

# 多模态数据融合在计算机智能检测中的特征提取方法研究

张俊

杭州旭辐检测技术有限公司, 浙江杭州, 310022;

**摘要:** 随着人工智能技术的快速发展和传感设备的多样化, 计算机智能检测系统对数据处理能力提出了更高要求。传统单一模态数据在面对复杂环境与多维信息时存在信息缺失、特征表达有限等问题, 难以支撑高精度检测任务。因此, 多模态数据融合技术逐渐成为研究热点, 尤其在特征提取阶段的优化对于提升智能检测性能至关重要。本文聚焦于多模态数据融合在智能检测中的特征提取策略, 从模态间协同编码、跨模态对齐机制及深度融合网络设计三个方面展开系统研究, 分析不同模态数据(如图像、文本、语音、点云等)在时空尺度、结构特征与语义层面的互补性, 提出一套可泛化、可迁移的多模态特征提取方法体系。研究表明, 通过有效建模模态间的相关性与差异性, 可以显著提升检测模型的鲁棒性与泛化能力, 为多源信息融合场景下的智能检测系统提供理论支持与实践路径。

**关键词:** 多模态融合; 智能检测; 特征提取; 跨模态对齐; 协同学习; 深度网络

**DOI:** 10. 64216/3080-1508. 25. 02. 019

## 1 多模态特征融合的基本原理与挑战

### 1.1 多模态数据特征的异质性分析

多模态数据最为突出的特点就在于其来源相当多样, 并且结构呈现出异质性, 不同模态的数据在维度分布、表达方式、语义层级以及时间序列属性等诸多方面都有着明显的差异存在。就比如说, 图像数据一般展现出来的是二维空间结构的形态, 它带有十分直观的纹理以及空间方面的特征; 而语音, 它实则是一种时间序列信号, 重点在于其频率成分以及变化的趋势; 至于文本数据, 它则属于离散符号序列这一类别, 其信息密度颇高, 只是在感知的直观性方面有所欠缺。这样的模态间的异质性状况, 使得要达成统一的特征表达以及共享编码这件事变得颇为复杂起来。特别是在特征提取的这个阶段当中, 不同模态所需要的网络结构以及处理机制彼此之间存在着不小的差异, 是没办法直接去套用单一的策略的。所以, 研究者务必要从模态间的相关性这个角度切入进去, 构建起那种具有适应性的特征表示的方法, 进而实现结构的匹配、语义的对齐以及分布的重建等操作, 最终实现在异质空间当中提取出具备共识性的融合特征这样的目标。

### 1.2 模态间互补性与冗余信息的平衡

在多模态系统当中, 各模态的数据之间常常存在着颇为显著的互补关系。就好比在夜间或者存在遮挡的环境之下, RGB 图像的质量会出现下降的情况, 这个时候, 红外图像或者深度图像便能够对其不足之处予以弥补。再比如在进行语义识别的时候, 图像是可

以提供出上下文相关的视觉方面的信息的, 而文本, 则能够在增强语义理解的准确性上发挥作用。不过, 模态之间的这种互补性也并非一直都是起到正向增强的效果, 也有可能存在着冗余的情况, 甚至还会出现信息相互冲突的状况, 如此一来便会对模型的判断产生影响。所以说, 在特征提取的这个过程之中, 怎样去识别并且融合那些有价值的信息, 同时又能够对冗余的或者不确定的信息加以抑制, 这可就是设计出有效融合机制的关键所在了。当前经常会用到的一些策略包含有通过注意力机制来加以引导的特征筛选的方式、多尺度融合结构以及模态选择门控机制等。借助于对不同模态的贡献度进行动态调整, 进而实现特征维度以及语义层次的自适应融合, 这样便可以提高整体特征表达所具有的鲁棒性以及判别力了。

### 1.3 多模态融合中的对齐问题与信息缺失

多模态特征融合得以开展的一个前提条件在于各模态数据要实现时空对齐。可在实际的应用当中, 模态之间常常会出现不少问题, 比如采样频率不一样、数据延迟没办法同步以及存在缺失模态等情况, 这就致使对齐工作面临困难, 而且信息也变得不完整。特别是在动态场景或者是非结构化环境之中, 某些模态说不定会暂时失去效用, 甚至是完全缺失掉, 如此一来, 融合模型就没办法依靠完整的信息去开展推理判断的工作了。那么, 在存在缺失情况或者是对齐做得不够好的条件下, 怎样才能让融合特征一直保持有效性? 这已然成为当下研究领域里的一个重要方向了。针对这一情况, 其解决策略包含了好几个方面, 像采用那种基于注意力或者是图神经网络的跨模态对齐

机制,借助上下文信息来对缺失模态的表示进行重新构造,还有通过模态补全或者是推理机制去构建出一个较为鲁棒的融合表示空间等。这些方法在实际应用中确实有效地提升了特征提取工作的健壮程度,使得整个系统就算是处在部分信息缺失这样的状况下,依旧能够维持比较高的检测精度以及稳定性。

## 2 多模态特征提取的主流方法与进展

### 2.1 基于共享编码器的特征抽象机制

共享编码器结构属于多模态融合里极为基础的一类特征提取手段。其核心的想法是借助统一的编码架构,针对不同模态的输入来开展映射操作,进而生成能够在同一个潜在空间当中进行比较的向量表示形式。这类方法往往在前面的部分先对各个模态做预处理以及初级的特征提取工作,接着再利用共享的网络架构,像是 Transformer、ResNet 或者 AutoEncoder 这些,去对其实施特征的抽象处理以及统一的编码流程。在展开训练的这个过程当中,通过引入跨模态对比损失又或者是共享注意力机制的方式,来强化不同模态之间语义方面的耦合程度,这样就能获取到更加紧凑并且在语义上保持一致的特征向量了。此类方法所具备的优势就在于它的结构并不复杂,训练起来也比较稳定,而且还适用于那些模态之间存在一定同质性的具体场景,在图文检索、视觉问答、视频语义识别等诸多任务当中都得到了广泛的运用。不过,在异质模态融合以及信息缺失处理这两个方面,它依然还是存在着一定的局限性的。

### 2.2 基于跨模态对齐的特征匹配策略

当不同模态的数据存在较强异构性的情况时,基于对齐的特征提取方法往往会表现得更为有效。这类方法一般会专门去设计跨模态对齐模块,借助注意力机制、图匹配算法或者判别网络等手段,来对各个模态数据的特征分布状况、语义关系情况以及时序结构情形进行对齐匹配方面的操作。就好比在图像和文本融合的时候,交叉注意力模型能够达成图像区域与文本单词之间那种精细的语义对齐效果,进而提升多模态特征所具有的一致性以及互补性;而在点云与图像融合的情境下,则可以通过构建模态图结构的方式,实现空间结构和纹理细节的融合补充这一目的。这些对齐机制一方面提升了特征的可比性,另一方面也增强了系统在多模态复杂数据场景当中的适应能力,它们已然成为了当前多模态检测系统中实现性能提升的关键所在。

### 2.3 基于协同学习的深度融合网络

协同学习属于一种凭借多网络联合优化来开展多模态特征提取的策略,其会构建多个专属于模态的网络以及共享判别模块,在维持模态独立性的同时,促使特征表达朝着互补空间逐步靠近。像 Dual-Encoder 模型、Co-Attention 机制以及多模态 Transformer 架构等都属于典型结构,它们借助引入模态协同损失、自监督对比机制以及互信息最大化目标的方式,达成多模态特征彼此间的互动以及增强效果。而且,有部分研究还进一步提出了多层协同结构,能够在不同的抽象层级上搭建起模态之间相互交流的桥梁,让低层次的感知特征与高层次的语义信息一同实现协同优化,进而提升整个网络的表达水平。协同学习机制一方面提高了融合的效率,另一方面也明显增强了模型针对模态缺失以及非一致分布情况的适应能力,已然成为当前多模态特征提取领域中极为重要的发展方向之一。

## 3 融合特征在检测系统中的实际应用与优化

### 3.1 多模态融合在目标检测中的实践案例

多模态融合于目标检测领域的应用已然彰显出颇为可观的实际价值,特别是在复杂环境之中,相较于单一模态的检测系统,其呈现出了更高的准确程度以及更强的鲁棒特性。就拿自动驾驶来举例说明,车辆的感知系统往往会把摄像头所获取的图像、激光雷达的点云以及毫米波雷达的数据融合到一起,凭借着不同模态之间所具备的互补性质,进而达成全天候、全角度开展目标检测这项任务的目的。在这样的系统里面,图像数据能够给予纹理还有颜色方面的相关信息,点云数据,则可以提供极为精准的三维结构以及空间深度方面的信息,而毫米波雷达即便是在恶劣的天气状况之下,依旧拥有较为稳定的穿透能力。研究人员一般会采用多通道并行网络去处理各种各样的模态信息,并且借助中间层融合机制从中提取出对于目标识别而言最具判别性的混合特征。

例如,就 KITTI 数据集的多模态检测模型来讲,引入了点云图像联合特征的那种模型,在检测精度这块明显要比单模态模型表现得更好,尤其在针对小目标以及有遮挡目标的检测场景当中,其优势就更为突出了。这样的一种应用实践,一方面验证了融合特征提取方法是行之有效的,另一方面也给多模态技术在工业级检测系统当中得以落地实施打下了现实层面的基础。

### 3.2 融合模型的计算效率优化策略

虽说多模态特征融合在很大程度上提升了检测

的精度,可它随之带来的计算成本以及模型复杂度方面的问题,是绝对不能被忽视掉的。多模态系统通常会涉及到多个输入的通路,还有多个编码的网络,再加上其有着颇为复杂的融合结构,这就致使整个模型在进行推理的时候,会出现资源占用量偏高、延迟比较大的情况,如此一来,就很难去满足边缘设备或者移动平台对于实时检测所提出的需求。所以说,针对融合模型展开的计算优化策略,已然成为了系统设计环节当中极为重要的一个部分。当下主要的优化方向涵盖了:对网络架构进行轻量化的设计、开展模型压缩与剪枝方面的工作、制定模态选择的策略以及确立分层融合的机制等等。

在这之中,轻量化设计能够通过引入像 MobileNet、ShuffleNet 这类轻型结构来替换重型编码网络。模型压缩,是借助知识蒸馏、参数剪枝等手段去削减冗余计算。至于模态选择策略,它可以让系统依据环境条件,动态地激活或者关闭某些模态通路,以此达成计算资源依照需求来进行分配的目的。另外,部分研究提出把融合操作挪到模型中后期的高抽象层,以此来实现“后融合”策略,进而降低早期卷积计算时出现的模态冗余情况。这些优化策略一方面保障了融合模型在资源受限的场景下能够有较好的运行效率,另一方面也给它的广泛部署开拓出了可行的路径。

### 3.3 面向实际场景的融合系统鲁棒性增强

在实际的应用场景当中,多模态融合系统不得不去面对各式各样难以把控的因素,像模态输入出现缺失、传感器存在误差、受到外部的干扰以及模态质量参差不齐等诸多问题,而这无疑对特征提取机制的鲁棒性提出了更为严苛的要求。从一个方面来讲,模态的缺失极有可能致使特征空间变得不完整,进而对模型决策的稳定性产生影响;从另一个方面来看,在特定的一些条件之下,某些模态或许会引入噪声,甚至还可能会引入错误信息,从而引发误检或者漏检的情况出现。所以说,强化融合模型在异常状况下的稳定性以及容错能力,已然成为了极为关键的一项研究目标。当下的主流方法涵盖了基于注意力机制展开的模态质量评估、具备自适应特性的加权融合机制、着眼于模态鲁棒性的训练策略以及模态恢复模块等方面。

例如,在模型当中引入模态注意力机制的话,那么模型就能够实时地去评估每一种模态输入所具有的可信度,并且还可以动态地对其特征贡献权重做出

相应调整。在鲁棒性训练环节引入模态缺失模拟训练样本,如此一来,模型在训练阶段便拥有了应对不完整输入状况的能力。除此之外,有一些系统还专门设计了模态重构模块,一旦某一模态出现失效的情况,就能够借助其他模态的上下文信息来开展补偿或者重建方面的工作,进而维持住系统整体的稳定性。这些机制在很大程度上提升了多模态融合模型于非理想场景之下应用的可靠性,给智能检测系统在复杂环境里的稳定运行给予了相当坚实的保障。

## 4 结语

随着人工智能不断向实际应用领域延伸,如何从海量异构的多模态数据中高效提取特征,已成为智能检测系统面临的核心问题之一。本文围绕多模态数据融合的特征提取展开深入探讨,从异质性建模、对齐机制设计到协同学习网络构建,再到融合特征在目标检测中的落地应用与优化策略,系统梳理了多模态特征提取的关键路径与发展趋势。

研究表明,融合特征的有效建模不仅可以显著提升检测系统在复杂环境下的表现,也为处理信息缺失、模态不一致等现实问题提供了可行解决方案。尽管当前在模型效率、实时性与可扩展性方面仍存在挑战,但随着深度学习结构的演进与硬件计算力的提升,多模态融合技术将在智能安防、自动驾驶、智慧医疗、工业监测等领域发挥越来越重要的作用。未来的研究可以进一步从跨模态知识迁移、自监督多模态表示学习、模态间因果关系建模等方向展开,推动多模态特征提取迈向更高维度的智能演化阶段。

## 参考文献

- [1] 王梓衡,沈继锋,左欣,等. 基于特征级与决策级融合的农作物叶片病害识别[J]. 江苏大学学报(自然科学版). 2024,45(3). DOI:10.3969/j.issn.1671-7775.2024.03.006.
- [2] 宁大海,郑晟. 可见光和红外图像决策级融合目标检测算法[J]. 红外技术. 2023,45(3).
- [3] 陈辉,王硕,许家昌,等. 基于多尺度特征融合生成对抗网络的水下图像增强[J]. 计算机工程与应用. 2023,59(21). DOI:10.3778/j.issn.1002-8331.2304-0094.
- [4] 任泽裕,王振超,柯尊旺,等. 多模态数据融合综述[J]. 计算机工程与应用. 2021,(18). DOI:10.3778/j.issn.1002-8331.2104-0237.