doi:10.6041/j.issn.1000-1298.2025.01.001

农业领域多模态融合技术方法与应用研究进展

李道亮[1,2]　赵　晔[1,2]　杜壮壮[1,2]

(1. 中国农业大学信息与电气工程学院, 北京 100083; 2. 国家数字渔业创新中心, 北京 100083)

摘要: 多模态融合技术通过结合多源数据,可以克服单一模态的局限性。近年来,传感器以及遥感技术的发展为作物监测提供了更加丰富的数据源,光谱数据、图像数据、雷达数据以及热红外数据被广泛应用于作物监测中。通过利用计算机视觉技术以及数据分析方法,可以从中获取作物的表型参数、理化特征等信息,从而有助于评估作物的生长状况、指导农业生产管理。现有研究多数是基于单一模态数据展开,而单一模态的数据仅有一种类型的输入,缺乏对整体信息的理解,且容易受到单模态噪声的影响;部分研究虽然采用了多模态融合技术,但仍未能充分考虑模态间的复杂交互关系。为了深入分析多模态融合技术在农业领域应用的潜力,本文首先阐述了农业领域中多模态融合的先进技术与方法,重点梳理了多模态融合技术在作物识别、性状分析、产量预测、胁迫分析及病虫害诊断领域中的应用研究成果,分析了多模态融合技术在农业领域中存在的数据利用程度低、有效特征提取难、融合方式单一等问题,并对未来发展提出展望,以期通过多模态融合的方法推动农业精准管理、提高生产效率。

关键词: 多模态融合; 传感器; 遥感技术; 作物监测; 计算机视觉; 农业精准管理

中图分类号: S126　　**文献标识码**: A　　**文章编号**: 1000-1298(2025)01-0001-15　　**OSID**:

Advances in Multi-modal Fusion Techniques and Applications in Agricultural Field

LI Daoliang[1,2]　ZHAO Ye[1,2]　DU Zhuangzhuang[1,2]

(1. *College of Information and Electrical Engineering*, *China Agricultural University*, *Beijing* 100083, *China*
2. *National Innovation Center for Digital Fisheries*, *Beijing* 100083, *China*)

Abstract: Multi-modal fusion technology, by combining data from multiple sources, has been widely applied in fields such as medicine, autonomous driving, and emotion recognition to overcome the limitations of a single modality. In recent years, advancements in sensor and remote sensing technologies have provided richer data sources for crop monitoring, including spectral data, image data, radar data, and thermal infrared data. By utilizing computer vision and data analysis methods, information such as phenotypic parameters and physicochemical characteristics of crops can be obtained, helping to assess crop growth and guide agricultural production management. Most existing studies were based on single-modal data, which involved only one type of input and lacked an understanding of the overall information, making them susceptible to noise from a single modality. Although some studies employed multi-modal fusion technology, they still did not fully consider the complex interactions between modalities. To thoroughly analyze the potential of multi-modal fusion technology in crop monitoring, the advanced technologies and methods of multi-modal fusion in the agricultural field were firstly outlined, with a focus on its application in crop identification, trait analysis, yield prediction, stress analysis, and pest and disease diagnosis. The existing challenges were also discussed and an outlook on future developments was provided, aiming to promote precision agriculture management and improve production efficiency through multi-modal fusion methods.

Key words: multi-modal fusion; sensors; remote sensing technology; crop monitoring; computer vision; precision agriculture management

收稿日期: 2024-08-09　修回日期: 2024-09-01

基金项目: 国家自然科学基金项目(32373186)

作者简介: 李道亮(1971—),男,教授,博士生导师,主要从事农业智能信息处理与农业农村信息化研究,E-mail: dliang@cau.edu.cn

0 引言

作物的生长受到气象环境、栽培方法以及病虫害等多重因素的影响。通过定期监测可以掌握作物的生长状况,并能根据实际情况及时调整种植方案[1]。目前,智能化的监测方式已经取代了传统的人工监测,不仅节约了人力成本,提高了监测效率,也有助于实现更加精确的农业管理。然而,大部分研究基于单一模态数据,往往无法全面反映作物生长的复杂情况和潜在问题,从而影响决策的有效性。多源数据包括:遥感图像、环境传感器数据、气象数据、光谱数据等。多模态融合技术通过整合多源数据,能够提供更为精确的分析,从而改善监测的效果。

在作物监测中,基于成像技术获取的可见光图像数据是主要数据源[2]。其中,RGB 图像分辨率高,可提取作物纹理、颜色、形状特征,数据处理相对简单[3]。光谱图像可推导出与光合效率、叶绿素含量、植物胁迫等相关的指数,检测到更深层次的信息[4-5]。热成像技术可以获取作物叶片和冠层的温度,反映作物生理特性,在植物病害早期检测中具有潜力[6]。尽管基于单一视觉图像的监测模型在特定情况下表现出较高的准确性,但由于只关注一种数据模态,无法有效利用数据间的互补特性。其他跨模态信息如文本描述、气象特征、环境因素等未得到有效利用。此外,外部环境干扰、叶片果实重叠、光谱不可分等也影响单一图像评估结果[7-8]。因此,通过多模态融合技术整合不同数据源的信息,可为作物监测提供高可靠、高精度的解决方案。

目前,多模态融合技术在临床决策、自动驾驶、情感分析等领域已取得大量研究成果[9-10]。在农业领域,多模态融合技术也逐渐兴起,在灾害评估、作物识别等应用中展现巨大潜力[11-12]。然而,尽管多模态融合技术具备诸多优势,但这种组合信息的方法也会增加模型的复杂性。此外,融合模型的准确性也受到输入数据质量影响。如何有效提取模态数据特征、实现数据对齐及挖掘模态间相关性与互补性,仍需进一步研究与探讨。因此,本文总结并梳理多模态融合技术的发展及在作物监测领域中应用的相关文献,探讨现有技术不足并对未来发展提出展望,以期为研究人员提供参考。

1 农业领域多模态融合方法

多模态融合目标是将不同分布、来源和类型的数据或特征整合到一个统一的空间中,从而从各个数据模态中学习互补信息[13]。在这个空间中,多模态和跨模态信息都可以使用统一的方式表示,整个过程分为数据采集、特征提取以及多模态融合几个阶段,融合方法主要包括输入级融合、特征级融合、决策级融合以及混合级融合(图 1)。

在农业领域中,大多有关多模态融合的应用都是基于输入级融合、特征级融合以及决策级融合进行的,这些基于拼接式或手动选择权重的方式不能充分利用模态互补性,导致融合的效果不佳。从最近的文献中可以看到,基于混合融合的方法能够在一定程度上克服单一融合方法的不足[14]。此外,动态融合[15]、自适应逐级融合[16]等特征级融合方法的变体也能进一步提升融合效果,为农业领域的智能化和精准化发展提供有力支持。因此,主要介绍农业领域中几种先进的多模态融合方法,并对每种方法进行分析。

1.1 基于线性的融合

基于线性的融合方法可以用于同质或异质的数据中,通过计算两个模态特征之间的双线性交互,捕捉模态间更丰富的关系。YANG 等[17]在柑橘黄龙病的检测中,比较了特征加法、乘法和双线性融合方法,结果证明,与简单的特征加法或乘法方法相比,双线性融合能够更全面地捕获模态间的互补性,因此准确性最高。在跨模态数据的融合中,REN 等[18]基于线性方法将近红外光谱、电子眼、电子舌等多传感器数据特征进行融合,有效捕捉了跨模态间的互补关系。

相较于单模态信息,线性融合方法可以获取更加全面的特征表示[19]。然而这种方法无法充分利用两种信息在不同层次上的互补性,有研究提出利用半张量积多线性池化的方法将多模态信息映射到更紧凑的张量空间[20],从而灵活地处理模态间的复杂交互关系,解决了矩阵乘法中维度一致性的限制。

1.2 基于多流分支的融合

目前多模态特征融合架构主要分为单流和双流两种类型,与单流架构相比,双流架构更能充分利用模态间的互补特性,避免单流架构对多模态数据特征进行混合处理时可能出现的信息丢失或模态冲突问题[21-23]。在农业遥感领域,作物的几何特征、光谱特性都会受到时间的影响,因此可以通过双流架构提取光谱、时间和空间特征,提升监测的实时性与准确性[24]。CAI 等[25]构建了一种基于图注意力机制的三分支融合模型,用于高光谱(Hyperspectral imagery,HSI)和激光雷达(LightLaser detection and ranging,LiDAR)联合数据分类。融合阶段通过构建多源特征图来解决高光谱数据的长距离依赖,利用注意力机制聚合每个节点信息,虽然在一定程度上

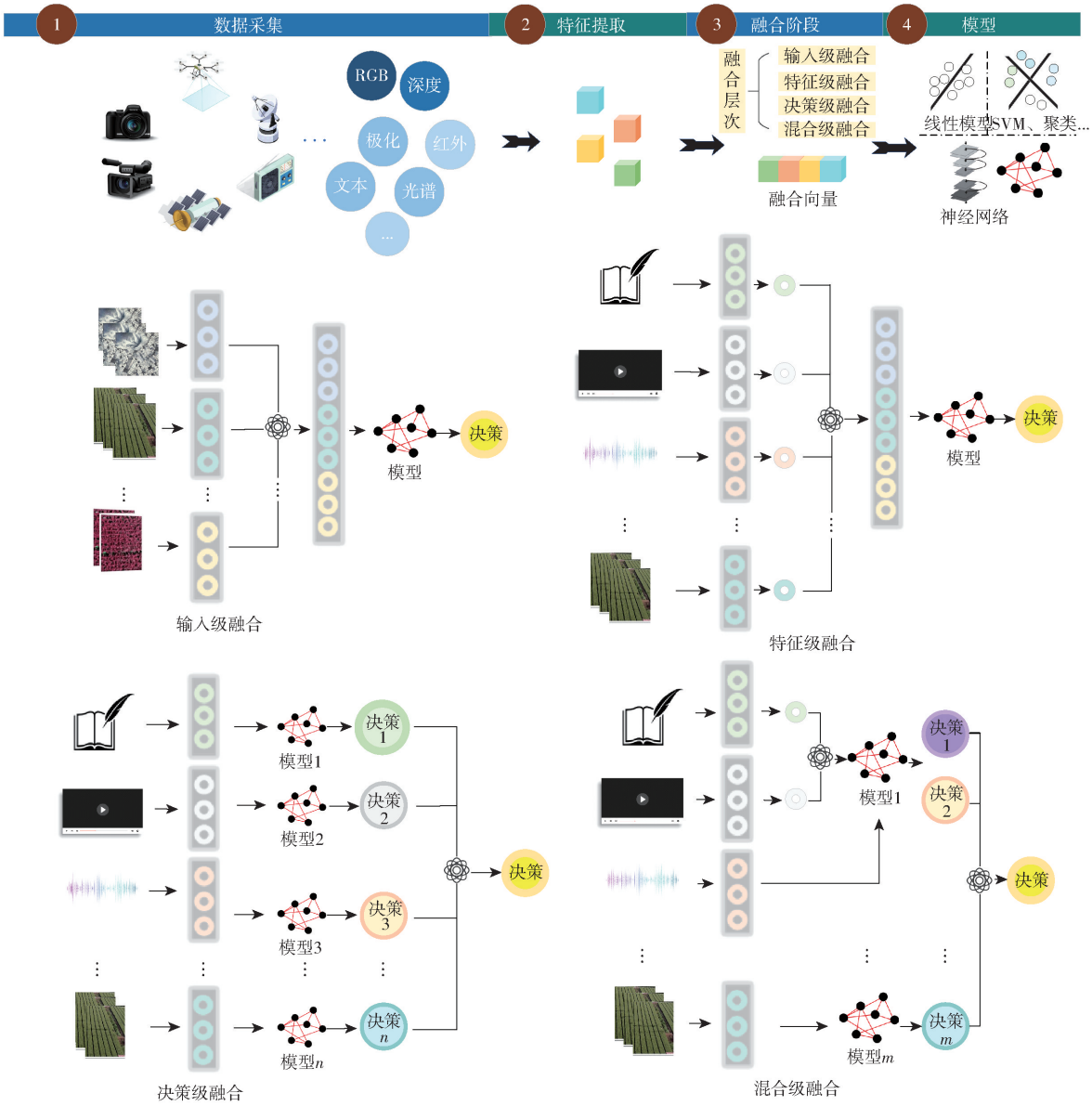


图 1　多模态融合过程

Fig. 1　Multi-modal fusion process flowchart

改进了模型的融合效果，但这种图注意力机制的构建和优化过程相对复杂，当处理大规模多模态数据时，计算成本较高。此外，噪声和不完全数据的处理以及如何构建不同模态数据之间的关系模型依然是有待解决的问题。

1.3　基于多阶段的渐进式融合

直接融合并行网络所提取的多模态特征图可能会出现信息排斥现象。而引入渐进式融合学习框架，可以在线性空间中显式提取模态的共享信息，以实现多模态特征的动态融合[26]。例如，LIU 等[27]提出了一种多模态分层融合方法，通过利用并行注意力机制对不同骨干网络的 C3 ~ C5 特征层进行分层融合，以增强多模态特征的融合效果。FAN 等[28]提出了一种基于对抗学习方法的多级交互式融合网络，用于 HSI 和 LiDAR 数据的分类，实现多源异构数据之间互补特征的有效结合。此外，研究中引入了监督 MLFFC 模块，能够进一步融合编码网络中学习到的所有层次特征之间的互补性，改进分类结果。

基于多阶段的渐进式融合通过多阶段学习，能够在一定程度上弥补不同模态间的差异性。但是模型的复杂性也会引起计算成本的增加。因此，未来的研究应着眼于优化网络结构和算法，在确保分类识别准确性的同时减少计算资源的消耗。

1.4　基于子空间的融合

基于子空间的融合方法（Subspace-based fusion methods，SFM）通过学习多模态的公共子空间，将原空间中不易学习或者难以识别的区域在子空间中展开，或者保留原空间优势，将样本映射到子空间中。基于视觉数据的可用性，子空间聚类在计算机视觉领域得到了广泛研究[20,29]。张立杰等[30]在苹果果

径与果形分级的研究中,通过输入级图像融合,将深度图像与 RGB 图像中的 GB 通道重组为 DGB 图像,实现了苹果果形的分级与信息输出(如图 2a)。XU 等[31]在麦田杂草检测中,通过 IHS 变换将深度图像编码为 DHS 图像,并结合多尺度检测和注意力机制进行多模态融合(如图 2b)。

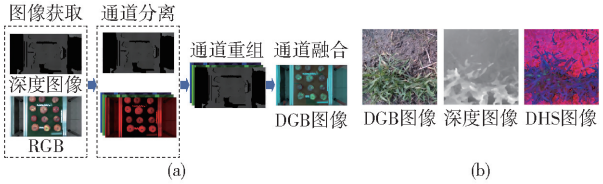


图 2 基于子空间的图像融合示例

Fig. 2 Example of image fusion based on subspace

在光谱数据中,鉴于其所具有的空间、光谱特征,可以通过将子空间聚类推广到双子空间聚类中来保留两种特性。这种方法不仅能够更好地捕捉光谱数据中的复杂关系,还能在数据维度上提供更为丰富的特征表达,使得聚类结果更具代表性。

1.5 基于 Transformer 的融合

基于 Transformer 的网络性能高,被广泛应用于计算机视觉任务[32-34]。Transformer 的多头注意力模块可以突出显示每种模态的具体特征,加强对关键特征的关注。胡学龙等[35]提出改进的 Fish-MulT 算法,对鱼类摄食强度进行识别。通过使用融合模态替代各支路的跨模态 Transformer,解决了在跨模态 Transformer 中,视觉模态的潜在适应性无法表征其他模态间适应性的问题。LOU 等[36]提出了一种基于 Transformer 和多层残差的多模态深度融合模型,用于评估杂草的竞争力。其中,多层残差融合模块用于对不同的分支进行浅层特征融合。既考虑到不同模态特征之间的相互关系,也平衡了细微的特征差异和更抽象的语义信息,最后,通过采用多层次分层深度融合方法整合浅层特征,进一步增强了对农田杂草竞争的综合评估能力。基于 Transformer 的机制能够有效融合不同模态中的特征表示,但在融合中还需要进一步考虑到模态对齐的效果,以防止对融合的结果产生影响。

表 1 列举了农业领域多模态融合的先进方法,总结了各方法在不同数据类型下的融合方式,并分析了每种方法的优缺点。从表 1 可以发现,多模态融合方法能够有效整合来自不同数据源的异构信息。然而,这些方法在实际应用中也面临诸多挑战。因此,未来的研究可以进一步优化融合模型,探索更加精确、高效的融合方法。

表 1 农业领域多模态融合先进方法

Tab. 1 Advanced multi-modal fusion methods in agricultural field

融合方法	数据类型	描述	优点	缺点	文献序号
基于线性的融合	RGB + 高光谱图像	软注意力机制降维,基于双线性融合特征	简单,易于实现,适用于同质或异质数据;有效捕捉模态间的互补关系	难以捕捉复杂关系,可能导致冗余,矩阵乘法存在维度问题	[17]
	光谱 + 多传感器	构建 CNN 分类模型			[18]
	X 射线 + RGB	建立 CNN-LSTM 检测模型			[19]
基于多流分支的融合	GF2 图像 + 多时相 Sentinel-2 图像	采用基于注意力的时空融合模块	最大化利用模态特征的互补性,减少信息丢失	处理大规模数据时计算成本高	[22]
	RGB + 深度图像	深度增强特征融合			[15]
	作物数据集:CropDeepv2	跨级融合			[25]
	高光谱图像 + 激光雷达	基于图注意力机制融合			[29]
基于多阶段的渐进式融合	RGB + 近红外 + 深度图像	采用多模态融合编码器,开发了 YOLO-DNA 轻量级模型	逐步解决模态差异,实现多模态特征的有效融合,提高分类识别精度	模型复杂性增加导致更高的计算成本	[26]
	RGB + 近红外	改进的注意力机制算法在 C3 ~ C5 特征层进行多模态特征融合			[27]
	高光谱图像 + 激光雷达	基于对抗学习的多级交互式融合网络			[28]
基于子空间的融合	RGB + 深度图像	改进型 SSD,网络轻量化	通过映射到公共子空间学习模态互补性	依赖于子空间选择和数据分布,对噪声敏感	[30]
	RGB + 深度图像	深度图像重新编码			[31]
基于 Transformer 的融合	水质 + 声音 + 视觉	跨模态融合	加强关键特征关注,平衡细微特征和抽象语义信息	存在潜在的过拟合,需要关注模态对齐	[35]
	激光雷达 + 高光谱图像 + 冠层高度	多层残差融合			[36]

2 多模态特征融合技术应用

作物的生长受到环境和栽培方法等多种因素的影响,准确并及时地获取作物的生长信息有助于评估其生长状况与产量潜力,并能及时采取相应措施[37-38]。传统的监测方法不仅耗时费力,而且在实时性上面临巨大的挑战[39]。随着计算机视觉技术的发展,基于图像的监测方法得到了广泛应用。该

方法通过成像技术获取作物的多源图像，再通过预处理、分割、特征提取等过程获取图像特征，最后结合机器学习和深度学习技术构建识模型[40]。常见的图像数据包括 RGB 图像[41]、光谱图像[42-43]、雷达图像[44]、热红外图像[45]以及深度图像[46]。然而，单一模态的图像数据都存在自身的局限性，表 2 展示了作物生长监测中常用图像数据的优点与缺点。

表 2　作物监测中常用的图像数据类型与特点

Tab. 2　Types and characteristics of image data commonly used in crop monitoring

图像数据	优点	缺点
RGB 图像	易于获取，空间分辨率高，可提供颜色、纹理、几何特征，有助于分析作物的生物特性	光照不均、作物果实叶片重叠遮挡、背景干扰等问题影响图像分割精度；提供的信息有限
高光谱图像	包含丰富的光谱信息，可以反映作物内部状态，分辨率高，光谱曲线完整	图像获取操作复杂，特征过多导致难以筛选有用特征
多光谱图像	数据获取成本低，波段多	图像分辨率低，光谱曲线不完整
雷达图像	提供作物的水分含量、农田的结构和土壤水的可用性信息等，反映垂直、水平结构信息	数据获取成本高
SAR 图像	采集作物全天候数据，提供大量的纹理和结构信息	易受环境干扰，空间分辨率低
热红外图像	提供作物表面的温度变化	图像存在背景噪声，图像分辨率低，获取信息单一
深度图像	反映植株三维立体特征，提供作物位置、轮廓、场景范围等空间深度信息，有利于区分重叠目标	图像存在孔洞问题，导致信息丢失

多模态融合技术可以将从不同单模态图像中提取的信息整合到统一的表示中，增强信息的互补性，从而获得更准确的作物状况分析，为在复杂的户外农业环境中完成视觉任务提供了更大的可能[47]。此外，一些非图像数据如文本数据，可以提供丰富的语义信息，通过与图像数据的融合，可以进一步增加样本数据的多样性，提高模型的准确性。

2.1　作物识别

作物监测首要任务是作物识别，之前的研究大多基于 RGB 图像进行。然而，杂草干扰、光照变化以及植株密集生长所带来的叶片遮挡问题都会影响模型识别精度。随着传感器技术发展，通过融合深度数据、红外数据等额外信息，可以获取目标轮廓、位置、温度等数据，从而有助于区分具有相似纹理与颜色特征的目标个体[48-49]。目前，多模态融合技术正逐渐成为作物识别领域的研究热点，相关研究成果如表 3 所示[50-55]。

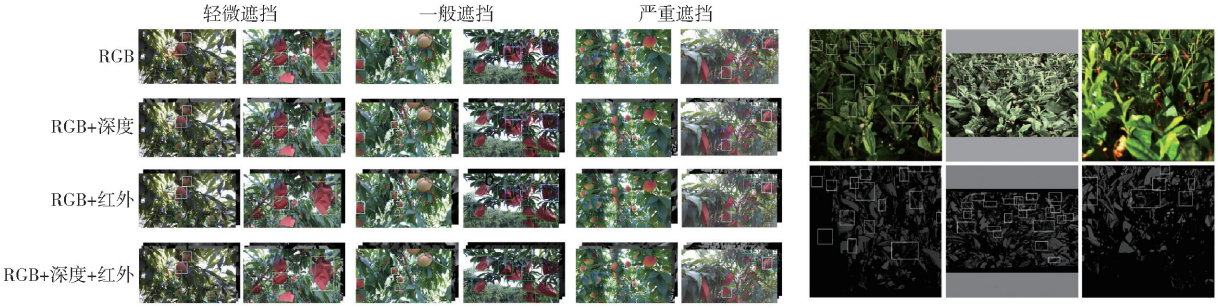
表 3　基于多模态融合的作物识别研究成果

Tab. 3　Research on crop identification based on multi-modal fusion

研究对象	数据类型	描述	性能	文献序号
番茄识别	RGB + 深度	提出 RD - SSD 模型，由并行子网分别提取特征，拼接特征后输入识别模型	mAP 为 91.47%	[50]
茶苗识别	RGB + 深度	提出 YOLO - RGBDtea，设计并行特征提取网络，引入自注意力机制对特征优先级排序，设计跨模态空间注意力融合模块(CSFM)单向融合特征	mAP 为 92.4%	[12]
桃子识别	RGB + 深度 + 红外	引入注意力机制实现融合，设计轻量级 YOLO v5s 模型优化检测	裸桃 mAP 为 98.6%；套袋毛桃 mAP 为 88.9%	[51]
小麦识别与分类	光谱 + RGB + 雷达	分别提取特征并拼接，评估随机森林等机器学习算法和深度循环网络算法分类效果	mAP 为 94.3%	[54]
小麦识别与分类	光学 + 雷达	提出 DOCC 框架，引入注意力模块	mAP 为 94.3%	[55]

WANG 等[50]提出了一种基于彩色和深度图像融合的模型(RGB - D single shot multibox detector，RD - SSD)，用于生长阶段番茄的识别。通过融合纹理特征与深度特征，在番茄成熟期的识别精度达到 91.47%。罗庆等[51]利用 RGB、红外及深度图像构成的多模态图像集实现桃子的检测，通过在 YOLO v5s 模型的基础上添加方向感知和位置敏感的注意力机制，使得模型可以在一个空间方向捕捉长距离依赖，同时在另一个空间方向保留准确的位置信息。研究发现，在人工光照条件下，添加深度图像有助于减少模型误判，而在明亮光照环境下，增加红外图像可以有效区分图像中的水果和背景。当同时使用所有成像模式时，在任

何照明环境中都可以获得最佳结果,如图 3a 所示。这种基于多模态视觉数据的方法对于裸桃和套桃的分类精度分别达到 98.6% 和 88.9%,相比仅使用单一 RGB 图像,精度分别提高 5.3% 和 16.5%。然而,RGB-D 传感器成像受到原理和精度的限制,深度图像容易出现孔洞问题[52-53]。为了减少低质量深度信息带来的不利影响,WU 等[12]基于 YOLO v7 开发了一种增强的端到端 RGB-D 多模态茶苗检测网络 YOLO-RGBDtea,通过调整大小、缩放、纵横比失真和随机水平翻转等数据增强措施扩展数据库,如图 3b 所示。此外,该模型采用添加了自注意机制的并行轻量级深度图像特征提取主干网络,通过引入单向互补的多模态融合方法,实现深度特征与 RGB 特征的集成,在面对复杂背景下的茶树苗检测时,准确率达到 92.40%。光学遥感数据在晴天条件下可以很好地捕捉到地表的光谱信息,而雷达遥感数据不受天气和光照的影响,能够在多云和夜间条件下进行观测。为了利用雷达数据特征,冯权泷等[54]融合光学影像与雷达影像的光谱、纹理与极化特征,对小麦进行识别分类。研究表明,基于随机森林的识别方法在融合了时序 Sentinel-1 雷达数据与时序 Sentinel-2 光学数据上的识别精度最高,达到 94.3%,而单独使用 Sentinel-1 雷达数据和 Sentinel-2 光学数据总体精度分别为 87.38% 和 93.95%。LEI 等[55]提出了多模态融合框架 DOCC,用于冬小麦的识别与分类。该框架的提取网络由 4 个卷积层和可替换通道注意模块组成,可以实现光学与影像特征的自动提取,在所有多模态融合的遥感数据上获得了最高 F1 分数。



(a) 不同遮挡场景中使用不同模态组合的检测效果 (b) RGB-D数据增强

图 3 作物识别中多模态融合的处理与应用效果

Fig. 3 Multi-modal fusion processing and application effect in crop identification

分析上述文献发现,RGB 与深度图像的融合可能受到深度传感器的固有限制,导致深度图像中的孔洞问题仍对识别效果产生负面影响。通过加入自注意力机制和跨模态空间注意力模块,可以在一定程度上提高基于 RGB 与深度图像融合的效果。未来的研究应着眼于开发更轻量级的模型,进一步优化融合算法,提升多模态数据的利用效率和识别精度。

2.2 性状分析

性状分析是了解植物生理过程的前提,也是精准管理作物生产过程的关键[56]。目前主流方法是基于光学遥感对作物形态、生化、生理学以及性能特征进行预测分析[57]。表 4 展示了多模态融合在性状分析中的相关研究成果。

表 4 基于多模态特征融合的性状分析研究成果

Tab. 4 Research on crop identification based on multi-modal feature fusion

研究对象	数据类型	描述	性能	文献序号
生菜叶面积、直径、干湿质量、干质量估计	RGB + 深度	基于 ResNet50 提出生菜表型参数估算模型	R^2 大于 0.81	[58]
冬小麦冠层结构参数与叶面积指数估计	RGB + 多光谱	基于 ResNet 50 提取图像特征,构建 MMF 模型,融合图像与 VI 特征	mAP 为 92%	[59]
蔬菜表型性状	LiDAR 点云 + 多光谱	特征级融合,利用支持向量机(SVM)和随机森林(RF)算法估计	mAP:79%(西红柿);84%(茄子);76%(卷心菜)	[60]
柑橘叶面积指数估计	RGB + 点云	双分支提取网络,融合特征输入 VPNet,完成预测	R^2 为 0.861	[61]

朱逢乐等[58]利用 ResNet50 提取生菜的 RGB 与深度图像特征,并将拼接后的生菜表型参数特征向量输入回归网络,实现生菜表型参数的估算。通过加入深度图像的方式融入生菜的三维立体形态信息,弥补了二维 RGB 图像的不足,使得模型在对生菜空间维度特征的估算上更加准确。结果表明,融合模型对生菜叶面积、直径、干湿质量以及干质量的估算决定系数均高于 0.94,相较于传统的估算方法

有了显著提升。CHENG 等[59]提出融合图像特征与多光谱图像特征的多模态融合模型 MMF－MTL，对冬小麦的生长进行监测。通过结合遥感图像以及多光谱图像提供的植被指数，可以同时利用小麦的空间特征与光谱信息。此外，多任务学习的方式能够共享特征，加快深度学习模型的训练速度，提升整体模型的准确性。与 ResNet 50 模型相比，MMF－MTL 对小麦的叶面积指数（LAI）、地上生物量（AGB）、株高（PH）和叶片叶绿素含量（LCC）的 R^2 分别提高 0.123 1、0.164 2、0.045 5、0.245 9。LU 等[61]基于 RGB 与点云数据构建多模态回归模型（Visual-physical network，VPNet），通过利用 RGB 数据提供的颜色、纹理信息，以及点云数据提供的位置、大小、方向信息，实现了柑橘冠层结构参数与叶面积指数在更精细空间尺度上的估计。相对于 PCNet，VPNet 的 MAE 未改变，MSE 下降 42.9%，R^2 由 0.805 增加至 0.861。

总体而言，多模态融合技术的发展提高了作物表型参数估计的准确性和全面性。一些研究试图通过融合 RGB 与光谱数据来提高模型评估的潜力。但生物物理特性的可辨别特征主要与作物的结构-几何属性有关，且不同类型作物在反射域中存在固有的光谱相似性，因此基于光学数据融合的方法为生物物理表征提供可量化特征的能力有限。上述研究中采用加入点云数据方法在反演作物的精确垂直结构中具备潜力。但点云数据包含了高密度的三维信息，在与不同数据源的对齐和融合中存在难题。未来需要开发更高效的多模态融合算法，并探索光探测数据和测距数据、多视角立体视觉数据（Multi-view stereo，MVS）以及合成孔径雷达数据（Synthetic aperture radar，SAR）等的融合潜力。

2.3　产量预测

产量预测关系到粮食安全保障以及可持续性的农业管理[62]，通过利用多源数据特征可以提高模型的预测精度，保障作物的稳产高产。其中，归一化差异植被指数（Normalized difference vegetation index，NDVI）反映了植被绿色程度，展现了植被发育状况，是评估作物产量的有效指标[63]。可见光/近红外光谱通过光谱数据提供对作物内部质量的洞察。此外，土壤状况以及其他影响因素也与作物产量之间存在非线性关系。表 5 展示了基于多模态融合的产量预测部分研究成果。

表 5　基于多模态融合的产量预测研究成果

Tab. 5　Research on yield prediction based on multi-modal fusion

研究内容	数据类型	描述	性能指标	文献序号
大豆产量预测	RGB＋多光谱＋热红外	基于 DNN 评估了输入级与中级特征融合	R^2 为 0.691（DNN－F1），R^2 为 0.72（DNN－F2）	[64]
小麦产量估计	多光谱＋光强＋土壤特性	融合前茬作物的长势信息与当季作物不同时期的多时相数据，基于 GPR 估算产量	R^2 为 0.92	[65]
小麦产量估计	多光谱＋热红外＋坡度因子	特征级融合，MLR、PLSR、RFR 估算产量	R^2 为 0.865	[66]
产量预测	卫星图像＋气象数据	提出 MMST－ViT，引入注意力机制与多模态对比学习融合特征	R^2 为 0.843	[67]

研究表明，通过利用不同传感器系统的冠层光谱、结构、热红外和纹理信息的组合可以改善农业应用中的植物性状估计结果。MAIMAITIJIANG 等[64]将 RGB、多光谱、热红外数据特征作为输入，构建了基于深度神经网络（Deep neural network，DNN）的输入级融合模型 DNN－F1 和基于特征级融合的模型 DNN－F2。两种模型 R^2 分别达到 0.691、0.72，展现出 DNN 模型的强适应性。轮作体系的前茬作物信息能够表征土壤肥力状况，作为现有模态数据的补充。将前茬作物的长势信息与当季作物不同时期的多时相数据融合，能够进一步提高产量预测的精度。李阳等[65]基于玉米作物信息、土壤特性信息、施肥信息以及小麦的长势信息构建多时相多模态参数的小麦产量估计模型，相较于其他融合方法，该模型 R^2 提高 2% ~41%。张少华等[66]基于多光谱、热红外和坡度因子，使用随机森林算法构建小麦的多模态产量预测模型，通过充分利用冠层温度、地形及植被指数多源数据特征，提高了估算模型精度，为作物的产量估算提供了技术支持。由于作物生长对生长季节的天气变化非常敏感，产量的及时预测依旧存在挑战。LIN 等[67]基于视觉 Transformer 模型（Vision transformer，ViT）提出多模态时空视觉模型 MMST－ViT，对作物产量进行预测，如图 4 所示。短期的天气变化可能对作物的瞬时生长状态产生直接影响，而长期的气候趋势则影响作物的整个生长周期。利用多模态多头注意力机制捕获天气因素的影响，实现了视觉遥感数据和气象数据的有效融合，并能够同时捕捉短期气象变化和长期气候变化对作物

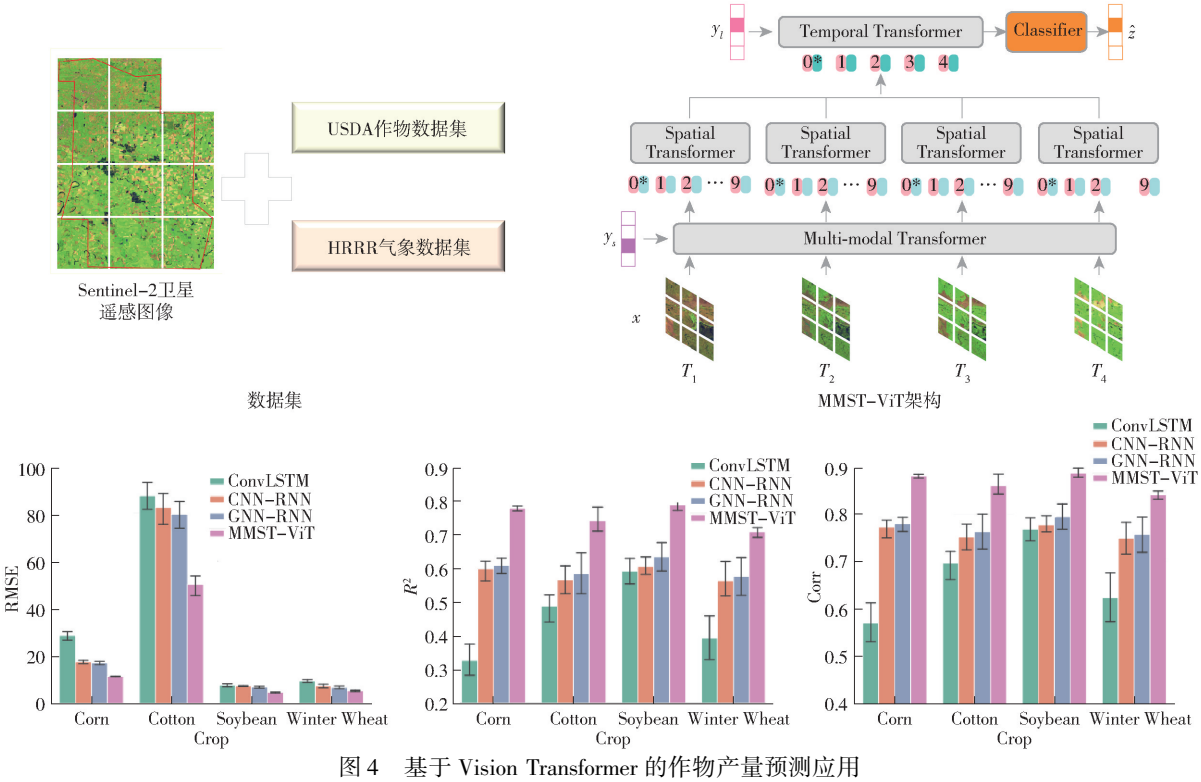


图 4 基于 Vision Transformer 的作物产量预测应用

Fig. 4 Application of Vision Transformer-based crop yield prediction

产量的影响。此外通过多模态对比学习方式,避免了 ViT 模型过拟合问题。最终模型均方根误差(RMSE)达到 3.9,R^2 达到 0.843。

总体而言,土壤信息、气象信息等环境因子在作物产量的评估中具有重要的作用。此外,前茬作物信息表征了土壤的肥力状况,是土壤信息的补充,能够提高土壤状况的量化能力。多时相数据能够综合考虑作物生长周期中的各个阶段。而在提取到的作物遥感信息中,基于单一光谱信息的产量预测结果优于 RGB 与热红外信息。在一定程度上,融合的参数越多,模型评估的更准确。但是,多参数的数据处理和融合过程相对复杂,计算资源需求高。机器学习的方法难以提取有效特征。通过引入多头注意力机制,可以提高模型对多模态数据的处理能力,为作物产量及时预测提供了参考。未来研究中,可以探索更轻量级的深度学习模型。

2.4 病虫害诊断

植物病虫害会影响作物生长发育过程,导致作物减产[68]。因此,加强病虫害的监测对提高农作物的品质意义重大。病虫害主要作用于作物的代谢系统,会导致作物外部形态以及内部生物理化特征变化,进而导致光谱特性改变[69-71]。随着技术的发展,利用监测传感器技术以及相关数据处理技术可以提取植物病害病理特征。表 6 展示了多模态融合在病虫害诊断中的部分研究成果。

表 6 基于多模态融合的病虫害诊断研究成果

Tab. 6 Research on pest and disease diagnosis based on multi-modal fusion

研究内容	数据类型	描述	性能	文献序号
柑橘黄龙病预测	RGB + 高光谱	双分支结构,基于 ResNet50 提取图像特征,引入注意力机制对高光谱数据降维,基于双线性融合特征	mAP 为 97.89%	[17]
番茄健康状态分类	热 + RGB + 深度图像	图像配置,提出局域和全局的特征提取	AP 为 89.93%	[72]
水稻病害诊断	气象 + RGB	构建 Rice-Fusion 双分支。MLP 提取文本特征,CNN 提取图像特征,特征串联后进行预测	ACC 为 95.31%	[35]
荔枝枯树病预测	环境 + 光谱	LSTM 提取序列化特征,SPA 提取光谱特征,基于贝叶斯网络设计自适应权重,进行决策级融合	AP 为 89.58%	[73]
病害分类	RGB + 文本	提出 BimodalFINet,通过双分支结构提取语义和空间位置特征,特征层拼接融合,输入超图神经网络	AP 为 93.22%	[74]
黄瓜病害识别	RGB + 文本	结合图像-文本多模态对比学习以及图像自监督对比学习,关注不同模态的相关性	ACC 为 94.84%	[75]

YANG 等[17]基于 RGB 与高光谱图像构建多模态网络架构，对柑橘黄龙病症状进行分级诊断。网络分别通过 ResNet50 与添加注意力机制的神经网络提取叶片的图像特征和光谱特征。研究通过比较 3 种不同的早期融合结果后发现，双线性早期融合方法的拟合速度最快，识别准确率最高。病虫害的感染可能改变植物的呼吸作用和蒸散速率，从而导致热蒸发水平的明显变化，而热成像技术可以在早期发育阶段有效地监测温度的不规则性[76]。然而，与植物热成像相关的主要问题之一是作物的冠层结构、环境条件以及植物区域与相机深度（距离）也会导致作物的温度变化，影响诊断结果。此外，在作物疾病早期，难以依靠肉眼捕捉作物的疾病状态，因此仅靠颜色特征也不能提供良好的检测结果。为了克服该问题，RAZA 等[72]融合热、可见光以及深度图像特征，对番茄植株的健康状况进行分类。结果表明，在植株接种病毒 70 d 后，相较于单一特征或两个特征的组合方法，该方法分类准确性提高 5%，证明了在作物早期疾病检测的巨大潜力。环境变化也是植物病虫害发生和传播的关键因素，对准确识别病害类型至关重要[77]。PATIL 等[73]采用特征串联的方式融合农业气象文本信息（温度、相对湿度、土壤湿度及土壤养分含量等环境属性值）与水稻的图像特征，构建水稻病害诊断的特征级多模态融合框架，最终实现了 95.31% 的准确率，显著优于仅使用图像数据的 CNN 模型（82.03%）以及仅使用环境数据的 MLP 模型（91.25%）。通过异质模态数据的融合克服了光照变化以及复杂背景给图像数据检测带来的影响。LU 等[78]利用物联网设备以及光谱仪分别采集荔枝生长阶段的环境数据和叶片的光谱数据，基于贝叶斯网络构建多模态决策级融合的荔枝枯树病预测模型。由于荔枝霜枯病与温度、湿度等环境因素之间存在很强的相关性，充沛的降雨和高湿度都会导致荔枝霜枯病的发生和蔓延。因此通过结合叶片生长的环境数据与高光谱数据，可以从多个角度提高荔枝霜枯病预测的精度，预测准确率达到 89.58%。

有研究指出，由于害虫可能食用植物的叶子和茎，从而引起植物叶面积和生物量的减少，导致光谱特异性的缺失[79]。此外，部分植物的病虫害类型缺乏可识别的特征。因此，仅基于光谱特性或图像特征限制了病虫害识别的性能。通过融合图像和文本特征，可以利用文本模态丰富的语义信息，与图像特征形成互补，从而提高特征集的表达能力[80]。张净等[74]分别利用基于递归神经网络的文本处理模型（Text recurrent neural network，TextRNN）与结合了 ResNext 和通道注意力机制（Channel attention，CA）的图像处理模型 ResNext50－CA 提取双分支文本与图像特征，再基于超图神经网络实现特征的融合，对 5 种农作物的 7 种病害进行分类测试，平均准确率达到 93.22%。王春山等[81]构建双模态的病害识别模型，对黄瓜、番茄、桃进行病害识别。最优模型组合的准确率达到 99.47%，证明了图像–文本的融合在作物病害识别中的有效性。然而，这些方法将图像和文本映射到独立的特征空间，没有充分利用不同模态之间的数据相关性。CAO 等[75]在对黄瓜病害进行识别的过程中，通过结合图像–文本多模态对比学习以及图像自监督对比学习的训练方式，缩短相匹配的图像与文本以及同类别图像之间的距离，充分利用了数据间的互补性，模型在黄瓜病害数据集上的识别准确率达到 94.84%。刘立波等[82]基于 Transformer 模型构建枸杞虫害的图文检索模型，通过注意力机制聚合特征向量挖掘数据的语义特征，利用跨模态联合损失函数挖掘模态之间语义特征关系，最终平均精度提高 1.1% ~19.5%。

总体而言，现阶段大部分的病害识别模型都是基于图像模态构建的，但作物的病害症状并不都能通过表层特征显露出来，有些病害的特征还存在一定的相似性。此外，光线变化或背景干扰会影响图像的采集质量。因此，仅基于图像模态的病虫害诊断具有局限性。上述研究中通过图像与文本的融合提供了丰富的语义特征，为病虫害诊断提供了新的方向。但图像与文本的融合属于一维和二维向量的跨模态融合，融合过程更需要关注模态之间的相关性。未来的工作在收集作物各种病害的平衡数据集的基础上，需要关注如何充分利用模态之间的相关性，以提高小样本下不同类型疾病识别的准确性。

2.5　胁迫分析

尽早发现作物的胁迫状况可以最大限度地减少生产力损失[83]。胁迫会引起作物的叶片和表面结构的变化，改变植物叶子或树冠光反射。常见的胁迫类型主要包括水分胁迫与营养胁迫。表 7 为多模态融合在胁迫分析中的部分研究成果。

水分胁迫是导致农业产量损失的主要因素，也是精准灌溉的先决条件[84]。以往的研究主要依赖二维热红外图像进行水分胁迫检测，忽视了叶片位置和检测位置的影响。WANG 等[85]利用三维重建技术开发了一种结合热红外成像和双目立体视觉系统的低成本三维运动机器人系统，对马铃薯植株的深度、温度和 RGB 颜色信息进行自动采集。图 5a 展示了数据的采集与处理过程。系统首先通过图像配准和几何校准技术融合多模态数据，生成三维点

表 7 基于多模态融合的胁迫分析研究成果

Tab. 7 Research on crop stress analysis based on multi-modal fusion

研究内容	数据类型	描述	性能指标	文献序号
马铃薯水分胁迫检测	深度 + 温度 + RGB	开发三维(3D)运动机器人系统采集 RGB 与热数据,采用融合点云方式进行输入级融合		[85]
冬小麦干旱检测	图像 + 文本	构建 S - DNet 模型,SPEI 与 DenseNet - 121 双分支提取特征,自适应加权实现决策级融合	ACC 为 96.4%	[86]
水分胁迫估计	环境 + 图像	提出多模态滑动窗口的支持向量回归(SW - SVR),基于 DNN 与 ROAF 提取特征	MAE 降低 20%	[87]
水分胁迫估计	环境 + 图像	基于聚类的下降(C - Drop),加入长短期记忆层(LSTM)与动态注意力机制	MAE 和 RMSE 降低 21%	[88]
营养诊断	图像 + 光谱	提出多级融合方法,模型引入跳过连接	ACC 为 88.8%	[89]
小麦氮含量预测	光谱 + 卷积	获取小麦冠层光谱数据,CNN 提取冠层多模态特征,构建 SVR、PLSR、PSO - SVR 模型,评估预测结果	R^2 为 0.896	[90]

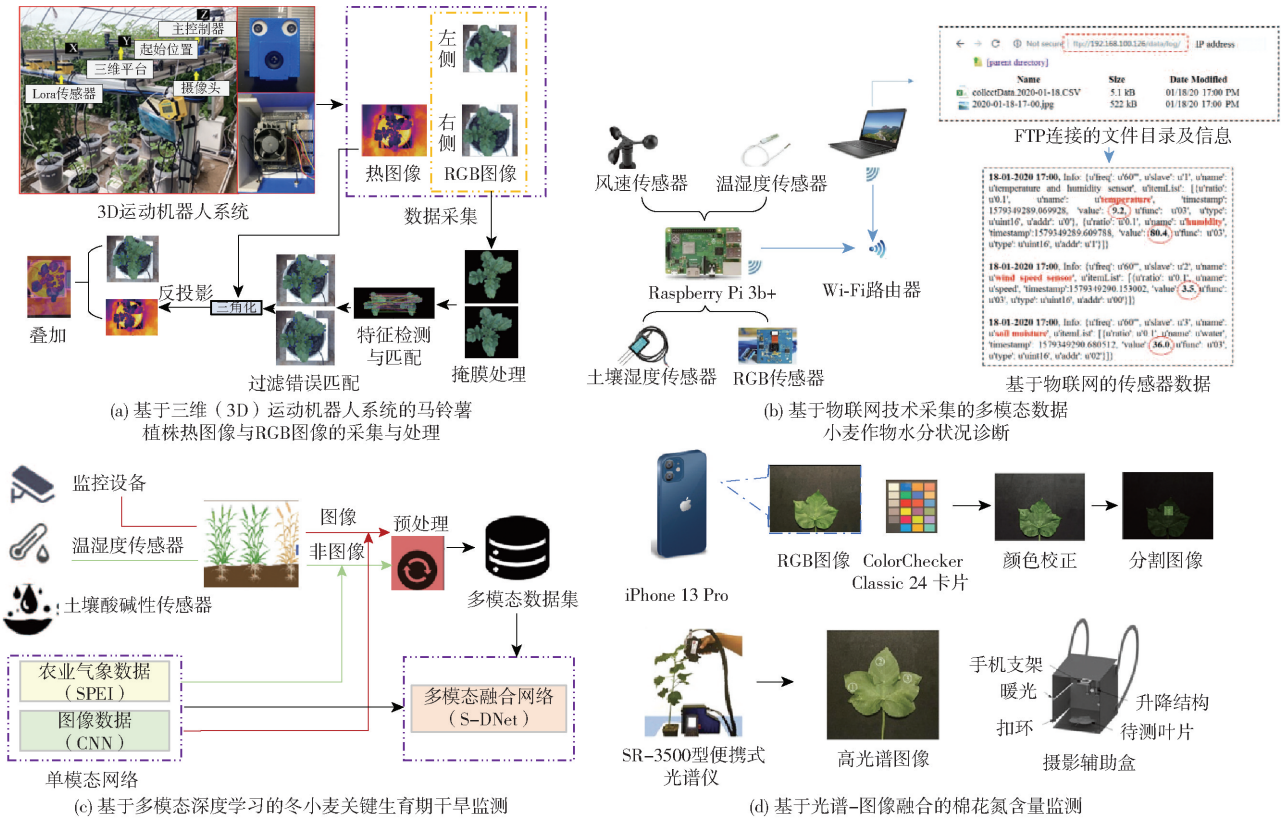


图 5 基于多模态融合的作物胁迫分析应用

Fig. 5 Application of crop stress analysis based on multi-modal fusion

云,然后再通过计算作物水分胁迫指数(Crop water stress index,CWSI),实现马铃薯植株水分胁迫状态的三维检测。

除了监测作物表型结构的变化外,环境气象数据如温度、相对湿度、太阳辐射量与叶片的蒸腾作用密切相关,可以提供水分胁迫分析的互补信息。ELSHERBINY 等[91]利用物联网技术采集到的多源传感器信息对小麦水分状况进行诊断,如图 5b 所示。通过将深度学习网络与多模态数据(包括环境变量和图像数据)相结合,为作物水分状况评估提供了一种低成本的高效方法。YAO 等[86]提出一种多模态深度学习模型(Sensor-data network,S - DNet),用于冬小麦干旱胁迫的高精度识别和监测,图 5c 展示了应用流程。模型采用决策级融合方法,对分别基于 DenseNet - 121 的图像分支与 SPEI 文本分支得到的预测结果进行自适应加权融合,从而充分利用图像与非图像数据特征,平均干旱识别准确率达到 96.4%,提高了小麦关键生育期干旱胁迫评估的准确性和效率。虽然 CNN 在基于图像的分析领域出现了大量成果,但植物图像数据中存在大量与水分胁迫预测不相关的信息,导致 CNN 不能正确提取所需特征。KANEDA 等[87]提出了一种基于

多模态滑动窗口的支持向量回归(Sliding window support vector regression for multimodal data, SW - SVR)的新型植物水分胁迫预测方法,更适合具有时间相关特征的数据学习,但具体性能需要进一步研究。WAKAMORI 等[88]提出了一种基于聚类的下降多模态神经网络,通过加入长短期记忆层与动态注意力机制,解决了叶片枯萎和环境数据与茎径变化存在的时间依赖关系,优化了水分胁迫估计的特征。与单模态方法相比,植物水分胁迫估计精度平均绝对误差和均方根误差均降低 21% 左右。

作物营养的缺乏或过剩状况都会对作物的品质造成影响[84]。作物早期的营养胁迫症状一般不明显,这使得及时诊断和处理变得更加困难。而在作物成熟阶段,由于作物的株高不再发生明显变化,因此作物营养状况的外在变化主要反映在以冠层面积的横向扩张上。此外,冠层光谱中的反射光谱数据也可以用于预测作物营养状况。为了弥补单一数据源的不足,QIN 等[89]结合决策森林结构(Decision forest, DF)与堆叠集成学习方法,开发图像-光谱多级融合模型,实现棉花氮含量的预测,图 5d 展示了数据采集与处理流程。研究表明,二级决策级融合遵循决策级融合的构建概念,通过整合特征级和决策级融合的预测值,形成最终的联合决策结果,表现出显著的抗噪性。R^2 提高 0.118 ~ 0.125,有效提高了棉花叶片氮素营养监测模型的回归精度。GAO 等[90]提出了一种基于冠层光谱特征与卷积特征融合的小麦氮含量预测模型。其中,卷积特征代表了冠层反射光谱的深层语义信息,可以弥补光谱特征鲁棒性的不足。使用多通道的 CNN 提取一维光谱数据中的卷积特征,再通过特征优化和池化层对多模态特征进行降维处理,从而显著提高了 LNC 估计模型的准确性。结果表明,基于融合特征的 PSO - SVR 模型的 R^2 达到 0.896。

总体而言,胁迫会导致叶片的形态发生变化,因此评估过程中通过引入三维运动数据可以进一步提高估计精度。此外,还需要充分考虑时间序列信息与环境特性。在融合异构数据中,多级融合模型通过逐级融合信息的方式,能够缓解不同模态的数据在时间、空间或语义上的不匹配问题,增强特征的表达能力。未来的研究采用多级融合的方法,探索融合深度信息、气象数据以及时间数据等的应用潜力。

3　展望

在农业生产过程中,综合利用多源数据能有效监测作物生长状况,然而,现阶段该领域相关研究较少,现有研究中依然存在一定的局限性。主要包含以下方面:

(1)跨模态数据的应用范围小、数据的利用程度低。具体来说,多模态融合技术目前主要是融入构造文本描述、物联网技术获取的气象数据用于病虫害诊断方面,在更加广阔的其它农业方向上的应用涉及较少,数据的综合利用程度较低,未来可以考虑引入视频、时间序列数据等。

(2)图像数据的特征提取过程主要基于卷积神经网络。光谱数据处理没有自监督模型,有效特征的提取存在难题。一些基于输入级图像融合的应用中,应针对信息缺失进行相关调整。例如 RGB - D 的融合需要关注深度图像的孔洞问题。一些反演作物三维结构数据如点云数据,受到农作物高度和复杂土壤背景的影响[60]。此外,不同图像信息间的融合还需要重点关注对齐操作。

(3)多源数据的融合能够提高单模态监测的准确性。然而,盲目地融入过多数据也会导致冗余,反而影响模型精度,此外,针对不同的应用场景,数据选取的侧重点不同。例如,在表型性状分析中,需要关注作物的结构特性。而在生物量的分析中,这种外部形态特征可能难以满足要求,而需要挖掘作物更深层的特征。未来需要探索更多不同数据与预测分类结果之间的相关性,例如多时相数据、前茬作物长势信息等,形成有效的数据组合。

(4)大多数研究采用特征级融合,混合级融合相关研究较少。特征级融合中对不同模态数据的处理还停留在简单的特征拼接上,没有关注到模态之间的特异性,对于跨模态数据的处理也是基于分离空间进行的。因此,未来可以通过对比学习、注意力机制等进行相关改进,此外,还可以通过分级融合的方式进行模态间的深度融合。

参　考　文　献

[1]　岳学军,宋庆奎,李智庆,等. 田间作物信息监测技术的研究现状与展望[J]. 华南农业大学学报,2023,44(1):43 - 56.
YUE Xuejun, SONG Qingkui, LI Zhiqing, et al. Research status and prospect of crop information monitoring technology in field [J]. Journal of South China Agricultural University, 2023, 44(1): 43 - 56. (in Chinese)

[2]　王鹏新,田惠仁,张悦,等. 基于深度学习的作物长势监测和产量估测研究进展[J]. 农业机械学报,2022,53(2):1 - 14.
WANG Pengxin, TIAN Huiren, ZHANG Yue, et al. Crop growth monitoring and yield estimation based on deep learning: state of the art and beyond[J]. Transactions of the Chinese Society for Agricultural Machinery, 2022, 53(2): 1 - 14. (in Chinese)

[3] 马彦鹏,边明博,樊意广,等. 基于冠层光谱和覆盖度的马铃薯叶片钾含量估算方法[J]. 农业机械学报, 2023,54(12):226 - 233.
MA Yanpeng,BIAN Mingbo,FAN Yiguang,et al. Estimation of potassium content in potato leaves based on canopy spectrum and coverage[J]. Transactions of the Chinese Society for Agricultural Machinery,2023,54(12):226 - 233. (in Chinese)
[4] 唐子竣,向友珍,王辛,等. 利用相关矩阵法优化光谱指数的冬油菜氮素营养诊断[J]. 农业工程学报, 2023, 39(17):97 - 106.
TANG Zijun,XIANG Youzhen,WANG Xin,et al. Nitrogen nutrition diagnosis of winter oilseed rape using spectral indexes optimized by correlation matrix method[J]. Transactions of the CSAE, 2023,39(17):97 - 106. (in Chinese)
[5] BERGER K, VERRELST J,FÉRET J B,et al. Crop nitrogen monitoring: recent progress and principal developments in the context of imaging spectroscopy missions[J]. Remote Sensing of Environment, 2020,242:111758.
[6] MIRZAEV K G, KIOURT C. Machine learning and thermal imaging in precision agriculture[C]// International Conference on Information, Intelligence, Systems, and Applications, 2023:168 - 187.
[7] 李庆松,康丽春,饶洪辉,等. 基于改进 YOLO v4 - Tiny 的自然环境下油茶果识别方法[J]. 中国农机化学报,2023,44(10):224 - 230.
LI Qingsong,KANG Lichun,RAO Honghui,et al. Recognition method of *Camellia oleifera* fruit in natural environment based on improved YOLO v4 - Tiny[J]. Journal of Chinese Agricultural Mechanization,2023,44(10):224 - 230. (in Chinese)
[8] 袁洪波,赵努东,程曼. 基于图像处理的田间杂草识别研究进展与展望[J]. 农业机械学报,2020,51(2):323 - 334.
YUAN Hongbo,ZHAO Nudong,CHENG Man. Review of weeds recognition based on image processing[J]. Transactions of the Chinese Society for Agricultural Machinery,2020,51(2):323 - 334. (in Chinese)
[9] HUANG S C,PAREEK A,ZAMANIAN R,et al. Multimodal fusion with deep neural networks for leveraging CT imaging and electronic health record:a case-study in pulmonary embolism detection[J]. Scientific Reports,2020,10(1):22147.
[10] ZHANG X,YIN X,GAO X,et al. Adaptive entropy multi-modal fusion for nighttime lane segmentation[J]. IEEE Transactions on Intelligent Vehicles,2024,31:1 - 13.
[11] 沈伟豪,钟燕飞,王俊珏,等. 多模态数据的洪涝灾害知识图谱构建与应用[J]. 武汉大学学报(信息科学版),2023,48(12):2009 - 2018.
SHEN Weihao, ZHONG Yanfei,WANG Junjue, et al. Construction and application of flood disaster knowledge graph based on multi-modal data[J]. Geomatics and Information Science of Wuhan University,2023,48(12):2009 - 2018. (in Chinese)
[12] WU Y,CHEN J,WU S, et al. An improved YOLO v7 network using RGB - D multi-modal feature fusion for tea shoots detection[J]. Computers and Electronics in Agriculture,2024, 216:108541.
[13] ZHANG Na,LIU Juan,JIN Yu,et al. An adaptive multi-modal hybrid model for classifying thyroid nodules by combining ultrasound and infrared thermal images[J]. BMC Bioinformatics,2023,24(1):315.
[14] ALGHOWINEM S,GOECKE R,COHN J F,et al. Cross-cultural detection of depression from nonverbal behaviour [C]//2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG). IEEE,2015:1 - 8.
[15] RONG J,ZHENG W,QI Z,et al. RTMFusion: an enhanced dual-stream architecture algorithm fusing RGB and depth features for instance segmentation of tomato organs[J]. Measurement,2024,239(15):115484.
[16] ZHONG Z,LIU X,JIANG J,et al. High-resolution depth maps imaging via attention-based hierarchical multi-modal fusion[J]. IEEE Transactions on Image Processing,2021,31: 648 - 663.
[17] YANG D,WANG F,HU Y, et al. Citrus huanglongbing detection based on multi-modal feature fusion learning[J]. Frontiers in Plant Science,2021,12:809506.
[18] REN G,WU R,YIN L,et al. Description of tea quality using deep learning and multi-sensor feature fusion[J]. Journal of Food Composition and Analysis,2024,126:105924.
[19] 李善军,宋竹平,梁千月,等. 基于 X-ray 和 RGB 图像融合的实蝇侵染柑橘无损检测[J]. 农业机械学报,2023,54(1):385 - 392.
LI Shanjun,SONG Zhuping,LIANG Qianyue,et al. Nondestructive detection of citrus infested by *Bactrocera dorsalis* based on X-ray and RGB image data fusion[J]. Transactions of the Chinese Society for Agricultural Machinery,2023,54(1):385 - 392. (in Chinese)
[20] LIU F,CHEN J,LI K,et al. STP - MFM: semi-tensor product-based multi-modal factorized multilinear pooling for information fusion in sentiment analysis[J]. Digital Signal Processing, 2024,145:104265.
[21] LIU Z,CHENG J, LIU L,et al. Dual-stream cross-modality fusion transformer for RGB - D action recognition[J]. Knowledge - Based Systems,2022,255:109741.
[22] CAI Z,HU Q,ZHANG X,et al. Improving agricultural field parcel delineation with a dual branch spatiotemporal fusion network by integrating multimodal satellite data[J]. ISPRS Journal of Photogrammetry and Remote Sensing,2023,205:34 - 49.
[23] BALTRUSAITIS T, AHUJA C, MORENCY L P. Multimodal machine learning: a survey and taxonomy[J]. IEEE Trans. Pattern Anal. Mach. Intell., 2019,41(2):423 - 443.
[24] ZHANG Y,CHEN W,CHENG C. Multi-modal network based on spatio-temporal and attention for emotion recognition[C]// IEEE International Conference on Bioinformatics and Biomedicine,2023:2364 - 2367.
[25] CAI J, ZHANG M, YANG H, et al. A novel graph-attention based multimodal fusion network for joint classification of hyperspectral image and LiDAR data[J]. Expert Systems with Applications,2024,249:123587.
[26] CHEN W,RAO Y,WANG F,et al. MLP-based multimodal tomato detection in complex scenarios: insights from task-specific

analysis of feature fusion architectures[J]. Computers and Electronics in Agriculture,2024,221:108951.
[27] LIU C,FENG Q,SUN Y,et al. YOLACTFusion: an instance segmentation method for RGB - NIR multimodal image fusion based on an attention mechanism[J]. Computers and Electronics in Agriculture,2023,213:108186.
[28] FAN Y, QIAN Y,GONG W,et al. Multi-level interactive fusion network based on adversarial learning for fusion classification of hyperspectral and LiDAR data[J]. Expert Systems with Applications,2024,257:125132.
[29] KONG J,WANG H,WANG X,et al. Multi-stream hybrid architecture based on cross-level fusion strategy for fine-grained crop species recognition in precision agriculture[J]. Computers and Electronics in Agriculture,2021,185:106134.
[30] 张立杰,周舒骅,李娜,等. 基于改进 SSD 卷积神经网络的苹果定位与分级方法[J]. 农业机械学报,2023,54(6):223 - 232.
ZHANG Lijie,ZHOU Shuhua,LI Na,et al. Apple location and classification based on improved SSD convolutional neural network[J]. Transactions of the Chinese Society for Agricultural Machinery,2023,54(6):223 - 232. (in Chinese)
[31] XU K,YUEN P,XIE Q,et al. WeedsNet:a dual attention network with RGB - D image for weed detection in natural wheat field [J]. Precision Agriculture,2024,25(1): 460 - 485.
[32] ZHONG Z,LIU X,JIANG J,et al. High-resolution depth maps imaging via attention-based hierarchical multi-modal fusion[J]. IEEE Transactions on Image Processing,2021,31: 648 - 663.
[33] 文含,付忠良,赵莹,等. 基于多模态融合注意力的肝细胞癌疗效预测方法[J]. 计算机应用,2023,43(2):41 - 46.
WEN Han,FU Zhongliang,ZHAO Ying,et al. Prediction method of hepatocellular carcinoma efficacy based on multimodal fusion attention[J]. Journal of Computer Applications,2023,43(2):41 - 46. (in Chinese)
[34] 石岩,任宇琪,王思远,等. 自适应融合气体-光谱双模态信息花生产地溯源方法[J]. 农业机械学报,2024,55(4):176 - 183.
SHI Yan,REN Yuqi,WANG Siyuan, et al. Adaptive fusion of gas spectral bimodal information for peanut origin traceability [J]. Transactions of the Chinese Society for Agricultural Machinery,2024,55(4):176 - 183. (in Chinese)
[35] 胡学龙,朱文韬,杨信廷,等. 基于水质-声音-视觉融合的循环水养殖鱼类摄食强度识别[J]. 农业工程学报,2023,39(10):141 - 150.
HU Xuelong,ZHU Wentao,YANG Xinting, et al. Identification of feeding intensity in recirculating aquaculture fish using water quality - sound - vision fusion[J]. Transactions of the CSAE,2023,39(10):141 - 150. (in Chinese)
[36] LOU Z,QUAN L,SUN D,et al. Multimodal deep fusion model based on transformer and multi-layer residuals for assessing the competitiveness of weeds in farmland ecosystems[J]. International Journal of Applied Earth Observation and Geoinformation, 2024,127:103681.
[37] LIN Z, FU R, REN G,et al. Automatic monitoring of lettuce fresh weight by multi-modal fusion based deep learning[J]. Frontiers in Plant Science,2022,13:980581.
[38] CHEN P,WANG F. Effect of crop spectra purification on plant nitrogen concentration estimations performed using high-spatial-resolution images obtained with unmanned aerial vehicles[J]. Field Crops Research,2022, 288:108708.
[39] 张竞成,袁琳,王纪华,等. 作物病虫害遥感监测研究进展[J]. 农业工程学报,2012,28(20): 1 - 11.
ZHANG Jingcheng,YUAN Lin,WANG Jihua,et al. Research progress of crop diseases and pests monitoring based on remote sensing[J]. Transactions of the CSAE,2012,28(20):1 - 11. (in Chinese)
[40] 贾少鹏,高红菊,杭潇,等. 基于深度学习的农作物病虫害图像识别技术研究进展[J]. 农业机械学报,2019,50(增刊):313 - 317.
JIA Shaopeng,GAO Hongju,HANG Xiao,et al. Research progress on image recognition technology of crop pests and diseases based on deep learning[J]. Transactions of the Chinese Society for Agricultural Machinery,2019,50(Supp.):313 - 317. (in Chinese)
[41] 韩文霆,李敏,陈微. 作物数字图像获取与长势诊断的方法研究[J]. 农机化研究,2012,34(6):1 - 6.
HAN Wenting,LI Min,CHEN Wei. Methods of image acquisition and analysis for crop conditions diagnose[J]. Journal of Agricultural Mechanization Research,2012,34(6):1 - 6. (in Chinese)
[42] 刘爽,谭鑫,刘成玉,等. 高光谱数据处理算法的小麦赤霉病籽粒识别[J]. 光谱学与光谱分 析,2019,39(11):3540 - 3546.
LIU Shuang,TAN Xin,LIU Chengyu,et al. Recognition of fusarium head blight wheat grain based on hyperspectral data processing algorithm[J]. Spectroscopy and Spectral Analysis,2019,39(11):3540 - 3546. (in Chinese)
[43] 王楠,李震,李佳盟,等. 融合多光谱成像与深度学习的作物植株叶绿素检测系统研究[J]. 农业机械学报,2023,54(2):260 - 269.
WANG Nan,LI Zhen,LI Jiameng,et al. Fusing multispectral imaging and deep learning in plant chlorophyll index detection system[J]. Transactions of the Chinese Society for Agricultural Machinery,2023,54(2):260 - 269. (in Chinese)
[44] 郭交,朱琳,靳标. 基于 Sentinel - 1 和 Sentinel - 2 数据融合的农作物分类[J]. 农业机械学报,2018,49(4):192 - 198.
GUO Jiao, ZHU Lin,JIN Biao. Crop classification based on data fusion of Sentinel - 1 and Sentinel - 2[J]. Transactions of the Chinese Society for Agricultural Machinery,2018,49(4):192 - 198. (in Chinese)
[45] 马晓丹,刘梦,关海鸥,等. 基于热红外图像处理技术的农作物冠层识别方法研究[J]. 光谱学与光谱分析,2021,41(1):216 - 222.
MA Xiaodan,LIU Meng,GUAN Haiou,et al. Recognition method for crop canopies based on thermal infrared image processing technology[J]. Spectroscopy and Spectral Analysis,2021,41(1):216 - 222. (in Chinese)
[46] 江晓庆,肖德琴,张波,等. 基于 Kinect 的农作物长势深度图像实时获取算法[J]. 广东农业科学,2012,39(23):195 - 199.

JIANG Xiaoqing, XIAO Deqin, ZHANG Bo, et al. Real-time obtaining algorithm for the range image of crop growth based on Kinect[J]. Guangdong Agricultural Sciences, 2012, 39(23): 195 – 199. (in Chinese)
[47] KARMAKAR P, TENG S W, MURSHED M, et al. Crop monitoring by multimodal remote sensing: a review[J]. Remote Sensing Applications: Society and Environment, 2023, 33: 101093.
[48] 张羽丰,杨景,邓寒冰,等. 基于 RGB 和深度双模态的温室番茄图像语义分割模型[J]. 农业工程学报, 2024, 40(2): 295 – 306.
ZHANG Yufeng, YANG Jing, DENG Hanbing, et al. Semantic segmentation model for greenhouse tomato images based on RGB and depth bimodality[J]. Transactions of the CSAE, 2024, 40(2): 295 – 306. (in Chinese)
[49] XU K, ZHANG J, LI H, et al. Spectrum-and RGB – D-based image fusion for the prediction of nitrogen accumulation in wheat [J]. Remote Sensing, 2020, 12(24): 4040.
[50] WANG Y, CHEN Y, WANG D. Recognition of multi-modal fusion images with irregular interference[J]. PeerJ Computer Science, 2022, 8: 1018.
[51] 罗庆,饶元,金秀,等. 基于改进 YOLO v5s 和多模态图像的树上毛桃检测(英文)[J]. 智慧农业(中英文), 2022, 4(4): 84 – 104.
LUO Qing, RAO Yuan, JIN Xiu, et al. Multi-class on-tree peach detection using improved YOLO v5s and multi-modal images [J]. Smart Agriculture, 2022, 4(4): 84 – 104. (in Chinese)
[52] 周宏平,金寿祥,周磊,等. 基于多模态图像的自然环境下油茶果识别[J]. 农业工程学报, 2023, 39(10): 175 – 182.
ZHOU Hongping, JIN Shouxiang, ZHOU Lei, et al. Recognition of *Camellia oleifera* fruits in natural environment using multi-modal images[J]. Transactions of the CSAE, 2023, 39(10): 175 – 182. (in Chinese)
[53] 邓向武,齐龙,马旭,等. 基于多特征融合和深度置信网络的稻田苗期杂草识别[J]. 农业工程学报, 2018, 34(14): 165 – 172.
DENG Xiangwu, QI Long, MA Xu, et al. Recognition of weeds at seedling stage in paddy fields using multi-feature fusion and deep belief networks[J]. Transactions of the CSAE, 2018, 34(14): 165 – 172. (in Chinese)
[54] 冯权泷,任燕,姚晓闯,等. 基于多源光学雷达数据融合的黄淮海平原冬小麦识别[J]. 农业机械学报, 2023, 54(2): 160 – 168.
FENG Quanlong, REN Yan, YAO Xiaochuang, et al. Identification of winter wheat in Huang – Huai – Hai plain based on multi-source optical radar data fusion[J]. Transactions of the Chinese Society for Agricultural Machinery, 2023, 54(2): 160 – 168. (in Chinese)
[55] LEI L, WANG X, ZHONG Y, et al. DOCC: deep one-class crop classification via positive and unlabeled learning for multi-modal satellite imagery[J]. International Journal of Applied Earth Observation and Geoinformation, 2021, 105: 102598.
[56] 岑海燕,朱月明,孙大伟,等. 深度学习在植物表型研究中的应用现状与展望[J]. 农业工程学报, 2020, 36(9): 1 – 16.
CEN Haiyan, ZHU Yueming, SUN Dawei, et al. Current status and future perspective of the application of deep learning in plant phenotype research[J]. Transactions of the CSAE, 2020, 36(9): 1 – 16. (in Chinese)
[57] SUN D, ROBBINS K, MORALES N, et al. Advances in optical phenotyping of cereal crops[J]. Trends in Plant Science, 2022, 27(2): 191 – 208.
[58] 朱逢乐,严霜,孙霖,等. 基于深度学习多源数据融合的生菜表型参数估算方法[J]. 农业工程学报, 2022, 38(9): 195 – 204.
ZHU Fengle, YAN Shuang, SUN Lin, et al. Estimation method of lettuce phenotypic parameters using deep learning multi-source data fusion[J]. Transactions of the CSAE, 2022, 38(9): 195 – 204. (in Chinese)
[59] CHENG Z, GU X, DU Y, et al. Multi-modal fusion and multi-task deep learning for monitoring the growth of film-mulched winter wheat[J]. Precision Agriculture, 2024, 25: 1 – 25.
[60] NIDAMANURI R R, JAYAKUMARI R, RAMIYA A M, et al. High-resolution multispectral imagery and LiDAR point cloud fusion for the discrimination and biophysical characterisation of vegetable crops at different levels of nitrogen[J]. Biosystems Engineering, 2022, 222: 177 – 195.
[61] LU X, LI W, XIAO J, et al. Inversion of leaf area index in citrus trees based on multi-modal data fusion from UAV platform [J]. Remote Sensing, 2023, 15(14): 3523.
[62] 宋成阳,耿洪伟,费帅鹏,等. 基于多源数据的小麦品种产量估测研究[J]. 光谱学与光谱分析, 2023, 43(7): 2210 – 2219.
SONG Chengyang, GENG Hongwei, FEI Shuaipeng, et al. Study on yield estimation of wheat varieties based on multi-source data[J]. Spectroscopy and Spectral Analysis, 2023, 43(7): 2210 – 2219. (in Chinese)
[63] ANUP K P, LIM C, RAMESH P S, et al. Crop yield estimation model for Iowa using remote sensing and surface parameters [J]. International Journal of Applied Earth Observation and Geoinformation, 2006, 8(1): 26 – 33.
[64] MAIMAITIJIANG M, SAGAN V, SIDIKE P, et al. Soybean yield prediction from UAV using multimodal data fusion and deep learning[J]. Remote Sensing of Environment, 2020, 237: 111599.
[65] 李阳,苑严伟,赵博,等. 基于多时相多参数融合的麦玉轮作小麦产量估算方法[J]. 农业机械学报, 2023, 54(12): 186 – 196.
LI Yang, YUAN Yanwei, ZHAO Bo, et al. Estimation of wheat yield in wheat-maize rotation based on multi-temporal and multi-parameter fusion[J]. Transactions of the Chinese Society for Agricultural Machinery, 2023, 54(12): 186 – 196. (in Chinese)
[66] 张少华,段剑钊,贺利,等. 基于无人机平台多模态数据融合的小麦产量估算研究[J]. 作物学报, 2022, 48(7): 1746 – 1760.
ZHANG Shaohua, DUAN Jianzhao, HE Li, et al. Wheat yield estimation from UAV platform based on multi-modal remote sensing data fusion[J]. Acta Agronomica Sinica, 2022, 48(7): 1746 – 1760. (in Chinese)
[67] LIN F, CRAWFORD S, GUILLOT K, et al. MMST – ViT: climate change-aware crop yield prediction via multi-modal spatial-

temporal vision transformer[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision,2023:5774 - 5784.
[68] 赵法川,徐晓辉,宋涛,等. 融合多头注意力的轻量级作物病虫害识别方法[J]. 华南农业大报,2023,44(6):1 - 16.
ZHAO Fachuan,XU Xiaohui,SONG Tao,et al. A lightweight crop pest identification method based on multi-head attention[J]. Journal of South China Agricultural University,2023,44(6): 1 - 16. (in Chinese)
[69] 廖娟,陶婉琰,臧英,等. 农作物病虫害遥感监测关键技术研究进展与展望[J]. 农业机械学报,2023,54(11):1 - 19.
LIAO Juan,TAO Wanyan,ZANG Ying, et al. Research progress and prospect of key technologies in crop disease and insect pest monitoring[J]. Transactions of the Chinese Society for Agricultural Machinery,2023,54(11):1 - 19. (in Chinese)
[70] 刘青会. 遥感技术在豆类作物病虫害防治中的应用[J]. 农业工程技术,2022,42(21):27 - 28.
LIU Qinghui. Application of remote sensing technology in the control of pests and diseases of leguminous crops[J]. Agricultural Engineering Technology,2022,42(21):27 - 28. (in Chinese)
[71] JIANG X, ZHEN J, MIAO J, et al. Assessing mangrove leaf traits under different pest and disease severity with hyperspectral imaging spectroscopy[J]. Ecological Indicators,2021,129: 107901.
[72] RAZA S A,PRINCE G,CLARKSON J P,et al. Automatic detection of diseased tomato plants using thermal and stereo visible light images[J]. PloS One,2015,10(4):e0123262.
[73] PATIL R R,KUMAR S. Rice-fusion:a multimodality data fusion framework for rice disease diagnosis[J]. IEEE Access,2022,10:5207 - 5222.
[74] 张净,邵文文,刘晓梅,等. 基于超图的双模态特征融合的作物病害识别算法[J]. 江苏农业科学,2023,51(15):164 - 173.
ZHANG Jing, SHAO Wenwen, LIU Xiaomei, et al. Crop disease identification based on bimodal feature fusion and HGNN [J]. Jiangsu Agricultural Sciences,2023, 51(15):164 - 173. (in Chinese)
[75] CAO Y,CHEN L,YUAN Y,et al. Cucumber disease recognition with small samples using image-text-label-based multi-modal language model[J]. Computers and Electronics in Agriculture,2023,211:107993.
[76] ZHANG J,HUANG Y,PU R,et al. Monitoring plant diseases and pests through remote sensing technology:a review[J]. Computers and Electronics in Agriculture,2019,165: 104943.
[77] GUAN H,FU C,ZHANG G,et al. A lightweight model for efficient identification of plant diseases and pests based on deep learning[J]. Frontiers in Plant Science,2023,14:1227011.
[78] LU J,WU Z,LAN Y,et al. Study on the prediction model of litchi downy blight damage based on IoT and hyperspectral data fusion[J]. IEEE Internet of Things Journal,2024,16(11):27184 - 27200.
[79] ZHANG J,HUANG Y,YUAN L,et al. Using satellite multispectral imagery for damage mapping of armyworm (*Spodoptera frugiperda*) in maize at a regional scale[J]. Pest Management Science,2016,72(2):335 - 348.
[80] 周一帆,刘东洋,周宇平. 基于多模态特征对齐的作物病害叶片检测[J]. 中国农机化学报, 2024,45(7):180 - 187.
ZHOU Yifan,LIU Dongyang,ZHOU Yuping. Detection of crop disease leaf based on multi-modal feature alignment[J]. Journal of Chinese Agricultural Mechanization,2024, 45(7):180 - 187. (in Chinese)
[81] 王春山,赵春江,吴华瑞,等. 采用双模态联合表征学习方法识别作物病害[J]. 农业工程学报,2021,37(11):180 - 188.
WANG Chunshan,ZHAO Chunjiang,WU Huarui,et al. Recognizing crop diseases using bimodal joint representation learning [J]. Transactions of the CSAE,2021,37(11):180 - 188. (in Chinese)
[82] 刘立波,赵斐斐. 融合注意力机制的枸杞虫害图文跨模态检索方法[J]. 农业机械学报,2022,53(2):299 - 308.
LIU Libo, ZHAO Feifei. Cross-modal image and text retrieval method for *Lycium barbarum* pests by integrating attention mechanism[J]. Transactions of the Chinese Society for Agricultural Machinery,2022,53(2):299 - 308. (in Chinese)
[83] 付虹雨,王薇,卢建宁,等. 基于无人机多光谱的耐旱苎麻品种筛选方法[J]. 农业机械学报,2023,54(4):206 - 213.
FU Hongyu,WANG Wei,LU Jianning,et al. Screening of drought-tolerant ramie based on UAV multispectral imagery[J]. Transactions of the Chinese Society for Agricultural Machinery,2023,54(4):206 - 213. (in Chinese)
[84] 李美清,李晋阳,毛罕平. 基于光谱特征和生理特征的番茄磷营养诊断方法[J]. 农业机械学报,2016,47(3):286 - 291.
LI Meiqing, LI Jinyang, MAO Hanping. Tomatoes phosphorus nutrition diagnosis based on spectral and physiological characteristics[J]. Transactions of the Chinese Society for Agricultural Machinery,2016,47(3): 286 - 291. (in Chinese)
[85] WANG L,MIAO Y,HAN Y,et al. Extraction of 3D distribution of potato plant CWSI based on thermal infrared image and binocular stereovision system[J]. Frontiers in Plant Science, 2023,13:1104390.
[86] YAO J,WU Y, LIU J,et al. Multimodal deep learning-based drought monitoring research for winter wheat during critical growth stages[J]. PLoS One,2024,19(5):e0300746.
[87] KANEDA Y,SHIBATA S,MINENO H. Multi-modal sliding window-based support vector regression for predicting plant water stress[J]. Knowledge-Based Systems,2017,134: 135 - 148.
[88] WAKAMORI K,MIZUNO R,NAKANISHI G, et al. Multimodal neural network with clustering-based drop for estimating plant water stress[J]. Computers and Electronics in Agriculture,2020,168:105118.
[89] QIN S, DING Y, ZHOU T, et al. "Image-Spectral" fusion monitoring of small cotton samples nitrogen content based on improved deep forest[J]. Computers and Electronics in Agriculture,2024,221:109002.
[90] GAO Z, LUO N, YANG B,et al. Estimating leaf nitrogen content in wheat using multimodal features extracted from canopy spectra[J]. Agronomy,2022,12(8):1915.
[91] ELSHERBINY O,ZHOU L,HE Y,et al. A novel hybrid deep network for diagnosing water status in wheat crop using IoT-based multimodal data[J]. Computers and Electronics in Agriculture,2022,203:107453.