

DOI:10.19344/j.cnki.issn1671-5276.2025.05.044

用于无人机输电线检测的双目立体匹配算法

朱程^{1,2,3}, 安德钰^{2,3}, 赖佳华³

(1. 福建师范大学 计算机与网络空间安全学院, 福建 福州 350117;

2. 中国科学院大学 福建学院, 福建 福州 350025;

3. 中国科学院海西研究院 泉州装备制造研究中心, 福建 泉州 362216)

摘要:在无人机应用中,针对输电线等细小物体的高效检测问题,设计一种基于 Transformer 和几何编码的立体匹配算法,利用交叉注意力机制跨图像交互整合另一幅图像的信息,恢复精细的深度细节。利用几何编码将传统的成本体积编码变成几何编码体积和上下文信息,以精确估计输电线目标,帮助无人机避障。实验结果表明:相较于 SOTA 的立体匹配算法,该算法在输电线深度估计方面表现更优,能够精确地估计输电线目标。通过提高细小物体检测的准确性,为无人机在电力巡检等实际应用场景中提供了有力支持,推动了无人机技术在工程领域的发展。

关键词:立体匹配;无人机;Transformer;输电线检测;深度估计

中图分类号:V279 **文献标志码:**A **文章编号:**1671-5276(2025)05-0224-04

Binocular Stereo Matching Algorithm for UAV Power Line Detection

ZHU Cheng^{1,2,3}, AN Deyu^{2,3}, LAI Jiahua³

(1. College of Computer and Cyberspace Security, Fujian Normal University, Fuzhou 350117, China;

2. Fujian College, University of Chinese Academy of Sciences, Fuzhou 350025, China;

3. Quanzhou Institute of Equipment Manufacturing, Fujian Institute of Research on the Structure of Matter, Chinese Academy of Sciences, Quanzhou 362216, China)

Abstract: To address the challenge to the efficient detection of small objects such as power lines in UAV applications, this study designs a stereo matching algorithm based on Transformer and geometry encoding, which leverages a cross-image attention mechanism to integrate information from another image, facilitating the recovery of fine-depth details. Geometry encoding is employed to transform the conventional cost volume into a geometry encoding volume, encode geometric and contextual information for precise estimation of power line targets and aid of UAV obstacle avoidance. Experimental results demonstrate that compared to state-of-the-art stereo matching algorithms, the proposed algorithm exhibits superior performance in power line depth estimation, capable of accurately estimating the targets. By enhancing the accuracy of small object detection, this study provides robust support for UAV applications in real-world scenarios such as power line inspection, and contributes to the advancement of UAV technology in engineering domains.

Keywords: stereo matching; UAV; Transformer; power line detection; depth estimation

0 引言

近年来,随着传感技术和人工智能的迅猛发展,无人机在各个领域取得了显著进展,并在不同行业被广泛应用。然而,对于输电线等微小物体的准确检测仍然是一项巨大的挑战。在无人机应用中,精准检测输电线至关重要,可避免碰撞,确保飞行安全。双目立体匹配通过计算两张图像之间的差异来获取物体的三维信息,而输电线深度估计则是其关键应用之一,利用双目立体匹配技

术估算输电线的三维深度,帮助无人机规避飞行中的输电线障碍物,确保飞行安全。

在深度学习领域,基于 cost volume 的深度立体匹配方法得到了广泛应用。DispNetC^[1]作为首个端到端可训练的框架,利用左右特征图的点积构建 3D 成本体积。GwcNet^[2]引入了组相关性,构建了更紧凑的成本量。另一方面,基于变换的深度立体匹配方法在注意力机制下成为研究的另一方向。STTR 从序列到序列匹配像素,通过施加唯一性约束避免构造固定视差成本体积。在无人机检测领域,基于深度学习的方法如 YOLOv4^[3]

基金项目:福建省科技计划引导性项目(2022H0042)

第一作者简介:朱程(1997—),男,山东泰安人,硕士研究生,研究方向为计算机视觉、双目立体匹配,Rc_0507@163.com。

和 Faster-RCNN^[4] 成功应用于任务。

1 模型原理

本文设计了一种基于 Transformer 和几何编码的立体匹配模型,其局部-全局立体匹配网络如图 1 所示。该模型主要由两个关键组件组成:基于 Transformer 的特征增强模块和几何编码体编码模块。对于给定的校正图像对 I_l (左图像) 和 I_r (右图像),通过卷积神经网络 (CNN) 特征提取器提取 8 倍下采样的密集特征 F_l 和 F_r ,其中 H 、 W 和 D 分别表示高度、宽度和特征维度。随后,将这些提取的浅层特征输入通道注意力 (transformer cross attention, TCA) 模块,以进行特征增强,从而构建成本体积。最终,逐步将这些成本量送入轻量级的 3D 正则化网络,并通过 Self-Attention (SA) 模块处理不匹配区域,从而得到最终的视差图。这一流程的运用有助于提高立体匹配的准确性和鲁棒性。

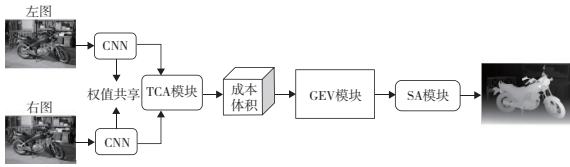


图1 所提算法的网络结构

受 Restormer^[5] 启发,本文提出了一个通道注意力(TCA)模块进行特征增强,利用滑动窗口基于交叉视图特征相似性有选择性地聚合信息。这使得 Transformer 模块能够有效地将编码器的低级图像特征与解码器的高级特征融合,从而有助于在图像中保留精细的结构和纹理细节。图 2 展示了 TCA 模块的详细结构。本文所提出的一维水平交叉注意力是为校正图像量身定制的,旨在同时捕捉长距离像素依赖关系并尽可能保留更多高频精细结构信息。这种设计不仅加速了模型的学习过程,还增强了对复杂场景的建模能力。其中 $\tilde{F}$ 是融合后的特征, F_l 和 F_r 分别是左、右特征, T 是聚合后的特征。

$$\tilde{F} = \tau[(F_l + T), (F_r + T)] \quad (1)$$

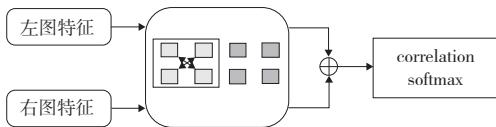


图2 TCA 模块

为了实现对尖锐边缘的准确建模和更好地处理弱纹理区域,保持通道中的高频信息变得至关重要。在整个结构中保持高分辨率是直观的方法,但由于计算成本极高,本文选择通过像素解缩将图像下采样缩小到原始大小的 1/8,同时扩展通道而不丢失任何高频信息。具体而言,原始图像的形状为 $[C, H \times r, W \times r]$,在像素解缩后被重新塑造为 $[C \times r^2, H, W]$ 。

根据 TCA 模块增强的左特征 F_l 和右特征 F_r ,构建了一个相关性成本体积。然而,这个成本体积仅基于特征相关性,缺乏捕获全局几何结构的能力。为了解决这个问题,采用了一个轻量级的三维正则化网络 R 来处理 C_{corr} ,得到几何编码体积 C_G 。

$$C_G = R(C_{\text{corr}}) \quad (2)$$

三维正则化网络 R 基于一个轻量级的三维 UNet,包含 3 个下采样块和 3 个上采样块。每个下采样块由两个 $3 \times 3 \times 3$ 的三维卷积组成。受 CoEx^[6] 工作的启发,使用从左特征 F_l 计算的权重来激发成本体积通道,以进行成本聚合。在三维正则化网络中,引导成本体积激励操作有效地推断和传播场景几何信息,从而得到几何编码体积,如图 3 所示。此外,计算相应的左、右特征之间的所有成对相关性,以获取全局特征相关性。

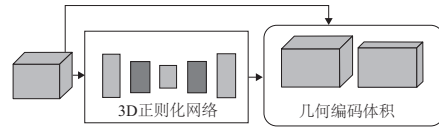


图3 GEV 模块

为了从几何编码体积 C_G 中回归出初始视差 d_0 ,采用 soft argmin 方法。视差 d_0 通过对几何编码体积 C_G 上的预设视差索引进行加权求和得到,使用 softmax 函数进行加权。进一步,使用 SA 模块更新视差 d_0 。该模块通过测量特征的自相似性,将高质量的预测传播到不匹配区域。这种方法增强了对遮挡、无纹理区域和弱纹理区域的鲁棒性。到目前为止,本文方法在保持简单高效的同时取得了竞争性的性能。然而,通过引入额外的细化步骤,可以进一步改进方法,以便在速度和准确性之间取得不同的平衡。

2 实验

2.1 数据集

为了提高对无人机输电线检测模型的有效

性,制作了一个仿真输电线数据集。使用 Unreal Engine 场景和 Airsim 工具生成了合成训练数据。每个场景都包含了虚拟无人机的左右图像对以及像素级精确的密集视差图。数据生成过程注重模拟真实场景中的挑战情况,并提供了多种增强功能。

为了增加输电线数据的内容多样性,在研究中采用了多个场景,包括 Urban City 和 Old Town,如图 4 所示。将 Airsim 的基本形状与 Unreal Engine 的线框修改器相结合,以创建复杂的光照环境。在场景中随机放置了不同类型的灯光,并使用真实世界图像作为物体和场景背景的纹理。此外,利用 Unreal Engine 中 Cycles 渲染器的光线追踪功能。为了验证立体匹配模型在输电线识别和泛化能力方面的有效性,在以下公开数据集上进行了比较实验:Scene Flow^[7]、Middlebury 2014^[8]、ETH3D^[9]、KITTI 2012^[10]和 KITTI 2015^[11]。

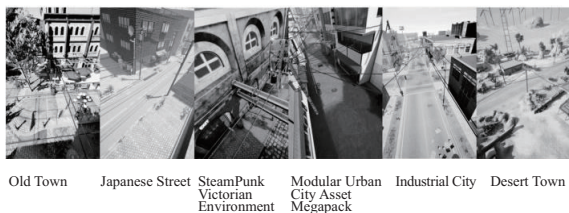


图 4 仿真输电线数据集

2.2 训练细节

采用了 PyTorch 框架来实现立体匹配网络,并在 NVIDIA RTX 3090 GPU 上执行了一系列实验。在训练过程中,选用 AdamW 优化器,并采用将梯度截断至范围 $[-1, 1]$ 的梯度裁剪方法。为了训练网络,在 Scene Flow 数据集上进行了长达 200 000 步的训练,使用的批处理大小为 8。在 KITTI 数据集上,对预先训练的 Scene Flow 模型进行微调,利用混合 KITTI 2012 和 KITTI 2015 的训练图像进行了 50 000 步的微调。为了增强数据多样性,采用随机裁剪图像大小至 320×736 ,并引入数据增强技术。实验中,设定索引半径为 4。为了优化训练过程,采用单周期学习率计划,初始学习率为 0.000 2,并在训练过程中进行了 22 次更新迭代。

1) 输电线数据集实验

根据实验参数设置,选择了一系列主流的立体匹配模型,并在输电线数据集上进行了对比实验,实验结果如表 1 所示。相较于最先进的 CREStereo 方法,本文设计的模型表现出较强的竞

争力,并且运行速度快了约 56%。与之前的轻量级立体模型 AANet 相比,本文设计的方法展现了更为卓越的性能。实验结果不仅显示了本文设计方法的竞争力,同时也揭示了其在速度和精度方面的显著优势。这表明,将立体匹配模型作为避障系统的图像识别模型能够实现最佳效果,从而在实时预警无人机飞行路径方面提供更为准确的保障,并最大程度地提升避障系统的效能。

表 1 输电线 Airsim 数据集不同模型性能评估

模型	D1	RMSE	运行时间/s
PSMNet ^[12]	1.89	2.22	0.38
GwcNet	1.14	2.00	0.36
AANet ^[13]	1.54	1.82	0.24
HITNet ^[14]	1.63	2.18	0.02
RAFT-Stereo ^[15]	1.12	1.95	0.35
CREStereo ^[16]	<u>0.98</u>	<u>1.71</u>	0.41
Ours	0.66	0.98	<u>0.18</u>

在对无人机飞行过程中随机选择的图像进行预警模型的测试中,验证避障模型对实际障碍物的推理性能,测试结果如图 5 所示。预警模型在输电线场景的推理性能维持在高水平,能够准确判断无人机在飞行中是否面临障碍物。这为操纵者提供了重要信息,使其能够及时注意旋翼是否存在碰撞危险,从而达到提前预警的目的。图 5 中的推理效果图生动展示了预警模型的实际应用效果。其能够有效地识别输电线场景中的障碍物,为无人机的安全飞行提供了关键的决策支持。

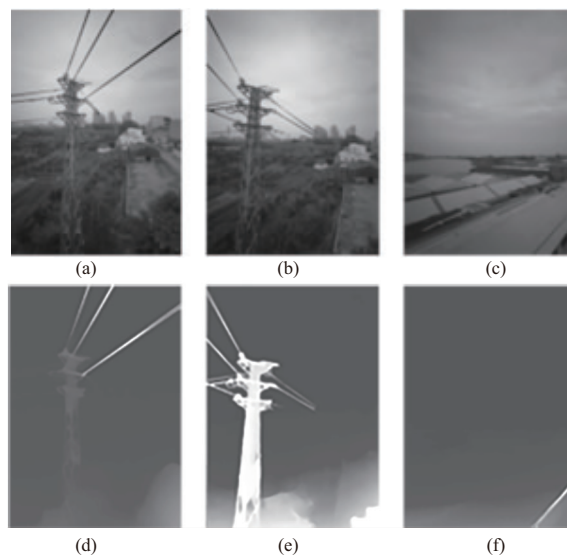


图 5 图片推理效果

这一结果不仅证明了本文设计方法的实用性,也强调了其在无人机避障系统中的作用。通过这一系列测试,本文设计的避障模型在实际应用中表现出色,为提高无人机操控者对潜在风险的感知能力提供了有效手段。这将有助于确保无人机的飞行安全,为相关领域的研究和应用提供了有力的支持。

2) 公开数据集实验

为了验证立体匹配模型在公开数据集上的效果,本文在不同的数据集上进行了验证。在 KITTI 2012 数据集上,相较于 GwcNet 和 AANet,本文设计的模型在 2 像素误差阈值下的 Out-Noc 指标分别提高了 19.19% 和 6.81%。在 KITTI 2015 数据集上,本文模型在 D1-all 指标上分别比 HITNet 和

ACVNet 高出 18.69% 和 2.42%。在 Middlebury 测试集上,本文设计的模型相较于 RAFTStereo 提高了 23% 的准确性,与经典的 PSMNet 相比,不仅准确性提高了 76.90%,而且速度更快。具体的定量比较结果如表 2 所示。值得注意的是,在与其他迭代方法如 CREStereo 的比较中,本文设计的模型不仅在性能上表现更出色,而且速度提高了 70.73%,在反射和细节区域表现极佳。在表 2 中对本文设计的模型和 RAFT-Stereo 在不适宜区域的性能进行了评估。由于 RAFT-Stereo 缺乏非局部知识,因此在处理不适宜区域的局部歧义时存在困难,而相比之下,本文设计的模型能够有效克服上述这些问题。

表 2 不同算法在公开数据集上的评估结果

模型	KITTI 2012			KITTI 2015			Middlebury	ETH3D	运行时间/s
	2-all	3-all	EPE-all	D1-bg	D1-fg	D1-all	Bad 2.0	Bad 1.0	
PSMNet	3.01	1.89	0.6	1.86	4.62	2.32	15.8	10.2	0.41
GwcNet	2.71	1.70	0.5	1.74	3.93	2.11	14.1	9.1	0.32
AANet	2.35	1.54	0.5	1.51	3.80	1.89	13.6	7.9	0.48
HITNet	2.65	1.89	0.5	1.74	3.20	1.98	13.9	8.9	0.02
ACVNet ^[17]	2.35	<u>1.47</u>	0.5	<u>1.37</u>	3.07	1.65	13.6	7.8	0.20
RAFT-Stereo	2.42	1.66	0.5	1.58	3.05	1.82	4.74	3.2	0.38
CREStereo	2.18	1.46	0.5	1.45	<u>2.86</u>	<u>1.69</u>	<u>3.71</u>	<u>2.8</u>	0.41
Ours	<u>2.19</u>	1.46	0.5	1.35	2.80	1.61	3.65	2.4	<u>0.12</u>

3 结语

本文提出了一种基于 Transformer 和几何编码的立体匹配算法,专注于解决无人机输电线深度估计问题。通过引入交叉注意机制和自注意层,有效地应对了无纹理区域和不匹配区域的挑战。同时,利用几何编码将成本体积编码变成几何编码体积,以实现输电线路目标的精确估计,从而协助无人机在复杂环境中实现避障。实验结果充分证明了本文设计的模型在输电线深度估计方面的卓越性能。与此同时,本文所提出的算法具有更为简洁和高效的模型设计与推理过程,为实际应用提供了更为可行的解决方案。本文设计的算法在无人机和自动驾驶等领域具有广泛的应用前景。其高效性和准确性使其成为一种理想的输电线深度估计方法。未来的工作将着重于进一步优化算法性能,探索更多实际应用场景,并深入研

究如何在复杂环境中提高模型鲁棒性的方法。

参考文献:

- [1] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [C]//Advances in Neural Information Processing Systems. [S. l.: s. n.], 2017: 5998-6008.
- [2] GUO X Y, YANG K, YANG W K, et al. Group-wise correlation stereo network [C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, CA, USA: IEEE, 2019: 3268-3277.
- [3] 郑伟,杨晓辉,吕中宾,等. 基于改进 YOLOv4 输电线关键部件实时检测方法[J]. 科学技术与工程, 2021, 21(24): 10393-10400.
- [4] 林刚,王波,彭辉,等. 基于改进 Faster-RCNN 的输电线巡检图像多目标检测及定位[J]. 电力自动化设备, 2019, 39(5): 213-218.

(下转第 272 页)

- 收稿日期:2024-06-24

- [12] CHANG J R, CHEN Y S. Pyramid stereo matching network [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA: IEEE, 2018: 5410-5418.
- [13] XU H F, ZHANG J Y. AANet: adaptive aggregation network for efficient stereo matching [C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, WA, USA: IEEE, 2020: 1959-1968.
- [14] TANKOVICH V, HÄNE C, ZHANG Y D, et al. HITNet: hierarchical iterative tile refinement network for real-time stereo matching [C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville, TN, USA: IEEE, 2021: 14357-14367.
- [15] LIPSON L, TEED Z, DENG J. RAFT-stereo: multilevel recurrent field transforms for stereo matching [C]//2021 International Conference on 3D Vision (3DV). London, United Kingdom: IEEE, 2021: 218-227.
- [16] LI J K, WANG P S, XIONG P F, et al. Practical stereo matching via cascaded recurrent network with adaptive correlation [C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans, LA, USA: IEEE, 2022: 16242-16251.
- [17] XU G W, CHENG J D, GUO P, et al. Attention concatenation volume for accurate and efficient stereo matching [C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans, LA, USA: IEEE, 2022: 12981-12990.

收稿日期:2023-12-08