line Segments) 作为评价标准。对于每一条 ground truth 中的线段,找出满足以下条件的线段:

$$|x_1 - \hat{x}_i|^2 + |x_2 - \hat{x}_j|^2 \leq V_L \quad (1)$$

其中  $V_L$  是预先设置的阈值。

在预测线段集合中:

(1) 如果解析线段匹配真值线段的集合为空,则将其指定为 False Negative ( $FN$ )

(2) 如果集合中有多条线段,则将概率最高的线段设置为 True Positive ( $TP$ ),其他的设置为 False Positive ( $FP$ )。

(3) 对于不属于任何真值线段对应集合的预测线段均被设置为  $FP$ 。

则 sAP 的计算方法如下:

$$\begin{cases} R = \frac{TP}{TP + FN} \\ P = \frac{TP}{TP + FP} \\ AP = \int_0^1 PdR \\ sAP = \frac{AP}{N} \end{cases} \quad (2)$$

根据  $V_L$  设置的值分别为 5、10、15，将 sAP 分别记为 sAP5、sAP10 与 sAP15。整体性能由不同阈值下的 sAP 的平均值来表示，记为 msAP。

### 3.2 实验数据及训练环境

实验中使用的数据集是当前专门用于直线检测的数据集 Wireframe Dataset 与 YorkUrban Dataset。为了解决样本数量少的问题，我们使用 python 脚本对数据进行增强操作，其中包括水平翻转、随机旋转、颜色扰动与更改通道顺序等。实验采用 Ubuntu22.04 操作系统、Pytorch 深度学习框架。硬件配置为 Nvidia GeForce RTX2080ti。

### 3.3 实验结果

实验表明改进的 HAWP 算法相较于原算法在 Wireframe Dataset 测试集的平均精度由 65.6%提升到了 68.4%。在 YorkUrban Dataset 测试集的平均精度由 32.8%提升到了 34.2%。表一是 LSD、L-CNN、HAWP 与改进 HAWP 在 Wireframe Dataset 测试集检测精度的对比。

表 1 不同算法在 Wireframe Dataset 测试集上的精度对比

方法	sAP5	sAP10	sAP15	msAP
LSD	/	/	/	/
L-CNN	57.9	62.8	65.0	61.9
HAWP	63.0	65.0	68.8	65.6
改进 HAWP	65.2	69.2	70.9	68.4

表二是 LSD、L-CNN、HAWP 与改进 HAWP 在 YorkUrban Dataset 测试集检测精度的对比。

表 2 不同算法在 YorkUrban Dataset 测试集上的精度对比

方法	sAP5	sAP10	sAP15	msAP
LSD	/	/	/	/
L-CNN	24.3	26.4	27.5	26.1
HAWP	30.4	33.3	34.8	32.8
改进 HAWP	31.8	34.7	36.0	34.2

实验结果表明，改进的 HAWP 在原算法基础上进一步提升了检测精度。

为了探究特征融合对模型精度的提升，我们将网络

Backbone 分别替换为 Resnet、Unet 与 HRnet 网络进行实验。表三是不同 Backbone 网络的模型在 Wireframe Dataset 测试集检测精度的对比。

表 3 不同的 Backbone 模型的精对比

Backbone	sAP5	sAP10	sAP15	msAP
Stacked Hourglass	64.0	68.0	69.8	67.2
Resnet	32.2	38.9	42.2	37.8
Unet	61.6	65.5	67.3	64.8
HRnet	64.4	68.4	70.1	67.6

Resnet 网络对图像进行下采样时没有进行特征信息的融合。Unet 先通过下采样降低分辨率再通过转置卷积还原分辨率，可以保留部分特征信息。HRnet 比 Stacked Hourglass 能够融合更多的特征信息。表三的实验结果表明，通过 Backbone 网络更好地融合特征图的信息可以提升检测精度。

图 5 是 LSD、L-CNN、HAWP 与改进的 HAWP 算法在单张弱光线复杂环境图片上的效果对比图。

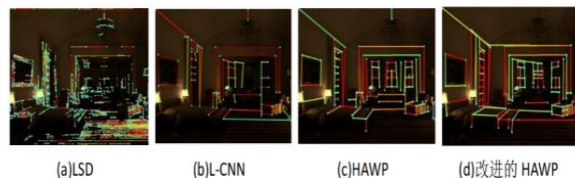


图 5 LSD、L-CNN、HAWP 与改进 HAWP 对比图

LSD 是经典的利用梯度进行计算的传统算法，在图 5(a)中 LSD 算法预测了大量的错误线段。L-CNN 是先预测连接点再根据连接点提议预测线段，在图 5(b)中 L-CNN 对光线较弱的纹理区域的预测效果不佳。图 5(c)与图 5(d)分别为 HAWP 算法和改进的 HAWP 算法的预测效果图。可以发现本文提出的改进的 HAWP 算法在弱纹理环境下的预测效果是几种算法中最好的。

### 4 结语

本文提出了基于 HAWP 改进的复杂环境直线检测算法，通过多尺度特征融合网络 HRnet 作为 Backbone 网络融合人图像多尺度特征信息。实验表明该算法在复杂环境下的检测效果良好，能够提升网络对于弱纹理区域直线预测的精度，相较于原版的 HAWP 算法，该算法的平均检测精度可达 68.4%，这对依赖于直线检测的定位与场景建模等领域具有重要意义。

### 参考文献

[1] Von Gioi R G, Jakubowicz J, Morel J M, et al. LSD: A line segment detector[J]. Image Proc

essing On Line, 2012, 2: 35-55.

[2]李玉峰, 张宸. 一种基于LSD改进的室内直线特征匹配算法[J]. 现代计算机, 2022, 28(09):15-21.

[3]Huang K, Wang Y, Zhou Z, et al. Learning to parse wireframes in images of man-made environments[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 626-635.

[4]Zhou Y, Qi H, Ma Y. End-to-end wireframe parsing[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019: 962-971.

[5]Xue N, Wu T, Bai S, et al. Holistically-att racted wireframe parsing[C]//Proceedings of th

e IEEE/CVF Conference on Computer Vision and P attern Recognition. 2020: 2788-2797.

[6]Newell A, Yang K, Deng J. Stacked hourglass networks for human pose estimation[C]//Comput er Vision-ECCV 2016, 2016: 483-499.

[7]Sun K, Xiao B, Liu D, et al. Deep high-reso lution representation learning for human pose estimation[C]//Proceedings of the IEEE/CVF con ference on computer vision and pattern recogni tion. 2019: 5693-5703.

作者简介: 张亮(1998—), 男, 汉族, 四川巴中人, 硕士研究生, 研究方向为人工智能。

王震(1999—), 男, 汉族, 四川巴中人, 硕士研究 生, 本文通信作者, 研究方向为计算机视觉。