ICS 35.240
CCS L70

SJ

中华人民共和国电子行业标准

SJ/T XXXXX—XXXX

人工智能 面向计算机视觉的自主学习系统技术规范

Artificial intelligence—Technical specification for self-learning system for computer vision

报批稿

XXXX－XX－XX 发布　　XXXX－XX－XX 实施

中华人民共和国工业和信息化部　　发布

目　次

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。

本文件由工业和信息化部科技司提出。

本文件由中国电子技术标准化研究院归口。

本文件起草单位：中国电子技术标准化研究院、杭州海康威视数字技术股份有限公司、赛西(深圳)电子信息产品标准化工程中心有限公司、中国电子工业标准化技术协会、清华大学、浙江大华技术股份有限公司、海信集团控股股份有限公司、杭州市北京航空航天大学国际创新研究院（北京航空航天大学国际创新学院）、北京海天瑞声科技股份有限公司、北京百度网讯科技有限公司、浪潮通信信息系统有限公司、浪潮云信息技术股份公司、浪潮电子信息产业股份有限公司、深圳思谋信息科技有限公司、合肥中科君达视界技术股份有限公司、中国电子技术标准化研究院华东分院、广电运通集团股份有限公司、马上消费金融股份有限公司、中移雄安信息通信科技有限公司、广东中科凯泽信息科技有限公司、昆仑数智科技有限责任公司、西北工业大学、上海计算机软件技术开发中心、中国南方电网有限责任公司超高压输电公司、中移（苏州）软件技术有限公司。

本文件主要起草人：马珊珊、任文奇、焦廉洁、钱晓东、迟子秋、郭阶添、朱江、任烨、谭文明、翁力帆、钟凯伦、李悦、陈庆帅、孙婷婷、赵梦芳、季向阳、沈芷月、聂简荻、方贵明、程淼、矫佩佩、康林林、郝玉峰、蒲逸凡、谭啸、陈坤斌、吴月升、肖红梅、梁秉豪、郑佳佳、邸贺亮、沈小勇、陈鹏光、严德斌、吴全进、严小格、黄宇恒、陈良旭、曾定衡、邓伟洪、郑庆国、吴军、王小宏、袁立杰、张艳宁、王鹏、陈敏刚、乔柱桥、闫伟。

人工智能　面向计算机视觉的自主学习系统技术规范

1　范围

本文件确立了面向计算机视觉的自主学习系统参考架构，规定了系统的功能和性能要求，描述了对应的测试方法。

本文件适用于面向计算机视觉的自主学习系统产品和服务的设计、开发、应用和测试评价。

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 41864-2022 信息技术 计算机视觉 术语

GB/T 41867-2022 信息技术 人工智能 术语

3　术语和定义

GB/T 41864-2022和GB/T 41867-2022界定的以及下列术语和定义适用于本文件。

3.1

自主学习 self learning

一种基于基线模型和基线标注数据，对生产环境中的无标注数据进行自动化标注，并通过在线方式持续优化和升级的机器学习技术。

3.2

自主学习系统 self learning system

基于自主学习技术开发的智能系统，能够在无人工干预的情况下，实现模型性能的自动优化提升。

3.3

基线标注数据 baseline labeled data

前期离线积累的、用于训练基线模型的有标签数据。

3.4

基线模型 baseline model

基于基线数据训练得到的初始模型。

3.5

在线标注数据 online labeled data

用户应用场景采集的数据经预处理后，生成的带伪标签或人工标注标签的数据。

3.6

半监督机器学习 semi-supervised machine learning

在训练过程中，能够同时使用标注数据和无标注数据进行训练的一种机器学习任务。

[来源：GB/T 41867-2022, 3.2.2]

3.7

伪标签 pseudo label

由模型推理生成的，而不是由人类标定的标签信息。一般而言伪标签性能指标越高，训练后的模型性能上限越高，但该标签会始终存在一定比例的噪声。

3.8

鲁棒训练 robust training

在训练数据的伪标签存在噪声的情况下，进行稳定有效模型训练的方法。

3.9

模型评价 model evaluation

根据模型类别选择合适的评价参数，对新训练的模型和已部署模型的性能优劣进行比较。

4 缩略语

下列缩略语适用于本文件。

AI：人工智能（ Artificial Intelligence）

AVS：音视频编码标准（Audio Video coding Standard）

BMP：位图（Bitmap）

CNN：卷积神经网络（Convolutional Neural Network）

DPI：每英寸点数（Dots Per Inch）

FADR：误报下降率（ False Alarm Decrease Ratio）

JPEG：联合图像专家组（Joint Photographic Experts Group）

MOT：模型优化周期（Model Optimization Term）

MPEG：动态图像专家组(Moving Picture Experts Group)

ONNX：开放神经网络交换格式（Open Neural Network Exchange）

PNG：便携式网络图形（Portable Network Graphics）

PS：节目流（Program Stream）

TIFF：标签图像文件格式（Tag Image File Format）

TPIR：正报提升率（True Positive Increase Ratio）

5 系统架构

5.1 概述

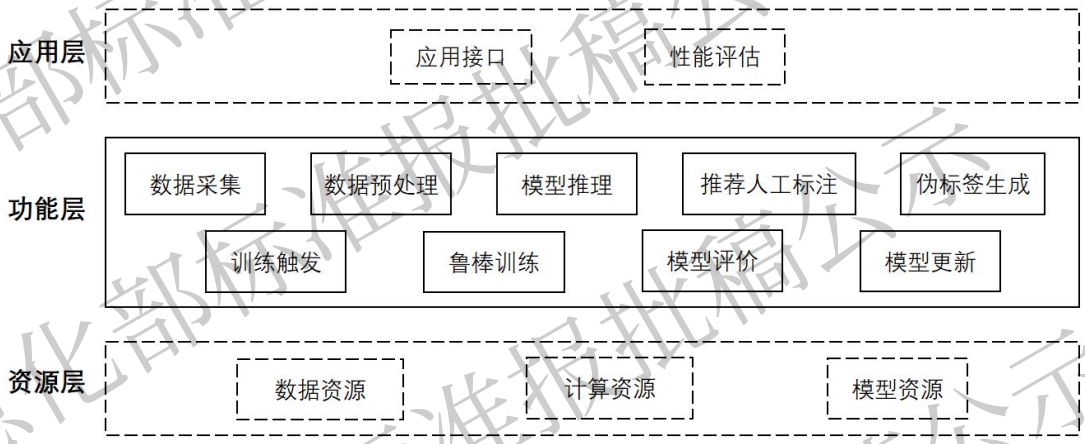

图 1 自主学习系统参考架构

自主学习系统参考架构见图1，包括资源层、功能层和应用层：

a) 资源层包含数据资源、计算资源和模型资源，分别用于支持自主学习系统的数据采集、模型训练和推理；

b) 功能层包括数据采集、数据预处理、模型推理、推荐人工标注、伪标签生成、训练触发、鲁棒训练、模型评价、模型更新等模块，是自主学习系统的核心功能模块；

c) 应用层包括应用接口、性能评估等模块，主要用于系统在计算机视觉任务场景下的部署应用和性能评估。

5.2 工作流程

自主学习系统在基线模型和基线数据的基础上，利用基线模型对在线采集的数据进行分析、自动生成标签，然后利用伪标签数据进行模型自主迭代学习。自主学习系统克服了传统方法数据依赖性强、场景泛化能力不足的问题，整个流程可在无需人工标注和干预下自动完成，也可引入推荐人工标注提高数据标签的准确性，实现模型性能的持续自动提升。

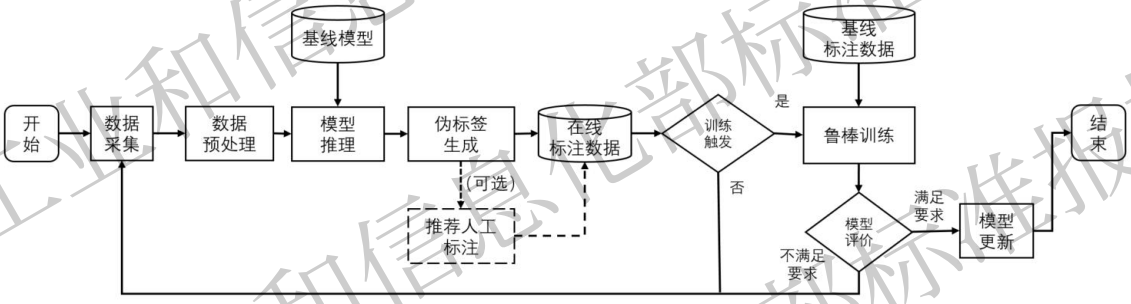

图 2 自主学习系统工作流程

自主学习系统工作流程见图2。其中：

a) 数据采集：通过数据采集单元（如摄像头）采集用户现场的图像或视频数据；

b) 数据预处理：对采集到的图像或视频进行数据增强、格式转换等操作；

c) 模型推理：将采集到的图像视频数据通过基线模型进行推理，获得推理结果；

d) 伪标签生成：基于推理结果，生成对应数据的伪标签，并存储到在线标注数据库中；

e) 推荐人工标注：本模块为可选模块，对高价值数据（如分类置信度较低的样本）进行人工推荐标注，以提高标签生成结果的准确性；
f) 训练触发：对在线数据进行分布统计，基于硬件资源监控信息和样本分布变化等信息判断是否满足触发条件，若满足则开始鲁棒训练，否则继续数据采集；
g) 鲁棒训练：加载基线标注数据和在线标注数据，在存在标签噪声条件下进行稳定有效训练，生成新的模型；
h) 模型评价：对训练完成后对模型性能指标进行评价分析，若满足要求则进行模型更新，否则继续数据采集；
i) 模型更新：结合模型评价结果，生成模型升级包并在设备上进行更新部署。

6 功能要求

6.1 数据采集

数据采集功能要求如下：
a) 应支持读取主流图像格式，包括但不限于 JPEG、PNG、TIFF、BMP 等格式；
b) 应支持读取主流视频格式，包括但不限于 H.264、H.265、AVS2、AVS3 等封装协议的 PS、MPEG、AVI 等格式；
c) 应具备读取分辨率在 128×128 DPI～4096×4096 DPI 范围内的图像；
d) 宜具备对视频进行抽帧并读取图像帧，抽帧策略具备自定义配置。

6.2 数据预处理

数据预处理功能要求如下：
a) 应具备图像变换能力，包括尺寸调整、旋转变换、裁剪处理、格式转换等基础操作；
b) 应具备图像质量增强能力，通过滤波算法优化图像的清晰度、几何形变、背景干扰和光照条件等问题，在不破坏图像边缘、轮廓、纹理等原有细节的条件下对噪声进行抑制；
c) 宜具备对图像成像质量进行评价，对图像中目标是否满足视觉辨识（包括人眼或机器视觉）的成像要求进行判断，过滤无法识别的低质量图像。

6.3 模型推理

模型推理功能要求如下：
a) 应具备进行典型计算机视觉任务处理能力，如检测、分类、识别、分割等算法；
b) 应具备主流的模型网络结构，如卷积神经网络（CNN）、Transformer 等深度学习模型；
c) 应支持主流模型格式（如 ONNX）的加载和推理。

6.4 推荐人工标注

本模块为可选模块，宜具备对高价值数据样本进行挑选和推荐，并进行人工复核标注。

6.5 伪标签生成

伪标签生成功能要求如下：
a) 应具备对线上无标签数据进行分析并生成业务相关的伪标签；
b) 分类的标签应采用规范的类别标识符，具备中文、英文或数字形式，如（人、车，0，1 等）；
c) 检测的标签应为目标框位置信息(x,y,w,h)和类别信息表示；

d) 分割的标签应为与图像对应的分割图像（分割图像像素值为类别阿拉伯数值）；
e) 识别的标签应为目标 ID 信息；
f) 宜具备伪标签过滤和优化策略，如强弱增强、阈值选择等。

6.6 训练触发

训练触发功能要求如下：

a) 应具备自动或手动触发自主学习训练功能，成功触发则进入训练环节，否则回到数据采集环节。触发条件可包括：
b) 样本数量变化，如新增样本数量超过现有数据集的 10%；
c) 样本分布变化，如样本分布 KL 散度变化超过 0.1；
d) 模型性能衰减，如在线评估的准确率下降超过预设阈值。

6.7 鲁棒训练

鲁棒训练功能要求如下：

a) 应具备基于无标签数据、伪标签在线数据和基线数据的可靠训练功能，保证在标签噪声干扰下的模型鲁棒性；
b) 应具备在边缘侧和中心侧的训练计算单元上进行鲁棒训练能力；
c) 宜具备在具备训练能力的端侧训练计算单元上进行鲁棒训练能力；
d) 宜具备 MeanTeacher、FixMatch、CoMatch, SimMatch 等半监督机器学习方法。

6.8 模型评价

模型评价功能要求如下：

a) 应具备查看基线模型和部署模型的情况，包括模型类型、模型运行状态等；
b) 应具备对训练生成的模型进行性能评价功能，在测试集上对新训练的模型和已有模型进行性能对比；
c) 应具备常用计算机视觉任务（如目标检测、图像/视频分类、图像分割等）的性能评价参数，包括但不限于分类准确率、召回率和误报率等。

6.9 模型更新

模型更新功能要求如下：

a) 应具备自动生成符合预期的训练模型的升级包，并提供给硬件设备升级功能。该模块可将评价指标符合预期的模型，采用自动化封装的方式转换成升级包，然后提供给硬件自动升级版本；
b) 宜具备差分更新、灰度发布、更新效果跟踪、应急回滚等机制。

7 性能要求

7.1 正报提升率（TPIR）

正报提升率用于衡量自主学习算法与基线模型分析中正确的数量提升的比例，在通常计算机视觉应用场景下，自主学习后正报提升率宜大于等于 10%。

7.2 误报下降率（FADR）

误报下降率用于衡量算法正确分析的数量占所有真实存在的正确样本的比例，在通常计算机视觉应用场景下，自主学习后误报下降率宜大于等于 10%。

7.3 模型优化周期（MOT）

模型优化周期是指自主学习系统进行一轮新模型训练和升级所需的时间，在通常计算机视觉应用场景下，模型优化周期宜不超过 14 天。

8 测试方法

8.1 功能测试方法

8.1.1 数据采集

数据采集采用下列测试方法：

a) 测试用户构造 JPEG、PNG、TIFF、BMP 等格式的图像样本，并检查系统是否正常读取；
b) 测试用户构造 H.264、H.265、AVS2、AVS3 等封装协议的 PS、MPEG、AVI 等格式视频，并检查系统是否正常读取；
c) 测试用户构造分辨率在 128×128 DPI～4096×4096 DPI 范围内的图像，并检查系统是否正常读取；
d) 测试用户自定义配置抽帧策略，并检查系统是否正常抽帧并读取图像帧。

8.1.2 数据预处理

数据预处理采用下列测试方法：

a) 测试用户对图像进行尺寸调整、旋转变换、裁剪处理、格式转换等操作，并检查图像是否正常；
b) 测试用户对注入噪声的图像进行滤波、噪声抑制等操作，并检查图像质量增强效果；
c) 测试用户对模糊、遮挡等低质量的图像进行成像质量评价，并检查评价的准确性。

8.1.3 模型推理

模型推理采用下列测试方法：

a) 测试用户加载标准图像数据集（如 ImageNet、MS COCO 等）进行模型推理，检查系统是否具备检测、分类、识别、分割等计算机视觉任务处理能力；
b) 测试用户加载 ResNet50、ViT 等不同网络结构的模型，检查系统是否正常加载模型。

8.1.4 推荐人工标注

测试用户输入低置信度样本（如置信度<0.6），检查系统能否主动推荐人工复核。

8.1.5 伪标签生成

伪标签生成采用下列测试方法：

测试用户输入无标签的测试图像进行模型推理，并检查系统是否构建了相应的伪标签：

a) 对于分类任务，检查标签是否采用规范的类别标识符；
b) 对于检测任务，检查标签是否为目标框信息和类别信息；
c) 对于分割任务，检查标签是否为对应的分割图像；

d) 对于识别任务，检查标签是否为目标 ID 信息

e) 检查是否具备强弱增强、阈值选择等伪标签过滤和优化策略。

8.1.6 训练触发

训练触发采用下列测试方法：

a) 测试用户加载超过 10%的新增样本数据，并检查是否触发训练；

b) 测试用户加载修改样本分布（KL 散度>0.1），并检查是否触发训练。

8.1.7 鲁棒训练

鲁棒训练采用下列测试方法：

a) 测试用户加载基线数据、无标签数据和伪标签在线数据，并检查模型是否正常训练；

b) 测试用户在边缘侧和中心侧的训练计算单元上进行训练，并检查模型是否正常训练；

c) 测试用户在具备一定算力的端侧训练计算单元上进行训练，并检查模型是否正常训练；

d) 测试用户启用 MeanTeacher、FixMatch、CoMatch, SimMatch 等半监督机器学习方法，并检查模型能否正常训练。

8.1.8 模型评价

模型评价采用下列测试方法：

a) 测试用户检查是否具备查看基线模型和部署模型的模型类型、模型运行状态等信息；

b) 测试用户加载测试集，并检查新训练的模型和已有基线模型的性能指标差异；

c) 测试用户检查是否具备目标检测、图像分类、分割等常用计算机视觉任务。

8.1.9 模型更新

模型更新采用下列测试方法：

a) 在满足模型升级的的条件下，测试用户检查系统是否正常生成了符合预期的训练模型升级包，并提供给硬件设备升级；

b) 测试用户检查是否具备差分更新、灰度发布、更新效果跟踪、应急回滚等机制。

8.2 性能测试方法

测试人员应根据不同计算机视觉任务场景制作测试数据集，如视频结构化、人脸识别、车辆识别、周界入侵检测等典型计算机视觉任务场景。训练场景与测试场景应保持一致。

测试人员应基于系统规格要求，搭建标准化的测试环境，确保软硬件配置满足测试需求，保证被测系统在环境中运行正常。

测试人员应确定选择第7章节中描述的若干测试指标作为测试目标，选择相应的测试数据集执行测试，按照标准格式记录和保存测试过程数据，生成规范化的测试结果报告。

8.2.1 正报提升率（TPIR）计算方法

计算公式如下：

$$TPIR = \frac{C_1 - C_0}{C_0} \times 100\% \quad \text{.....(1)}$$

式中：

TPIR——正报提升率；

C_0——基线模型分析结果中正确的数量；

C_1——为自主学习后模型分析结果中正确的数量。

8.2.2 误报下降率（FADR）计算方法

计算公式如下：

$$FADR = \frac{N_0 - N_1}{N_0} \times 100\% \quad (2)$$

式中：

FADR——误报下降率；

N_0——基线模型分析结果中错误的数量；

N_1——自主学习后模型分析结果中错误的数量。

8.2.3 模型优化周期（MOT）计算方法

计算公式如下：

$$MOT = T_0 + T_1 + T_2 + T_3 \quad (3)$$

式中：

MOT——模型优化周期；

T_0——数据准备周期；

T_1——模型训练周期；

T_2——模型评估周期；

T_3——模型更新周期。

附 录 A
（资料性）
自主学习系统参考部署方式

根据计算机视觉应用任务场景及计算设备算力的要求，自主学习系统可部署在端侧、边缘侧和中心侧的计算机视觉系统中。如图 A.1 所示。

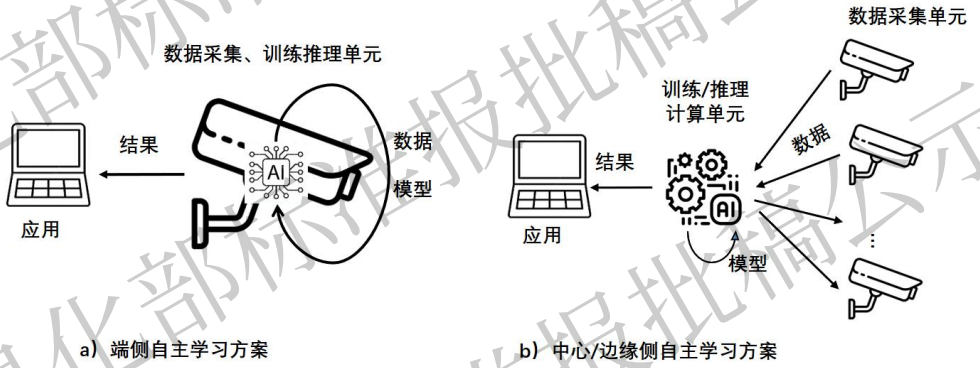

图 A.1 自主学习系统在端侧和中心/边缘侧部署方案

A.1 端侧部署方案

端侧部署方案如下：

a) 端侧数据采集单元应具备训练、推理能力（如具备 AI 算力）；

b) 数据采集单元直接基于原始数据推理并生成伪标签，经过推荐人工标注（可选）、训练触发、鲁棒训练、模型评价后，更新部署在设备上的模型。

A.2 中心/边缘侧部署方案

中心/边缘侧部署方案如下：

a) 数据采集设备可不具备训练推理能力，由专用的推理计算单元和训练计算单元执行推理和训练任务。推理和训练可以在一台设备或多台设备上分别进行；

b) 数据采集单元将原始数据上传至训练/推理计算单元，由推理计算单元生成伪标签数据，训练计算单元利用基线数据和在线数据进行训练触发、鲁棒训练、模型评价后更新模型，更新后的模型可部署在中心侧、边缘侧或端侧。

参 考 文 献

[1] GB/T 41867-2022 人工智能 术语

[2] GB/T 41864-2022 信息技术 计算机视觉 术语

[3] B. Zhou, J. Lu, K. Liu, Y. Xu, Z. Cheng and Y. Niu, "HyperMatch: Noise-Tolerant Semi-Supervised Learning via Relaxed Contrastive Constraint," 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, BC, Canada, 2023, pp. 24017-24026