ICS 35.240.60
CCS L 70

T/CS

团 体 标 准

T/CS 617—2026

多模态融合 感知数据质量评价指南

Multimodal fusion—Guidelines for quality evaluation of perception data

2026－08－28 发布 2026－08－28 实施

中国商品学会 发 布

目　　次

前言 II
引言 III
1 范围 1
2 规范性引用文件 1
3 术语和定义 1
4 总体原则 1
4.1 科学性原则 1
4.2 系统性原则 1
4.3 可操作性原则 1
4.4 可裁剪性原则 1
4.5 可扩展性原则 1
4.6 持续改进原则 2
5 多模态感知数据质量评价框架 2
5.1 评价框架概述 2
5.2 评价维度分类 2
5.3 层级与维度的矩阵关系 2
6 评价指标体系 3
6.1 通用数据质量指标 3
6.2 多模态特有质量指标 3
6.3 融合过程质量指标 3
6.4 应用适配质量指标 4
7 评价方法 4
7.1 评价流程 4
7.2 指标量化与权重确定 5
7.3 综合评价模型 5
7.4 综合评价等级 5
8 评价实施指南 6
8.1 评价准备 6
8.2 数据采集与预处理 6
8.3 评价执行 7
8.4 结果分析与改进建议 7
附录 A（规范性） 评价指标计算示例 8
附录 B（资料性） 典型应用场景评价要点 9
参考文献 10

前　言

本文件按照GB/T 1.1—2020《标准化工作导则　第1部分：标准化文件的结构和起草规则》的规定起草。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别专利的责任。

本文件由中国商品学会提出并归口。

本文件起草单位：广东智视数字技术有限公司、广东技术师范大学、广东工业大学、哈尔滨工业大学(深圳)(哈尔滨工业大学深圳科技创新研究院)、中山大学。

本文件主要起草人：房小兆、韩娜、王书贤、孙为军、文杰、黄超。

引　言

多模态融合感知系统通过整合来自不同类型传感器（如相机、激光雷达、毫米波雷达等）的感知数据，实现对环境更全面、更可靠的理解。这类系统在自动驾驶、智能机器人、工业检测等领域已得到广泛应用。

感知数据的质量直接决定了多模态融合系统的性能和可靠性。与单一模态数据的质量评价不同，多模态融合场景中的数据质量涉及模态间时空对齐、跨模态一致性、融合过程完整性等特有维度，需要建立专门的质量评价框架和方法。

本文件旨在为多模态融合感知数据的质量评价提供一套系统化、可操作的方法和指南，推动多模态感知技术的规范化发展。

多模态融合 感知数据质量评价指南

1 范围

本文件确立了多模态融合感知数据质量评价的总体原则与评价框架，给出了评价指标体系、评价方法及评价实施的指导。

本文件适用于多模态融合感知系统的数据质量评价，不适用于不涉及多模态融合过程的单一模态感知数据质量评价。

2 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中，注日期的引用文件，仅该日期对应的版本适用于本文件；不注日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

GB/T 36344-2018 信息技术 数据质量评价指标

GB/T 41867-2022 信息技术 人工智能 术语

3 术语和定义

GB/T 36344-2018、GB/T 41867-2022界定的术语和定义适用于本文件。

4 总体原则

4.1 科学性原则

评价指标宜基于明确的理论或工程依据，定义清晰、可量化、可验证，评价方法宜确保结果可复现。

4.2 系统性原则

评价宜覆盖数据采集、预处理、融合到输出的全生命周期，各维度相互独立、边界清晰，共同构成完整框架。

4.3 可操作性原则

指标宜具备明确的计算或评定方法，所需数据可获取、可度量，评价过程的时间与资源成本控制在合理范围。

4.4 可裁剪性原则

使用者宜结合自身应用需求和技术条件，合理选择适用的维度与指标，并可补充领域特有内容。

4.5 可扩展性原则

框架宜预留对新模态、新传感器、新技术的接入空间，指标体系可在不改变框架结构的前提下灵活增删。

4.6 持续改进原则

质量评价宜形成“评价—改进—再评价”的闭环，评价结果宜反馈至采集策略、传感器配置、融合算法等环节，驱动持续优化。

5 评价框架

5.1 评价框架概述

评价按数据流动的递进过程，自下而上划分为四个层级。低层级质量为高层级的基础，评价时宜逐层递进，低层异常时宜优先排查：

a) 原始数据层：评价传感器直接输出的原始信号质量。关注信号完整性、时间戳精度、传感器标定参数的有效性等基础物理层指标。
b) 单模态处理层：评价各模态独立完成预处理和特征提取后的中间结果质量。关注单模态的目标检测/分割精度、特征有效性及单模态置信度。
c) 多模态融合层：评价融合算法将多源数据整合后的输出质量。关注融合过程特有的维度，包括时空对齐精度、跨模态一致性、信息增益度、冲突消解度等。
d) 应用效果层：评价融合感知结果在具体业务任务中的实际表现。关注任务完成率、决策准确率、端到端延迟及场景适应性等应用级指标。

5.2 评价维度分类

从质量属性的角度，将所有评价指标划分为四类。每一类维度均有其适用的评价层级：

a) 通用数据质量类：继承 GB/T 36344-2018 的通用维度（规范性、完整性、准确性、一致性、时效性、可访问性），适用于全部四个层级。
b) 多模态特有质量类：针对多传感器系统特有的跨模态问题（时空对齐度、模态互补性、传感器冗余度、覆盖完整性），主要适用于原始数据层、单模态处理层和融合层。
c) 融合过程质量类：聚焦融合算法自身引入的质量问题（融合完整性、信息增益度、冲突消解度、融合延迟），专门适用于融合层。
d) 应用适配质量类：从应用任务需求出发（任务相关性、场景适应性、实时性、安全冗余度），专门适用于应用效果层。

5.3 层级与维度的矩阵关系

5.1的四层与5.2的四类交叉对应如下：

表 1 矩阵关系

层级与维度	原始数据层	单模态处理层	多模态融合层	应用效果层
通用数据质量类	● 适用	● 适用	● 适用	● 适用
多模态特有质量类	● 适用	● 适用	● 适用	—
融合过程质量类	—	—	● 适用	—
应用适配质量类	—	—	—	● 适用
注： “● 适用”表示该维度类别在该层级有对应的具体指标；“—”表示通常不涉及。				

6 评价指标体系

6.1 通用数据质量指标

通用数据质量指标参照 GB/T 36344—2018 的六个维度执行。各维度在多模态场景下的评价要点如下：

a) 规范性：各模态数据的文件格式、编码方式及字节序宜保持一致；传感器标定文件与时间同步配置文件的格式宜符合相应规范。
b) 完整性：宜重点评价各模态数据的缺失率（含帧丢失率、点云空洞率、信号中断率等）以及多模态数据在时间轴上的覆盖完整性。
c) 准确性：宜重点评价目标检测的定位精度、目标分类正确率，以及传感器测量值与真值之间的偏差。
d) 一致性：宜重点评价同一时刻不同模态对同一目标的感知结果一致性、时序数据中的目标追踪稳定性及语义一致性。
e) 时效性：宜重点评价端到端感知延迟、各模态数据间的同步延迟，以及数据更新频率与任务需求的匹配程度。
f) 可访问性：宜重点评价多模态数据的索引与检索效率，以及数据传输的带宽和可靠性。

注：-各指标的具体计算方法和评价准则见 GB/T 36344—2018 第5章至第7章。

6.2 多模态特有质量指标

6.2.1 时空对齐度

6.2.1.1 衡量各模态数据在时间和空间维度上的同步与对准程度，是多模态融合的基础性核心指标。

6.2.1.2 各模态数据在时间和空间维度上的同步与对准程度包括以下两个方面：

a) 时间同步精度：指各传感器数据帧采集时刻的一致性。评价时宜统计各传感器帧时间戳与参考时间的偏差分布及漂移情况。安全关键应用中，时间同步精度宜优于 1 ms。
b) 空间标定精度：指各传感器坐标系间转换关系的准确程度。评价时宜分析跨传感器目标投影至统一坐标系后的重投影误差，以及标定参数在温度、振动等环境变化下的稳定性。平均重投影误差≤2 像素可判定标定质量良好。

6.2.2 模态互补性

6.2.2.1 衡量多传感器系统在功能和性能上相互补充增益的程度。

6.2.2.2 评价时宜分析各模态在探测距离、视场角、分辨率及环境适应性等方面的组合覆盖能力，识别覆盖盲区与薄弱环节；同时评估单一模态失效时其他模态提供有效补充感知的能力，以及关键信息的冗余备份程度。

6.2.3 传感器冗余度

6.2.3.1 衡量多传感器系统中关键感知功能的冗余配置程度。

6.2.3.2 安全关键应用中，应评价同构冗余、异构冗余，以及部分传感器失效后系统维持安全运行的降级可用性。

6.2.4 覆盖完整性

6.2.4.1 衡量多传感器组合对目标区域的感知覆盖程度。

6.2.4.2 评价时宜计算有效探测范围对关注区域的覆盖比例，识别和量化感知盲区，并重点评估安全关键区域的冗余覆盖情况。

6.3 融合过程质量指标

6.3.1 融合完整性

6.3.1.1　衡量融合过程中多源有效信息被正确保留与利用的程度。

6.3.1.2　评价时宜计算信息保留率、误剔除率以及虚假融合率。融合完整性按下式计算：

$$融合完整性=\frac{各模态有效信息项数之和}{各模态有效信息项数之和-合理跨模态重复计数} \quad (1)$$

6.3.2　信息增益度

6.3.2.1　衡量引入多模态融合后相较于最优单模态系统的感知性能提升幅度。

6.3.2.2　评价时宜在相同测试集上对比融合系统与最佳单模态的精度增益、召回增益、鲁棒增益及覆盖增益。若增益为负值，宜重新审视融合策略的有效性。

6.3.3　冲突消解度

6.3.3.1　衡量融合系统正确识别并消解模态间矛盾感知结果的能力。

6.3.3.2　评价时宜构建包含已知冲突场景的测试数据集，计算冲突检出率和冲突消解准确率，并评估消解策略在不同场景下的合理性。

6.3.4　融合延迟

6.3.4.1　衡量从各模态数据到达融合模块至融合结果输出的时间间隔。

6.3.4.2　评价时宜计算标准测试条件下的平均延迟、极端场景下的最大延迟及延迟抖动（方差或标准差）。安全关键应用的融合延迟宜不超过 50 ms。

6.4　应用适配质量指标

6.4.1　任务相关性

6.4.1.1　衡量感知数据对特定应用任务的适配程度。

6.4.1.2　评价时宜评估感知输出信息类型与任务需求类型的匹配度、感知精度对任务精度要求的满足程度，以及感知输出中对任务决策有实际价值的信息占比。

6.4.2　场景适应性

6.4.2.1　衡量感知数据质量在不同场景下的稳定性和泛化能力。

6.4.2.2　评价时宜开展典型场景集的覆盖测试以分析质量波动，评估训练场景与测试场景存在分布差异时的性能下降幅度，以及系统对新场景的适应能力。

6.4.3　实时性

6.4.3.1　衡量从传感器曝光/采样时刻到融合结果可供下游任务使用的端到端总延迟。

6.4.3.2　评价时宜测量端到端延迟、延迟超标率，以及单位感知精度所消耗的计算资源效率。此指标与 6.3.4 融合延迟的区别在于：本指标涵盖数据采集、传输、处理至输出的全过程总延迟。

6.4.4　安全冗余度

6.4.4.1　衡量感知系统设计对安全性需求的保障程度。

6.4.4.2　安全关键应用中，宜评价维持最低安全功能所需的最小传感器集合及其组合、传感器或融合模块失效后系统进入安全状态的最大允许时间，以及感知系统对自身输出置信度的估计准确性。

7　评价方法

7.1　评价流程

多模态感知数据质量评价按以下六个步骤依次执行：

a） 明确评价目标：明确评价目标：确定评价目的、评价范围及目标质量等级。
b） 构建评价指标体系：根据评价目标与应用场景，从第 6 章中选取适用的评价维度和具体指标，设定各指标权重及合格判据。
c） 准备评价数据：收集或生成测试所需的原始传感器数据、真值标注数据、标定参数及同步配置信息。测试集宜覆盖正常、边缘及异常场景。
d） 执行质量评价：按既定指标与计算准则逐项评价，优先采用自动化评价工具以提高效率和可复现性。
e） 综合评价与分析：汇总各维度评分，形成综合质量结论，识别质量短板及改进方向。
f） 结果反馈与改进：将评价结果反馈至数据采集、传感器配置、融合算法设计等上游环节，形成“评价—反馈—改进—再评价”的闭环。

注： 上述步骤中，步骤 a）至 c）为评价准备阶段，步骤 d）至 f）为评价执行与闭环阶段。

7.2 指标量化与权重确定

7.2.1 指标量化方式

根据指标性质，采用以下方式之一：
a） 定量指标：通过公式计算得到连续数值，宜优先采用；
b） 分级指标：划分为若干等级（如 A/B/C/D 级）；
c） 二元指标：判定为“满足”或“不满足”特定要求。

7.2.2 权重确定方法

宜选用以下方法之一或其组合：
a） 德尔菲法：多轮专家咨询达成权重共识；
b） 层次分析法（AHP）：通过两两比较构建判断矩阵；
c） 数据驱动法：基于历史数据的敏感性分析确定权重；
d） 等权重法：当各指标重要性难以区分时采用。

7.3 综合评价模型

7.3.1 线性加权模型

当各评价指标之间相对独立时，宜采用线性加权模型进行综合评价：

$$Q = \Sigma(w_i \times q_i) \quad \cdots\cdots (2)$$

式中：
w_i为第i个指标的权重（满足$\Sigma(w_i=1)$）
q_i为第i个指标的归一化得分。

7.3.2 层次化综合评价

当评价体系包含多个层级时，逐层加权汇总：先计算各维度内指标得分，再计算各层级内维度得分，最后汇总为整体质量得分。

7.3.3 最差指标约束模型

对关键指标设定合格判定阈值，任一关键指标不达标时可直接判定为 D 级；全部达标后，再按线性加权模型评分。

7.4 综合评价等级

综合得分Q与质量等级的对应关系见表 2。

表 2 质量等级划分

等级	得分区间	说明
A级（优秀）	Q≥0.90	各维度均处于较高水平，满足严苛应用要求
B级（良好）	0.75≤Q<0.90	大部分维度表现良好，存在可接受改进空间
C级（合格）	0.60≤Q<0.75	基本维度达标，适用于一般应用场景
D级（待改进）	Q<0.60	存在显著质量缺陷，需系统性改进

8 评价实施

8.1 评价准备

8.1.1 评价团队

评价团队宜由感知系统工程师、数据工程师、融合算法工程师、领域专家及质量评价专家组成，或由具备相应能力的人员承担。

8.1.2 评价方案

评价方案宜至少包括以下内容：

a) 评价目标和范围；
b) 评价指标体系及权重配置；
c) 测试场景与数据需求清单；
d) 所需工具与设备；
e) 时间计划；
f) 评价报告模板。

8.1.3 评价环境

评价环境宜满足以下要求：

a) 计算设备的算力与存储能力宜满足多模态数据处理需求；
b) 软件工具宜经过验证且版本可控；
c) 测试数据宜与训练数据严格隔离，避免评价失真。

8.2 数据采集与预处理

8.2.1 测试数据采集原则

测试数据采集宜遵循以下原则：

a) 场景代表性：数据场景分布宜反映目标应用的实际分布，覆盖正常、边缘及异常场景；
b) 模态完整性：所有参与评价的模态均宜有完整的同步数据记录；
c) 标注准确性：真值标注精度宜高于感知系统目标精度至少一个数量级；
d) 数据可追溯性：宜记录采集时间、地点、环境条件等完整元信息。

8.2.2 数据预处理要求

数据预处理宜包括时间同步校验、空间标定校验、数据清洗及模态对齐等步骤，并对处理过程中发现的异常数据段进行标记。

8.3 评价执行

8.3.1 逐层递进评价

评价宜按照本文件第5章所述的四层模型从低到高逐层递进进行。较低层次的质量问题在较高层次可能被放大，当发现低层质量异常时，宜暂停后续评价并优先排查根本原因。

8.3.2 统计检验与不确定性评估

评价结论宜进行统计显著性检验（如 t 检验、方差分析等），以判断观测差异是否具有统计意义。评价结果宜附不确定性评估，说明评价结论的置信水平及主要不确定性来源。

8.4 结果分析与改进建议

8.4.1 分析维度

评价结果分析宜从以下维度展开：

a) 各质量维度的得分分布分析；
b) 得分显著偏低的维度识别（质量短板识别）；
c) 不同场景下质量得分的波动分析（场景敏感性分析）；
d) 质量退化或漂移趋势识别（时间趋势分析）。

8.4.2 改进建议

根据评价结果宜制定针对性改进建议：

a) 原始数据层短板——优化传感器选型、安装位置或标定方案；
b) 单模态处理层短板——优化感知算法或训练数据；
c) 融合层短板——优化融合策略或冲突消解机制；
d) 应用层短板——重新审视任务需求与感知能力的匹配度。

8.4.3 评价报告

评价报告宜包含以下内容：

a) 评价基本信息；
b) 指标体系及权重说明；
c) 测试数据与场景说明；
d) 各维度评价结果（含原始数据与图表）；
e) 综合评价结论与质量等级；
f) 质量短板与风险提示；
g) 改进建议；
h) 评价局限性说明。

附 录 A
（规范性）
评价指标计算示例

A.1 时空对齐度计算示例

A.1.1 时间同步精度评价示例

某多模态感知系统包含1个可见光相机（30 Hz）、1个激光雷达（10 Hz）和1个毫米波雷达（20 Hz）。以GPS时间为参考时钟，记录各传感器每帧数据的时间戳偏差Δt_i，计算均方根同步误差如下： 。

$$RMSE = \sqrt{(\Sigma(\Delta t_i)^2/n)} \quad (3)$$

示例：相机 RMSE=147 μs，激光雷达 RMSE=350 μs，毫米波雷达 RMSE=242 μs。若场景要求≤1 ms，则该系统时间同步精度达标。

A.1.2 空间标定精度评价示例

在标定板上设置已知空间位置的特征点集，将激光雷达点云中的特征点通过标定矩阵投影到相机图像平面，计算投影点与图像中对应特征点的像素距离（重投影误差）。

示例：20 个特征点，平均重投影误差 1.8 像素，最大 4.2 像素。通常认为平均重投影误差≤2 像素标定质量良好。

A.2 信息增益度计算示例

A.2.1 以自动驾驶3D目标检测任务为例，在相同测试集上对比三种配置的性能：

a) 相机单模态：$\Delta mAP = 0.62$
b) 激光雷达单模态：$\Delta mAP = 0.71$
c) 相机 + 激光雷达融合：$\Delta mAP = 0.79$

A.2.2 相对最佳单模态的信息增益度计算如下：

$$\Delta mAP = \frac{0.79-0.71}{0.71} \times 100\% = 11.3\% \quad (4)$$

结论：融合系统相较于最佳单模态获得显著性能增益，融合策略有效。

A.3 综合评价计算示例

A.3.1 某自动驾驶多模态感知系统综合评价：采用线性加权模型，各指标权重及归一化得分分别为——规范性(0.05,0.92)、完整性(0.10,0.88)、准确性(0.15,0.85)、一致性(0.10,0.78)、时效性(0.10,0.90)、时空对齐度(0.20,0.84)、融合完整性(0.10,0.80)、信息增益度(0.10,0.83)、场景适应性(0.10,0.76)。

A.3.2 综合得分:

$$Q = \sum_{i=1}^{n} w_i \times q_i = = 0.836 \quad (5)$$

质量等级判定为B级（良好）。质量短板为一致性（0.78）和场景适应性（0.76）偏低，宜优先改进。

附 录 B
（资料性）
典型应用场景评价要点

B.1 自动驾驶与智能网联汽车

B.1.1 典型模态组合：可见光相机（多目）+ 激光雷达 + 毫米波雷达（前向+角向）+ 超声波雷达 + GNSS/IMU。

B.1.2 评价重点：

a) 安全性维度权重宜显著高于其他维度；
b) 时空对齐度为核心指标——高速行驶（100 km/h）下，1 ms 时间误差对应约 0.03 m 位移误差；
c) 场景适应性应覆盖晴/阴/雨/雪/雾、日/夜/黄昏、城市/高速/乡村等；
d) 端到端延迟在紧急制动场景宜 ≤ 150 ms。

B.1.3 高优先级指标：准确性、时效性、时空对齐度、安全冗余度、场景适应性。

B.2 智能机器人与具身智能

B.2.1 典型模态组合：深度相机/RGB-D + 激光雷达（2D/3D）+ 触觉传感器 + 力/力矩传感器 + IMU + 关节编码器。

B.2.2 评价重点：

a) 实时性要求高——端到端延迟宜 ≤ 20 ms；
b) 近场感知精度关键——近距离目标位置和姿态估计精度宜优于 1 mm 和 1°；
c) 视觉-触觉-力觉同步精度直接影响操作成功率。

B.2.3 高优先级指标：准确性（近场）、时效性、时空对齐度、冲突消解度。

B.3 工业检测与智能制造

B.3.1 典型模态组合：可见光相机（面阵/线阵）+ 红外热像仪 + 3D结构光/激光轮廓仪 + 超声检测 + X射线/CT。

B.3.2 评价重点：

a) 准确性和一致性权重极高——缺陷检测误检率和漏检率直接关系产品质量；
b) 多模态互补性价值突出——不同模态对不同类型缺陷各有优势；
c) 高速产线时间同步精度宜 ≤ 100 μs。

B.3.3 高优先级指标：准确性、一致性、规范性、融合完整性。

B.4 场景裁剪与适配指南

其他应用领域参考以下步骤裁剪适配：

a) 明确本领域典型模态组合和核心任务；
b) 从第 6 章指标体系中选择适用指标，删除无关指标；
c) 补充领域特有指标；
d) 通过专家咨询确定指标权重；
e) 设定关键指标合格判据；
f) 在不少于 3 个实际项目中试用验证后优化指标配置。

参 考 文 献

[1] GB/T 41867—2022 信息技术 人工智能 术语

[2] ISO/PAS 8800:2024 Road vehicles — Safety and artificial intelligence

[3] ISO/IEC 5259 (all parts) Artificial intelligence — Data quality for analytics and machine learning (ML)

[4] 《中华人民共和国数据安全法》（2021年）