

基于光谱空间融合的高光谱目标检测

摘要

基于深度学习的高光谱图像 (HSI) 分类和目标检测技术因其在图像内容分析、解译以及更广泛的高光谱应用中的重要作用, 已引起广泛关注。然而, 当前的高光谱目标检测方法主要侧重于光谱或空间信息, 往往忽视了这两方面信息之间宝贵的互补关系。在本研究中, 提出了一种新颖的光谱-空间聚合 (S2ADet) 目标检测器, 该检测器能够有效利用高光谱图像中丰富的光谱和空间互补信息。S2ADet 包括一个高光谱信息解耦 (HID) 模块、一种双流特征提取网络以及一个单阶段检测头。HID 模块通过波段选择和主成分分析聚合光谱和空间信息, 从而处理高光谱数据并减少冗余。在获取的光谱和空间聚合信息的基础上, 提出了一种特征聚合双流网络, 用于交互光谱-空间特征。此外, 为了解决现有数据库的局限性, 标注了一个名为 HOD3K 的大规模数据集, 该数据集包含 3,242 幅高光谱图像, 这些图像拍摄于多种真实场景, 涵盖了三个目标类别。这些图像的分辨率为 512×256 像素, 覆盖了 470 nm 至 620 nm 的 16 个波段。在该数据集上进行训练, 取得了稳健且可靠的结果。

关键词: 高光谱图像目标检测; 深度学习; 特征融合; 光谱-空间聚合

1 引言

目标检测是一项关键任务, 旨在识别图像中属于特定类别的目标。它在多个领域中具有广泛的应用, 包括遥感、自动驾驶和医学分析。在过去的几十年中, 为提升 RGB 图像的目标检测性能, 人们付出了显著努力。然而, 与 RGB 数据相比, 高光谱数据具有多方面的优势, 它能够捕捉物体材料的更多内在属性, 从而实现更细粒度的目标检测。因此, 高光谱检测已在学术界引起了广泛关注。

高光谱检测在遥感应用中具有举足轻重的作用, 通过光谱特征识别特定材料或目标。Rui 等人、Chang 等人和 Dong 等人 [5] 的最新研究主要依赖于像素级光谱信息或相邻像素之间的相关信息。这些方法可以分为两大类: 统计分布模型和子空间模型。根据 Shang 等人和 Chang 等人 [3] 的描述, 统计分布模型假设图像背景符合多变量正态分布。这些模型通过将像素与背景分布的中心进行比较, 检测异常像素, 从而定位可能的目标所在区域。另一方面, 子空间模型通常涉及设计一个线性滤波器, 在满足约束方程的同时最小化输出能量。这种方法能够识别包含感兴趣信号的子空间, 从而实现精确的高光谱目标检测。

然而, 目前大多数高光谱检测方法主要依靠特定图像波段的光谱信息来识别材料, 忽略了空间维度中包含的关键语义信息。相比之下, 现有的基于深度学习的高光谱图像目标检测方法侧重于空间信息, 没有考虑丰富的可用光谱语义信息。此外, 相邻波段中的高光谱图像

高度相似，并且相邻波段之间包含大量冗余信息，简单地将它们输入特征提取器会显著影响检测性能。为了应对这一挑战，提出了 S2ADet [7]，这是一种新颖的物体检测器，采用双流光谱空间特征聚合方法来检测高光谱数据中的物体。S2ADet 利用互补的空间和语义信息来学习更好的对象语义特征。

2 相关工作

目标检测旨在对检测到的目标执行边界回归和类别分类任务，并已广泛应用于自动驾驶和异常检测等领域。根据检测头中涉及的阶段数，目标检测方法可分为两阶段方法和单阶段方法。自 Girshick 等人提出 RCNN 方法以来 [6]，两阶段目标检测方法迅速发展，并基于 Faster RCNN [4] 的一系列检测器相继被提出，取得了显著的成果。尽管两阶段检测器的精度高于单阶段检测器，但检测过程中的候选区域生成步骤降低了目标检测的速度。为了满足需要高检测速度的场景应用，一些高效的单阶段目标检测器应运而生，并在工业场景中得到广泛使用。近年来，无锚框方法引起了广泛关注，提出了用锚点代替锚框进行目标检测。为了实现端到端目标检测，提出了 DETR [2]，该方法通过可学习的 Query 实现目标检测。

2.1 HOD3K 数据集

现存的目标检测数据集存在着数据量少且数据类型分布不均匀的问题，为了解决这一问题，提出了一个新的高光谱目标检测数据集，称为 HOD3K，并配备了可用于实际应用的注释，以满足各种实际应用的需求。图 1 显示了 HOD3K 数据集和现有的代表性数据集 (San Diego 和 HOD-1) 的示例图像，其包含的图像和注释数量是目前最主要的高光谱数据集 (HOD-1) [9] 的十倍，同时包含 3242 张自然场景的高光谱图像。每张高光谱图像包含 512×256 像素，具有 16 个波段，覆盖 470nm 到 620nm 的光谱范围。包含三种类别（人、车和自行车），共计 15,149 个物体，平均每张图像包含 4.67 个物体。具体而言，数据集包括 12,144 个人物体，817 辆车和 2,188 辆自行车。下表显示了每个类别标签数量的分布。HOD3K 数据集专为高光谱图像中的目标检测设计，使其成为研究大规模场景检测的宝贵资源。表 1 为 HOD3K 数据集数据分布。

表 1. HOD3K 数据集数据分布.

Dataset	Annotation	People	Car	Bike	All
HOD3K	Number	12,144	817	2,188	15,149
	Ratio (%)	80.2	14.4	5.4	100

2.2 基于空间-光谱信息融合的高光谱目标检测

现有的高光谱目标检测方法主要依赖于像素级的光谱信息或相邻像素之间的简单相关信息。为了利用深度学习挖掘高光谱图像中的深层信息，Yu 等人提出了一个基于 CNN 的目标检测器，该检测器通过 3D 卷积层将高光谱图像整合到现有的目标检测模型中。该方法忽略了光谱和空间维度信息中丰富的上下文语义互补信息。为了解决这些局限性，本文提出了一

种高光谱信息解耦（HID）模块，从光谱和空间的角度对高光谱数据进行解耦，并设计了一种双流网络用于光谱和空间维度信息的交互。因此，我们旨在聚合空间维度中丰富的上下文语义信息与光谱维度的信息。

3 本文方法

3.1 本文方法概述

图 1 描绘了该新型目标检测器 S2ADet 的整体架构，该检测器利用高光谱图像中的丰富光谱和空间信息。为此，我们采用了特征金字塔网络（FPN）和受单阶段目标检测器启发的检测头。为了简洁起见，图中省略了 FPN 和检测头的展示。S2ADet 的结构包含高光谱信息解耦（HID）模块、双流网络、光谱-空间聚合（SSA）模块和单阶段检测头。HID 用于去除冗余信息。还设计了一个嵌入的双流特征提取网络（SSA 模块），用于交互高光谱图像的光谱和空间维度特征。

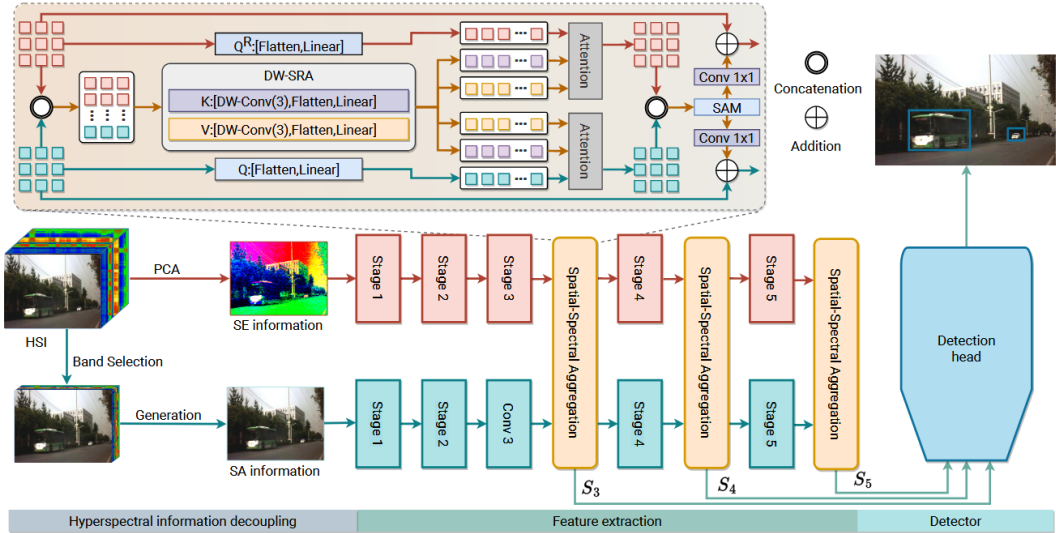


图 1. S2ADet 网络结构

3.2 高光谱信息解耦（HID）模块

高光谱成像仪在捕获光谱波段时表现出多样性，这导致了高光谱图像中信息的高度冗余。因此，直接使用探测器（ H ）处理高光谱图像会限制探测器的性能。为了解决这一问题，我们从光谱和空间的角度对高光谱图像进行解耦，并通过一个高光谱信息解耦（HID）模块对原始数据进行处理，如图 1 所示。该过程可以表示如下：

$$a = \text{SpatialFilter}(H), \quad e = \text{SpectralFilter}(H), \quad (1)$$

空间聚合信息 高光谱成像的特点是相邻光谱波段之间具有高度相关性，导致显著的空间数据冗余。为减少高光谱波段之间的冗余，我们引入了一种无参数的方法，即最优邻域重构（ONR）[8]，用于从高光谱数据中选择最具信息量的光谱波段。该波段选择步骤生成了一个保留最相关光谱信息的空间维度聚合图像。此外，为了进一步提炼空间信息，我们采用了一种称为色彩映射（color mapping）的技术来生成空间聚合信息（ a ）。

光谱聚合信息 高光谱图像包含的波段数量是自然图像的十倍以上，这显著降低了目标检测的速度。为了解决高光谱图像中光谱数据冗余带来的这一挑战，我们采用主成分分析 (PCA) [1]，以光谱维度为基础对这些图像进行压缩。通过降维，我们获得了光谱聚合信息 (e)。

3.3 光谱-空间聚合模块

在构建高光谱信息解耦模块后，可以获得 SE 和 SA 信息。为了在光谱和空间两个维度上提取交互图像的特征，我们设计了一个双流网络。如图 1 顶部所示，该网络对 SE 和 SA 信息进行特征提取，并结合注意力操作，构成了光谱-空间聚合 (SSA) 模块。

输入 $a^i \in \mathbb{R}^{n \times d}$ 是来自 SA 信息的特征输入，而 $e^i \in \mathbb{R}^{n \times d}$ 表示 SE 信息。其中， n 表示 patch 的数量，等于图像高度 (h) 和宽度 (w) 的乘积，而 d 表示特征的维度。首先，为了实现光谱-空间图像特征的上下文信息聚合，我们将两个图像特征连接起来，如下所示：

$$f^i = \text{Concat}(a^i, e^i). \quad (2)$$

然后，将 f^i 分别作为 $f^i W_K \in \mathbb{R}^{n \times d \times 2}$ 和 $f^i W_V \in \mathbb{R}^{n \times d \times 2}$ ，其中 a^i 为 $a^i W_Q \in \mathbb{R}^{n \times d_v}$ ， e^i 为 $e^i W_Q \in \mathbb{R}^{n \times d_v}$ ，输入至分割注意力模块。

由于通过高光谱特征计算注意力直接会带来较高的计算开销，我们采用空间降维注意力来减小 key 和 value 的维度。公式如下：

$$f^i \tilde{W}_K = \text{DWConv}(f^i W_K) \in \mathbb{R}^{r^2 \times d \times 2}, \quad (3)$$

$$f^i \tilde{W}_V = \text{DWConv}(f^i W_V) \in \mathbb{R}^{r^2 \times d \times 2}, \quad (4)$$

其中， r 是空间降维率。然后，将 $a^i W_Q$ 、 $e^i W_Q$ 、 $f^i \tilde{W}_K$ 和 $f^i \tilde{W}_V$ 分别输入注意力块单独计算：

$$\text{Attn}_1(a^i W_Q, f^i \tilde{W}_K, f^i \tilde{W}_V) = \text{Softmax}\left(\frac{\text{att}(a^i, f^i)}{\sqrt{d_k}}\right) f^i \tilde{W}_V, \quad (5)$$

$$\text{Attn}_2(e^i W_Q, f^i \tilde{W}_K, f^i \tilde{W}_V) = \text{Softmax}\left(\frac{\text{att}(e^i, f^i)}{\sqrt{d_k}}\right) f^i \tilde{W}_V, \quad (6)$$

其中

$$\text{att}(x, y) = (x W_Q)(y W_K)^T. \quad (7)$$

然后通过 FFN [1] 层添加非线性变换。变换器后的整体结构表达如下：

$$\begin{bmatrix} \tilde{a}^i \\ \tilde{e}^i \end{bmatrix} = \begin{bmatrix} \tilde{a}^i \\ \tilde{e}^i \end{bmatrix} + \text{Att}\left(\begin{bmatrix} a^i \\ e^i \end{bmatrix}\right) + \text{FFN}\left(\begin{bmatrix} \tilde{a}^i \\ \tilde{e}^i \end{bmatrix}\right). \quad (8)$$

我们通过结合变换器的全局和动态建模能力，将高光谱信息进行耦合处理。因此，该过程最终在光谱和空间维度上聚合了上下文信息。

在高光谱目标检测任务中，多任务损失函数在训练期间定义如下：

$$L = L_{\text{cls}} + L_{\text{box}}, \quad (9)$$

其中， L_{cls} 表示分类损失， L_{box} 表示边界框损失。该损失函数将边界框预测和类别预测解耦，与应用于目标检测的常规方法有所不同。先前工作的若干实验结果表明，该损失函数是获得良好目标检测结果的关键，适用于高光谱目标检测。

3.4 基于 Transformer 的光谱-空间聚合模块

在保证光谱和空间特征交互的前提下,进一步增强对光谱信息的提取,对原模型的 SSA 模块进行改进,将其注意力机制替换为两个并行的 Transformer 提取器,一个用于在通道维度进行提取特征,另一个用于在空间维度提取特征,但为保证两个维度信息的交互,在 Transformer 间引入交互式模态感知模块 (IMPM) 以提取和整合不同模态间的互补信息。具体结构如图 2

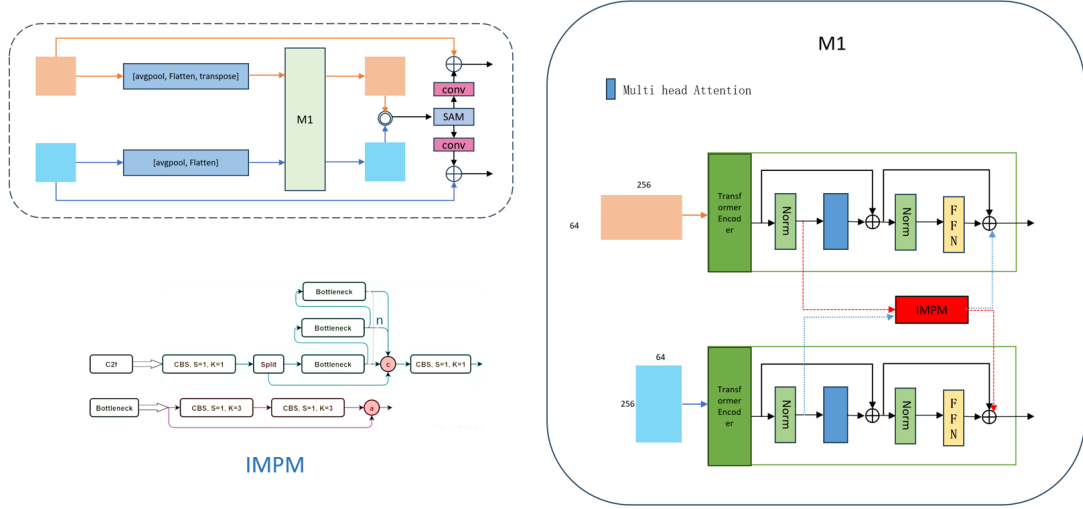


图 2. SSA 模块结构

4 复现细节

4.1 与已有开源代码对比

在为引入改进的 SSA 模块前,利用开源代码,在 HOD3K 上进行训练,训练轮次 100 轮, batch size 设置为 16,得到的结果如表 2 将引入了 Transformer 引入并改进 SSA 模块后,以

表 2. HOD3k 数据集上的性能

Class	Images	Labels	P	R	mAP@0.5	mAP@0.75	mAP
all	715	2846	0.828	0.773	0.841	0.562	0.509
people	715	2140	0.923	0.623	0.814	0.456	0.447
bike	715	63	0.599	0.857	0.780	0.472	0.426
car	715	643	0.963	0.839	0.929	0.757	0.654

同样的设置进行训练得到结果如下表 3,对比可知,引入了新的 SSA 模块后,其 Map 得到了一定的提升。这一段描述实验设置的文字。我们进行了相同设置下的训练,并记录了各类性能指标。

表 3. 以同样的设置进行训练性能

Class	Images	Labels	P	R	mAP@0.5	mAP@0.75	mAP@0.5:0.95
all	715	2846	0.891	0.749	0.856	0.549	0.510
people	715	2140	0.945	0.590	0.827	0.435	0.435
bike	715	63	0.768	0.841	0.826	0.453	0.447
car	715	643	0.960	0.816	0.916	0.758	0.648

4.2 创新点

- 设计了 HID 模块以解耦高光谱图像，从而减少冗余并聚合光谱和空间信息。
- 设计了一个双流网络，并通过 SSA 模块促进光谱和空间信息的聚合，从而提取精细特征。
- 设计了新型的 SSA 模块，进一步提高了模型精度。

5 实验结果分析

实验结果如图 3所示

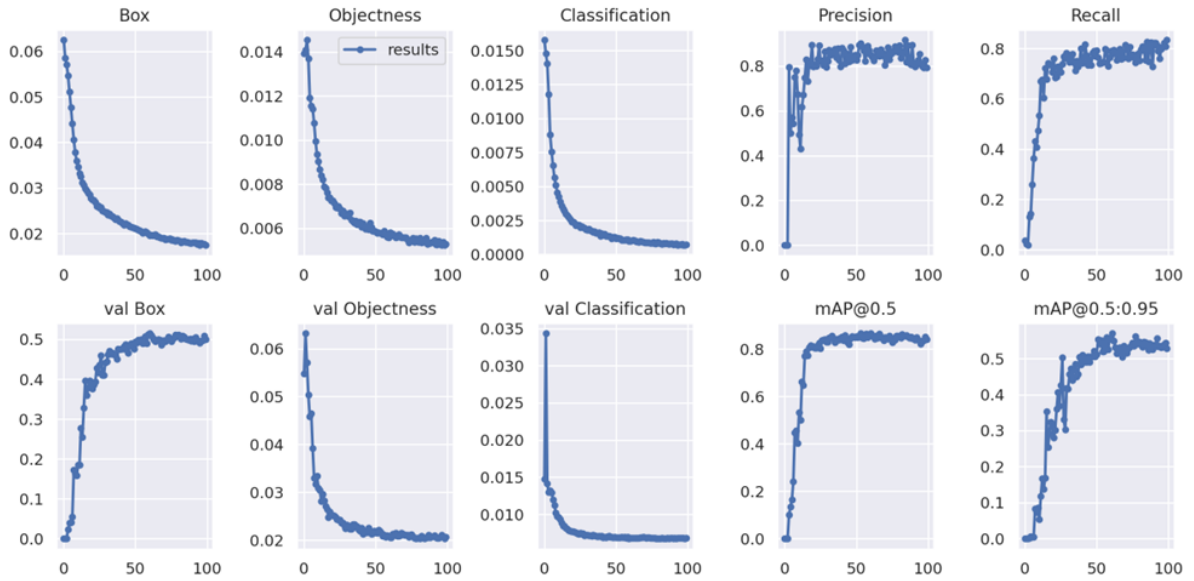


图 3. 实验结果示意

在测试集上的目标检测结果如图 4，该图显示了其中 16 张图片上的预测结果，其可以检测到大部分目标。但对于一些重合度较高的目标可能难以进行检测。如图 5



图 4. 测试集上的目标检测结果



图 5. 效果分析

6 总结与展望

本研究提出了一种新颖的基于光谱-空间聚合的高光谱图像目标检测方法 (S2ADet)，通过设计高光谱信息解耦 (HID) 模块和双流特征提取网络，有效利用了高光谱数据中的光谱与空间互补信息。在实验中，我们对比了不同模型的性能，结果显示，S2ADet 在多个指标上均优于已有方法，尤其是在 mAP@0.5 上获得了显著提升。引入的 SSA 模块通过改进注意力机制进一步提高了模型的检测精度。然而，本研究仍存在以下不足之处：

1. 数据集覆盖范围有限，仅包含特定场景和有限类别的高光谱目标，可能限制模型的泛化能力。
2. 在重合目标和小目标的检测任务上仍有改进空间，这可能需要引入更先进的特征增强机制。
3. 模型的复杂度较高，可能导致推理速度在实时场景应用中受到限制。

未来的研究方向可以集中于：扩展数据集规模与类别：采集更多样化的高光谱图像数据，涵盖更广泛的场景和目标类型，以提高模型的适用性和泛化能力。改进目标检测性能：设计更精细的网络结构以应对小目标和重叠目标的检测问题，同时探索多模态数据融合方法。提升推理速度：优化模型的计算效率，通过轻量化设计或知识蒸馏技术实现高效推理。

通过不断优化与改进，S2ADet 有望为高光谱图像目标检测的理论与实际应用提供新的解决方案，为相关领域的发展做出贡献。

参考文献

- [1] Hervé Abdi and Lynne J Williams. Principal component analysis. *Wiley interdisciplinary reviews: computational statistics*, 2(4):433–459, 2010.
- [2] Nicolas Carion, Francisco Massa, Gabriel Synnaeve, Nicolas Usunier, Alexander Kirillov, and Sergey Zagoruyko. End-to-end object detection with transformers. In *European conference on computer vision*, pages 213–229. Springer, 2020.
- [3] Chein-I Chang. Hyperspectral anomaly detection: A dual theory of hyperspectral target detection. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–20, 2021.
- [4] Xinlei Chen and Abhinav Gupta. An implementation of faster rcnn with study for region sampling. *arXiv preprint arXiv:1702.02138*, 2017.
- [5] Yanni Dong, Wenzhong Shi, Bo Du, Xiangyun Hu, and Liangpei Zhang. Asymmetric weighted logistic metric learning for hyperspectral target detection. *IEEE Transactions on Cybernetics*, 52(10):11093–11106, 2021.
- [6] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 580–587, 2014.
- [7] Xiao He, Chang Tang, Xinwang Liu, Wei Zhang, Kun Sun, and Jiangfeng Xu. Object detection in hyperspectral image via unified spectral-spatial feature aggregation. *IEEE Transactions on Geoscience and Remote Sensing*, 2023.
- [8] Qi Wang, Fahong Zhang, and Xuelong Li. Hyperspectral band selection via optimal neighborhood reconstruction. *IEEE Transactions on Geoscience and Remote Sensing*, 58(12):8465–8476, 2020.

- [9] Longbin Yan, Min Zhao, Xiuheng Wang, Yuge Zhang, and Jie Chen. Object detection in hyperspectral images. *IEEE Signal Processing Letters*, 28:508–512, 2021.