

团体标准

T/ITS XXXX-XXXX

高速公路人工智能视觉算法质量等级评估 指标和中试方法

Evaluation metrics for quality grading and pilot-scale testing methods of ai visual
Algorithms in freeway

(征求意见稿)

(本稿完成时间: 2026 年 7 月 21 日)

在提交反馈意见时, 请将您知道的相关专利连同支持文件一并附上。

2026-XX-XX 发布

2026-XX-XX 实施

中国智能交通产业联盟 发布

目 次

前言 III

1 范围 1

2 规范性引用文件 1

3 术语和定义、符号 1

4 缩略语 2

5 算法质量特性指标 3

6 算法等级评估 4

7 算法中试基准 4

8 算法中试平台功能要求 5

9 算法中试流程 6

附录 A（资料性）高速公路人工智能视觉算法中试报告（样例） 8

参 考 文 献 1

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这些专利的责任。

本文件由中国智能交通产业联盟（C-ITS）提出并归口。

本文件起草单位：

本文件起草人：

中国智能交通产业联盟

高速公路人工智能视觉算法质量等级评估指标和中试方法

1 范围

本文件确立了高速公路人工智能视觉算法中试体系，描述了算法质量特性评估指标及中试方法等内容。

本文件适用于指导高速公路行业人工智能算法使用方以及第三方等相关组织对人工智能视觉算法开展中试工作。

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 27964-2011 雾的预报等级

GB/T 28592-2012 降水量等级

GB/T 35273-2020 信息安全技术 个人信息安全规范

GB/T 40660-2021 信息安全技术 生物特征识别信息保护基本要求

GB/T 20480-2025 沙尘天气等级

GB/T 45225-2025 人工智能 深度学习算法评估

JT/T 1020-2016 交通运输信息系统 数据字典编制规范

T/AIA 016—2023 信创产品测评 通用要求

T/ITS 0247-2025 高速公路人工智能应用支撑平台通用要求

3 术语和定义、符号

下列术语和定义、符号适用于本文件。

3.1

高速公路人工智能视觉算法 freeway artificial intelligence visual algorithm

基于计算机视觉和深度学习技术，由深度学习模型及其配套逻辑构成的，专门用于高速公路环境中自动识别、检测、跟踪和分析车辆、行人及其他交通相关目标和行为的方法集合。

3.2

算法中试 algorithm pilot-scale test

对人工智能算法进行系统性验证与评估，以确认其在特定场景下的安全性、可用性 & 业务适用性，为算法上线准入、采购验收和责任界定提供核心依据的关键测试环节。

3.3

算法中试报告 algorithm pilot-scale testing report

准确、清晰、完整地总结反映算法中试活动与结果的一种文档。

3.4

算法中试基准 algorithm pilot-scale testing benchmark

由中试基准数据集、中试基准脚本以及中试基准环境组成的，满足标准化、可复现的算法中试评估体系。

3.5

中试基准数据集 pilot-scale testing benchmark dataset
人工智能算法中试所使用的标准化数据集合。

3.6

中试基准脚本 pilot-scale testing benchmark script
人工智能算法中试所执行的标准化程序代码。

3.7

中试基准环境 pilot-scale testing benchmark environment
人工智能算法中试所运行的系统软硬件配置。

3.8

算法中试平台 algorithm pilot-scale testing platform
用于支撑人工智能算法中试的集成化系统。

3.9

边缘情况 corner case
极端恶劣天气、重大交通事故、道路设施起火等发生概率低的情形或事件。

3.10

行业公开数据集 public industry dataset
由政府、研究机构或行业组织发布的，可供公众获取并直接用于深度学习模型训练或验证的数据集合。

3.11

自建测试数据集 custom-built test dataset
由企业或团体自行构建的，可直接用于深度学习模型验证的数据集合。

3.12

非对抗扰动 non-adversarial perturbation
对数据集添加自然噪声或环境变化等非刻意干扰，通常导致深度学习算法性能随机下降或输出不稳定，而非诱导其给出特定的错误结果的数据随机偏差。

3.13

负样本 negative sample
与视觉算法目标类别无关的数据实例。

4 缩略语

以下缩略语适用于本文件：

AI：人工智能（Artificial Intelligence）
ROC：受试者工作特性（Receiver Operating Characteristic）
TPR：真正率（True Positive Rate）
FPR：假正率（False Positive Rate）
AUC：ROC曲线下面积（Area Under the Curve）
PRC：精度召回率曲线（Precision-Recall Curve）
AP：平均精度（Average Precision）
CPU：中央处理器（Central Processing Unit）
GPU：图形处理器（Graphics Processing Unit）
NPU：神经网络处理器（Neural network Processing Unit）
TPU：张量处理器（Tensor Processing Unit）

IO: 输入输出 (Input-Output)

PDF: 便携文档格式 (Portable Document Format)

OFD: 开放版式文件 (Open Fixed-layout Documents)

HTML: 超文本标记语言 (HyperText Markup Language)

TOPS: 处理器运算能力单位, 万亿次运算操作每秒 (Tera Operations Per Second)

INT8: 8位整型数 (8-bit Integer)

5 算法质量特性指标

5.1 质量特性指标构成

高速公路人工智能视觉算法的质量特性包括功能性、性能效率、环境要素覆盖性和健壮性。在实施评估过程中, 应对不同类型的视觉应用算法组合不同质量特性指标指导具体评估。质量特性指标具体要求如下:

- a) 功能性指标: 对于不同类型视觉算法以及细分场景, 应以不同的功能性指标进行评估。参考 GB/T 45225 第 4.2 条规定, 功能性指标应包括但不限于准确率、精度、召回率、F1 值、平均精度等。
- b) 性能效率指标: 参考 GB/T 45225 第 4.3 条规定, 性能效率指标应包括但不限于:
 - 1) 时间效率: 宜使用平均响应时间、平均周转时间等指标来表示;
 - 2) 资源利用性: 宜使用处理器平均占用率、内存平均占用率、显存平均占用率等指标来表示。
- c) 环境要素覆盖性指标: 对于每一项环境要素, 应使用环境要素通过标志判断单项环境要素覆盖与否, 环境要素通过标志的计算方法见公式 (1), 环境要素应符合表 1 的要求。

$$Flag_i = \begin{cases} 1, & \text{if } \frac{N_{p_i}}{N_i} \geq \alpha_i \\ 0 & \end{cases} \dots\dots\dots (1)$$

式中:

$Flag_i$ ---第 i 个环境要素通过标志, 为Y代表通过, 为N代表不通过;

N_{p_i} ---算法推理结果正确的第 i 个环境要素用例数量;

N_i ---第 i 个环境要素用例总数量;

α_i ---第 i 个环境要素用例通过率阈值, 应根据实际应用需求设置。

表 1 环境要素覆盖性指标要求

环境要素大类	环境要素名称	通过标志
气象条件	能见度正常	Y/N
	雨	Y/N
	雪	Y/N
	雾	Y/N
	沙尘	Y/N
光照条件	正常	Y/N
	过亮	Y/N

环境要素大类	环境要素名称	通过标志
通行状况	过暗	Y/N
	正常通行	Y/N
	半幅封闭	Y/N
主体区域	路面	Y/N
	桥梁	Y/N
	隧道	Y/N
	立交	Y/N
	服务区/停车区连接线	Y/N
	收费站广场	Y/N

注 1:参考 GB/T 28592-2012, 环境要素雨应包括降雨量等级为小雨、中雨、大雨、暴雨、大暴雨及特大暴雨的情形。

注 2:参考 GB/T 28592-2012, 环境要素雪应包括降雪量等级为小雪、中雪、大雪、暴雪、大暴雪及特大暴雪的情形。

注 3:参考 GB/T 27964-2011, 环境要素雾应包括预报等级为轻雾、大雾、浓雾、强浓雾和特强浓雾的情形。

注 4:参考 GB/T 20480-2025, 环境要素沙尘应包括单地沙尘天气等级为浮尘、扬沙、沙尘暴、强沙尘暴和特强沙尘暴的情形。

d) 健壮性指标: 参考 GB/T 45225 第 4.7 条规定, 应对原始数据集进行图像加噪、多分辨率重塑等非对抗扰动处理后计算性能波动率。

5.2 算法质量特性指标选取原则

算法质量特性指标选取时应遵循以下原则:

- a) 场景适配性: 功能性指标选取时应符合视觉算法实际类型, 环境要素覆盖性指标选取宜充分结合算法实际部署的地域环境;
- b) 风险导向性: 当视觉算法涉及生命财产安全时, 相关功能性指标权重应为最高, 性能效率指标权重可适当提高。

6 算法等级评估

6.1 评估得分

高速公路人工智能视觉算法分析评估应符合GB/T 45225第6.4条规定, 根据选定算法质量特性指标计算评估得分。

6.2 等级判定

高速公路人工智能视觉算法评估等级应符合GB/T 45225第5条规定, 分为优越级、进阶级、条件级、受限级四个等级。等级判定应符合GB/T 45225第6.4条规定, 达到对应级别应同时大于或等于算法评估的总基准分值和各质量特性的基准分值的要求。

7 算法中试基准

7.1 中试基准数据集

中试基准数据集应满足包括但不限于以下要求:

- a) 充分性要求: 应严格围绕算法需求、实际使用场景进行构建, 构建时应满足:

- 1) 真实性: 应优先从高速公路真实运行环境采集数据, 确保数据集整体真实程度; 对于边缘情况, 可在高速公路封闭现场采用模拟的方式获取或通过高仿真技术生成。所有数据应注明数据真实性来源;
- 2) 多样性: 应满足环境要素覆盖性指标计算, 应包含进行非对抗扰动处理后的数据样本以计算健壮性指标。
- b) 来源要求: 中试基准数据集来源宜包括行业公开数据集和自建测试数据集
- c) 规模分布要求: 中试基准数据集规模分布应满足包括但不限于以下要求:
 - 1) 非边缘情况每类目标真实样本数应不少于 500 例;
 - 2) 边缘情况每类真实样本数不宜低于 50 例; 若真实样本数不足, 可采用“真实 + 现场模拟/高仿真生成”混合模式, 总数不宜低于 100 例, 真实样本数占总数比例应不低于 10%;
 - 3) 应避免单一类型样本数占比超 50%, 各类型样本分布应接近真实分布;
 - 4) 应包含负样本, 负样本总体占比应不低于 10%;
- d) 合规性与标注质量要求: 中试基准数据集合规性与标注质量应满足以下要求:
 - 1) 数据集真值标注准确率不应低于 98%, 宜通过至少两名标注员复核;
 - 2) 数据集的标注框、分割掩码、属性标签等标注应符合 JT/T 1020-2016 数据字典规范;
 - 3) 宜对人脸、车牌号等数据个人敏感信息进行去标识化处理, 处理应符合 GB/T 35273-2020 和 GB/T 4066-2021 要求。

7.2 中试基准脚本

中试基准脚本应满足包括但不限于以下要求:

- a) 应满足计算功能性指标, 包括但不限于准确率、精度、召回率和 F1 值等;
- b) 应满足计算性能效率指标, 包括但不限于平均响应时间、平均周转时间、处理器平均占用率、内存平均占用率和显存平均占用率等;
- c) 应满足计算环境要素覆盖性指标;
- d) 应满足计算健壮性指标。

7.3 中试基准环境

中试基准环境应提供不同高速公路行业在用主流软硬件配置, 分成以下两类:

注: 软硬件配置包括 CPU、算力卡 (GPU、NPU、TPU 等)、操作系统等。

- a) 边缘级推理基准环境: 宜满足 CPU 核心数不少于 16 核, 单核 CPU 主频不低于 1.9 GHz, 算力卡算力不低于 176 TOPS@INT8, 操作系统符合 T/AIA 016—2023 要求;
- b) 中心级推理基准环境: 宜满足 CPU 核心数不少于 48 核, 单核 CPU 主频不低于 2.6 GHz, 算力卡算力不低于 580 TOPS@INT8, 操作系统符合 T/AIA 016—2023 要求。

8 算法中试平台功能要求

8.1 测试环境管理

算法中试平台应提供测试环境的配置、部署、维护与监控功能, 要求如下:

- a) 应支持测试环境构建与配置, 宜使用容器化技术完成自动化部署, 宜对测试环境所使用的软件及依赖库版本进行记录和锁定;
- b) 应支持测试环境隔离与复用, 隔离方式应包括逻辑隔离和物理隔离, 环境配置宜支持模板导入、修改和导出;
- c) 应具备资源监控与日志管理功能, 支持实时采集 CPU、内存、磁盘 I/O 等系统资源指标, 并完整记录测试脚本执行过程及关键操作。

8.2 数据集管理

算法中试平台应提供数据集的概览、存储、标签化管理及版本控制功能，要求如下：

- a) 应支持数据集概览，展示数据集来源、规模、分布及标签信息，宜提供数据样本示例；
- b) 应支持数据集版本管理，数据集应具备明确的版本标识；
- c) 应支持数据集存储与备份，宜提供存储周期和备份策略。

8.3 数据隐私保护

算法中试平台应为数据集中个人敏感信息及其他受保护数据提供安全与合规保障功能，防止隐私数据被泄露或滥用。要求如下：

- a) 应支持数据脱敏与匿名化处理，示例数据应经过不可逆的脱敏或匿名化处理；
- b) 应支持访问控制与加密，应具备数据访问权限控制，对静态存储和动态传输的敏感数据宜采用加密等安全保护措施。

8.4 模型权重管理

算法中试平台应支持对高速公路人工智能视觉算法的模型权重文件进行管理并加密处理。要求如下：

- a) 模型权重文件应支持版本控制，应具备唯一的版本标识，宜关联测试脚本和中试基准数据集的版本及参数配置等信息；
- b) 模型权重文件应实现安全存储，并应具备访问权限控制；
- c) 应记录模型权重文件的名称、上传时间、哈希校验值等基本信息。

8.5 测试脚本管理

算法中试平台应支持测试脚本的开发、维护、调度执行与权限管理，要求如下：

- a) 应支持测试脚本编辑与版本控制，应提供脚本编辑工具，宜支持版本控制系统（SVN、Git等）进行版本管理；
- b) 应支持测试脚本调度与执行，应提供测试脚本定时调度、事件触发及手动执行功能；应支持测试脚本执行状态监控，并在执行失败时提供告警和日志记录；
- c) 应支持测试脚本权限控制，宜支持对测试脚本进行安全扫描。

8.6 测试报告管理

算法中试平台应支持测试报告自动生成和导出，要求如下：

- a) 应支持多种格式（PDF、OFD、HTML、Microsoft Word、Markdown等）导出，宜支持报告模板自定义。
- b) 应支持测试报告导出权限管理，导出操作应记录操作人、时间、IP地址、导出格式及文件哈希校验值。

9 算法中试流程

9.1 概述

高速公路人工智能视觉算法中试流程由试前准备、基准测试、结果评定和报告编制组成，过程应遵循科学性、公平性、可重复性及实用性原则，确保算法中试结果客观、准确、可重复。

9.2 试前准备

试前准备应包括但不限于：

- a) 中试基准数据集选定，应根据算法类型及实施场景，选择对应中试基准数据集；
- b) 质量特性指标选择，应根据算法质量特性指标结合具体场景要求设置；
- c) 中试基准环境选定，应根据算法运行场景，选择对应基准测试环境进行测试；
- d) 算法中试平台选定，应满足算法中试平台功能要求，宜依托T/ITS 0247-2025中所定义的高速公路人工智能应用支撑平台。

9.3 基准测试

基准测试应包括但不限于：

- a) 基准数据流测试，接入中试基准数据集进行测试，统计选定算法质量评估指标；
- b) 负样本专项测试，接入中试基准数据集中负样本进行测试，统计误报率；
- c) 边缘情况专项测试，接入中试基准数据集中边缘情况样本，统计漏报率。

9.4 结果评定

高速公路人工智能视觉算法中试结果应根据以下方法进行综合评定：

- a) 应根据算法等级评估方法，计算算法评估得分以及所处等级；
- b) 应设立“一票否决”项，当算法核心质量特性评估指标低于预设下限时，评估等级应定为受限级，并说明否决原因；
- c) 宜提供算法适用性说明，可提供算法优化建议。

9.5 报告编制

报告编制要求应包括：

- a) 中试报告内容应该包含中试基准情况、基准测试结果和最终评定结果；
- b) 中试报告应具备唯一标识性和跨平台可读性，便于审核、归档及交换；
- c) 中试报告应进行档案封存，应将中试报告、中试基准数据集版本、测试日志、模型哈希值打包加密，建立“一算法一档案”，保存期应不少于3年。

高速公路人工智能视觉算法中试报告示例见附录A。

附录 A

(资料性)

高速公路人工智能视觉算法中试报告（样例）

报告编号： TEST-2026-HW-AI-001

被测算法名称： 高速公路全域交通事件智能检测算法

版本号： V1.0.0

测试日期： 2026年6月1日 - 2026年6月30日

测试机构： [第三方测试机构名称]

(1) 评估摘要

- 核心结论： 算法综合评估得分为 88.5分，等级评定为“进阶级”。
- 适用性说明： 本算法适用于常规光照及气象条件下的事件检测场景，可覆盖所有选定通行条件及主体区域。在“雨+过暗”复合环境要素下，算法事件漏报率将显著上升，超出算法可处理上限。

(2) 测试环境和对象

- 被测算法信息：
 - 模型架构： YOLOv8-Trans + DeepSORT
 - 配置参数： 输入分辨率1080P@25fps，置信度阈值0.65，NMS阈值0.45
- 基准数据集信息：
 - 数据来源组成： 行业公开数据集HAVE-Dataset（30%）+ XX路段实路采集自建数据集V1.0（70%）。
 - 规模与分布： 总样本数12,000例。非边缘情况（包括异常停车、施工作业车辆停留、行人闯入、施工人员、摩托车、拥堵、抛洒物）每类>500例；边缘情况（包括道路火焰、逆行）采用“真实+模拟”混合模式共120例（真实占比15%）；负样本占比12%。
 - 标注质量： 真值标注经双人复核，涉及人脸/车牌均已脱敏。
- 基准环境信息： 运行于中心级推理基准环境。
- 选定环境要素：
 - 气象条件： 能见度正常、雨、雾
 - 光照条件： 正常、过亮、过暗
 - 通行状况： 正常通行、半幅封闭
 - 主体区域： 路面、桥梁、隧道、立交、服务区/停车区连接线、收费站广场

(3) 算法评估结果

质量特性 指标类型	指标名称	测试结果	评估得分	评估结果
功能性	准确率	95.20%	95.48	优越级
	召回率	85.65%	86.32	进阶级
	平均精度	76.53%	80.07	进阶级
性能效率	平均响应时间	18 ms	95	优越级
	处理器平均占用率	18.79%	90	进阶级
	内存平均占用率	20.90%	85	进阶级
	显存平均占用率	24%	85	进阶级
环境要素 覆盖性	气象条件-能见度正常	Y	84.97	进阶级
	气象条件-雨	N		
	气象条件-雾	N		
	光照条件-正常	Y		
	光照条件-过亮	N		
	光照条件-过暗	N		
	通行状况-正常通行	Y		
	通行状况-半幅封闭	Y		
	主体区域-路面	Y		
	主体区域-桥梁	Y		
	主体区域-隧道	Y		
	主体区域-立交	Y		
	主体区域-服务区/停车区连接线	Y		
	主体区域-收费站广场	Y		
健壮性	性能波动率	4.86%	87.61	进阶级
总体			88.5	进阶级

(4) 专项测试

- 负样本专项测试Top10误报类型：
 - 路面积水反光误报为白色抛洒物
 - 低速移动养护作业车辆误报为违停
 - 路面阴影/树影误报为行人或障碍物
 - 车辆远光灯眩光误报为火焰或异常光源
 -
- 边缘情况专项测试Top10漏报场景
 - 暴雨夜间无补光环境下远距离行人漏报
 - 隧道出入口暗光黑洞区域车辆漏检

- 浓雾天气下低对比度小型目标漏报
- 逆光强光背景下深色衣着行人漏检
-

(6) 最终结论和优化建议

- 评价结论：该算法总体达到“进阶级”水平。
- 优化建议：优化算法对光照条件为过亮、过暗及气象条件为雨雪时的鲁棒性，减少因环境干扰导致的漏检与误报。

测试负责人：[签名]

授权签字人：[签名]

测试机构盖章：[CMA/CNAS认证章 或 测试机构公章]

签发日期：2026年XX月XX日

本报告仅对“高速公路全域交通事件智能检测算法 V1.0.0”版本及本次中试基准条件负责。

若发生算法版本更新、部署硬件变更或核心场景重大变化，需重新进行测试评估。

未经本机构书面批准，不得部分复制本报告。

测试原始数据、模型哈希校验值及运行日志已按“一算法一档案”要求加密封存，保存期不少于3年。

参 考 文 献

- [1] GB/T 41867-2022 信息技术 人工智能 术语
- [2] GB/T 28789-2025 视频交通事件检测器
- [3] T/CCTAS 311—2026 高速公路视频人工智能分析系统指南

中国智能交通产业联盟

T/ITS XXXX-XXXX

中国智能交通产业联盟

中国智能交通产业联盟

标准

高速公路人工智能应用支撑平台通用要求

T/ITS XXXX-XXXX

北京市海淀区西土城路 8 号（100088）

中国智能交通产业联盟印刷

网址：<http://www.c-its.org.cn>

2025 年 11 月第一版 2025 年 11 月第一次印刷