

电镜扫描硅藻图像自动化检验 技术规范

Technical Standard for automatic recognition of digital scanning
electron microscopy images of diatoms

(报批稿)

在提交反馈意见时，请将您知道的相关专利连同支持性文件一并附上。

XXXX – XX – XX 发布

XXXX – XX – XX 实施

发 布

目 次

前 言 II

1 范围 3

2 规范性引用文件 3

3 术语和定义 3

4 方法原理 5

5 设备要求 5

6 自动化检验系统模型构建 5

 6.1 图像输入模块 5

 6.2 识别模块 5

 6.3 统计输出模块 7

7 实施步骤 7

 7.1 图像采集 7

 7.2 图像输入与识别 7

 7.3 统计输出 7

 7.4 结果判读 7

8 测试及性能要求 7

 8.1 测试要求 7

 8.2 识别查准率的计算 8

 8.3 识别查全率的计算 8

 8.4 不完整硅藻检出率的计算 8

 8.5 识别速度的计算 8

 8.6 漏检率的计算 9

 8.7 性能要求 9

参考文献 9

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。

本文件由江苏省司法厅提出并归口。

本文件起草单位：南京医科大学司法鉴定所、南京市公安局水上分局。

本文件主要起草人：陈峰、刘晓荣、陈鹏、陈治秋、李开、俞尤嘉、王增良、李荣、冀强、周世一、张志威、唐伟、陈吉、黄帅楠。

电镜扫描硅藻图像自动化检验 技术规范

1 范围

本文件规定了电镜扫描硅藻图像自动化检验的设备要求、自动化检验系统模型构建、实施步骤、测试及性能要求。

本文件适用于电镜扫描硅藻图像自动化检验系统的构建、测试与使用。

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。

GA/T 1662-2019 法庭科学 硅藻检验技术规范 微波消解-真空抽滤-显微镜法

T/CESA 1036-2019 信息技术 人工智能 机器学习模型及系统的质量要素和测试方法

T/CESA 1040-2019 信息技术 人工智能 面向机器学习的数据标注规程

AIOSS-01-2018 人工智能 深度学习算法评估规范

3 术语和定义

GA/T 1662-2019和T/CESA 1036-2019界定的以及下列术语和定义适用于本文件。

3.1

电镜扫描硅藻图像 digital scanning electron microscopy images of diatoms

通过微波消解-真空抽滤-电子显微镜法检验脏器样本或水体样本中的硅藻时，电子显微镜扫描富集膜所得的数字化图像。

3.2

电镜扫描硅藻图像自动识别系统 automatic recognition system of digital scanning electron microscopy images of diatoms

基于计算机技术，对电镜扫描硅藻图像进行获取、分析识别并输出图像中硅藻识别结果的系统。

3.3

深度学习模型 deep learning model

采用深度学习方法建立的输入与目标输出联系的计算模型。主要包含算法、超参数、参数、模型输入规范、模型输出规范五大要素。

3.4

训练 training

对于给定的数据集，生成和优化深度学习模型参数设置的过程。

3.5

验证 validation

对于给定的数据集，采用训练后的模型进行预测，由此进行选择并优化训练模型结构和超参数的过程。

3.6

测试 testing

对于给定的数据集，采用训练后的深度学习模型进行预测，由此评估训练模型性能的过程。

3.7

查准率 precision

正确识别为硅藻的数量占识别为硅藻数量的比率。

3.8

查全率 recall

正确识别为硅藻的数量占样本中所有硅藻数量的比率。

3.9

漏检率 miss rate

未被识别为硅藻的真实硅藻数量与样本中所有硅藻数量的比值。

3.10

完整硅藻 complete diatom

指在电镜图像中呈现边线连续、内部无缺损、形态完整的硅藻体。

3.11

不完整硅藻 incomplete diatom

指在电镜图像中呈现边线不连续、内部有缺损、形态不完整的硅藻体，其产生的主要原因有：水体中本身存在碎片化不完整硅藻；样本在消解过程中产生部分碎片化不完整硅藻；图像分割造成硅藻图像不完整。

3.12

碎片硅藻 fragmented diatom

指在电镜图像中呈现缺损面积大于 50%的不完整硅藻体。

3.13

识别速度 recognition speed

对于输入的电镜扫描硅藻图像，系统识别每张图像平均所用的时间。

3.14

边界框 bounding box

经过识别模型分析后，在图像中框出的系统认为是硅藻的区域。

3.15

置信度 confidence coefficient

识别图像中的硅藻时返回的该边界框内区域是硅藻的概率数值。

4 方法原理

本方法基于人工智能图像识别技术，构建电镜扫描硅藻图像自动识别系统（以下简称系统），对电镜扫描硅藻图像中的硅藻进行自动化定位、定量分析。

5 设备要求

电镜扫描硅藻图像自动化检验过程中使用的设备应符合如下要求：

- a) 硅藻检验基本设备和耗材：符合GA/T 1662-2019中5.1和5.2规定的要求。
- b) 扫描电子显微镜：符合GA/T 1662-2019中5.2.1.5.2规定的要求。
- c) 计算机：
 - 处理器（CPU）：英特尔（intel）i7-9700或以上；
 - 内存（RAM）：DDR4 2666 16GB或以上；
 - 硬盘：500GB SSD或以上；
 - 显卡：RTX 2080 8G或以上；
 - 电源：功率750W或以上。

6 自动化检验系统模型构建

6.1 图像输入模块

6.1.1 模块要求

图像输入模块应能够批量导入电镜扫描硅藻图像，并输送至识别模块。

6.1.2 输入图像要求

待识别图像按照GA/T 1662-2019规定的方法进行硅藻检验和扫描电镜图像采集所得。推荐滤膜扫描区域直径16mm，扫描放大倍数1500×。

6.2 识别模块

6.2.1 模块要求

识别模块应能够利用训练好的自动识别模型识别输入的硅藻电镜图像，用边界框框出系统认为是硅

藻的区域并返回系统认为是硅藻的置信度。识别模块的组成和构建流程见图 1。

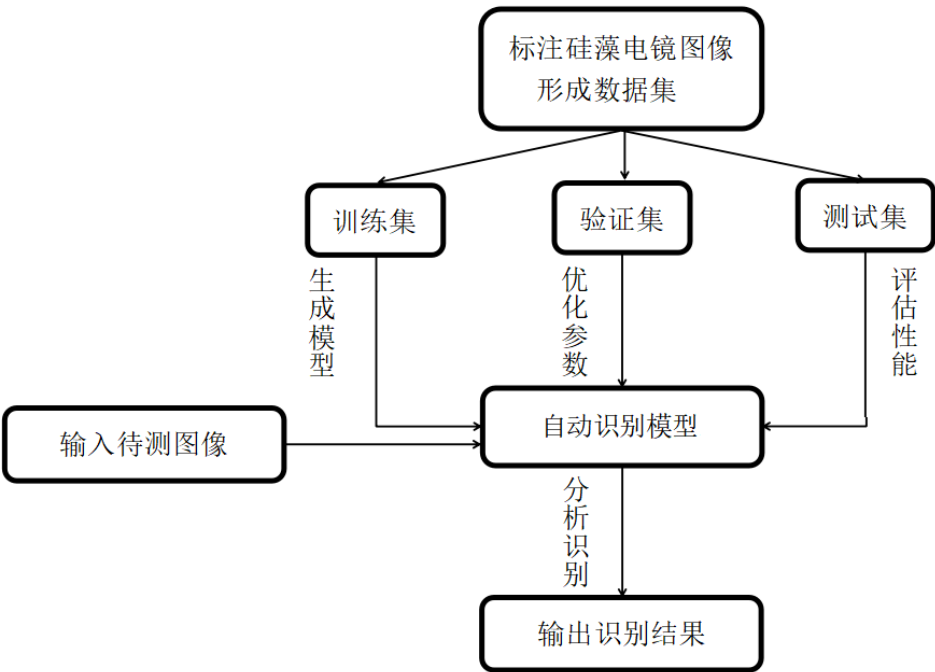


图1 识别模块组成和构建流程

6.2.2 数据集的建立

用于自动识别模型构建的数据集的建立应符合如下要求：

- a) 建立数据集所使用的电镜扫描硅藻图像可来自于本实验室仪器平台以及其他来源（如外部数据库等）。每张原始图片包含的硅藻数量应不少于1个。
- b) 原始图片应可进行像素大小调整，使得图片大小符合人工智能模型输入规格要求。调整后的图片应进行像素预处理（如均一化、标准化、灰度化等），降低不同数据源所带来的像素差异性。
- c) 按所选人工智能模型的数据输入规范选择符合要求的数据标注格式，标注图像形成数据集，标注规范符合T/CESA 1040-2019的规定。
- d) 标注者应为有丰富硅藻检验经验的法医学工作者。
- e) 标注结果应为带有标签的数据，包括“硅藻”这一标签，以及标签对应的图像空间位置。
- f) 标注图像数量宜大于1000张，且应涵盖当地所有常见硅藻类别，每种类别不少于100张。
- g) 对标注图像进行扩增，包括图片翻转、旋转、调整对比度及亮度等，最终形成数据集。

6.2.3 人工智能模型选择

可选择深度学习图像分类模型（ResNet-101、VGG-16、Inception-v3、MobileNet等）或者目标检测模型（Faster-RCNN、Yolo V3、SSD等），或其他适当的人工智能模型对数据集进行模式特征学习。选择的模型应为能够将硅藻与背景杂质相区分的模型架构。

6.2.4 模型训练与评估

按7:2:1的比例将数据集划分成互斥的训练集、验证集、测试集，分别对所选模型进行训练、训练过程中的模型效能监测和评估、模型训练完毕后的性能泛化指标评价。按T/CESA 1036-2019的规定对模型质量进行评估，并对参数进行测试和优化，形成最终的自动识别模型。

6.3 统计输出模块

统计输出模块应能够输出硅藻识别的结果，包括识别出硅藻的数量、边界框及置信度，并将识别出硅藻的图像与未识别出硅藻的图像分别分类存放。

7 实施步骤

7.1 图像采集

操作人员应按照GA/T 1662-2019规定的方法进行硅藻检验和电镜扫描图像采集。以某地水样或某脏器样本作为一个样品进行检验，每个样品使用一张滤膜对硅藻进行富集，并使用电镜扫描采集图像。推荐滤膜扫描区域直径16mm，扫描放大倍数1500×，每个样品富集滤膜的电镜扫描图像分别存放在独立的文件夹中。

7.2 图像输入与识别

图像输入与识别的过程应符合如下要求：

- a) 系统利用识别模型对输入图像中的硅藻进行识别，生成识别结果并进行保存。
- b) 识别结果包括在图像中显示识别出硅藻的边界框，并提供置信度。
- c) 若系统具有设置自选阈值功能，将置信度大于阈值的硅藻图像作为筛选结果进行保存。

7.3 统计输出

识别结果统计输出的过程应符合如下要求：

- a) 识别模块运行完毕，系统自动输出识别结果，包括图像上识别出硅藻的边界框和置信度。
- b) 系统将识别出硅藻的图像与未识别出硅藻的图像分类存放。
- c) 系统统计识别出硅藻的个数并记录存档。

7.4 结果判读

识别结果判读的过程应符合如下要求：

a) 系统对于输入图像的自动识别结果（包括系统按照阈值筛选所得结果）分为检出硅藻和未检出硅藻两类，图像分别存入两个独立的文件夹。

b) 点击查看检出硅藻的图像所在的文件夹，系统在每张图像上标注检出的单个硅藻所在的边界框，并在边界框右上角标注置信度，作为系统认为是硅藻的可能性，操作人员可根据边界框和置信度对识别结果进行验证。

c) 对于不完整硅藻（定义见3.12），系统视为硅藻一同检出。

d) 操作人员统计系统识别所得硅藻数量时，缺损面积小于50%的不完整硅藻计数为完整硅藻，缺损面积大于50%的不完整硅藻计数为碎片硅藻，分别计算完整硅藻和碎片硅藻数量。

e) 操作人员可依据一定扫描区域内的硅藻数量推断整个水样或脏器样本中硅藻的相对数量。

8 测试及性能要求

8.1 测试要求

系统测试的过程应符合如下要求：

- a) 系统正式投入使用前应利用真实样本对系统性能进行测试，性能达到本标准8.7中的规定，则认为识别结果具有可信度。
- b) 测试人员应为具有丰富硅藻检验经验、能够辨认硅藻形态的相关人员。
- c) 测试人员应按本标准7.1的要求采集电镜硅藻图像作为测试图像集，以每一个硅藻（包含完整硅藻和不完整硅藻）的位置作为一个“样本”，完整记录测试图像集中真正硅藻的位置及样本数量。将测试图像集输入系统识别后，以预测出的每一个边界框作为一个“样本”，人工判断预测样本是否为真正硅藻，记录数据，并与测试前数据进行比较。

8.2 识别查准率的计算

识别查准率计算按式（1）进行。

$$P = \frac{TP}{TP+FP} \dots\dots\dots(1)$$

- 式中：
- P* ——识别查准率；
 - TP* ——识别为真正硅藻的样本数；
 - FP* ——识别为硅藻而实际非硅藻的样本数。

8.3 识别查全率的计算

识别查全率计算按式（2）进行。

$$R = \frac{TP}{TP+FN} \dots\dots\dots(2)$$

- 式中：
- R* ——识别查全率；
 - TP* ——识别为真正硅藻的样本数；
 - FN* ——识别为非硅藻而实际为真正硅藻的样本数。

8.4 不完整硅藻检出率的计算

不完整硅藻检出率计算按式（3）进行。

$$D = \frac{TD}{TD+LD} \dots\dots\dots(3)$$

- 式中：
- D* ——不完整硅藻检出率；
 - TD* ——识别为真正不完整硅藻的样本数；
 - LD* ——识别为非硅藻而实际为真正不完整硅藻的样本数。

8.5 识别速度的计算

识别速度计算按式（4）进行。

$$V = \frac{T}{N} \dots\dots\dots(4)$$

- 式中：
- V* ——识别速度；
 - T* ——完成识别的总响应时间；

N ——识别图像总张数。

8.6 漏检率的计算

按设定阈值筛选出置信度高于阈值的硅藻样本后，漏检率计算按式（5）进行。

$$L = \frac{TL}{TL+TI} \dots\dots\dots(5)$$

式中：

- L ——漏检率；
- TL ——未被筛选出的真正硅藻的样本数；
- TI ——被筛选为真正硅藻的样本数。

8.7 性能要求

测试后的电镜扫描硅藻图像自动识别系统性能应达到如下要求：

- a) 识别查准率：水、肺等硅藻含量高的样本，识别查准率不低于90%。
- b) 识别查全率：水、肺等硅藻含量高的样本，识别查全率不低于90%；肝、肾等硅藻含量少的样本，识别查全率不低于95%。
- c) 不完整硅藻的检出：对于不完整硅藻，系统应视为硅藻一同检出。
- d) 识别速度：识别一张图像的平均需要时间不大于1s。
- e) 阈值设置和筛选：系统支持按自选阈值筛选置信度达到阈值的硅藻，当阈值设置在0.5以上时，筛选硅藻的漏检率不超过10%。

参 考 文 献

- [1] 陈吉, 刘晓荣, 杨佳雯, 陈治秋, 王诚, 欧梦媛, 吴嘉仪, 俞尤嘉, 李开, 陈鹏, 陈峰. 基于 YOLOv3 的硅藻电子显微镜图像识别模型的构建及应用. 法医学杂志, 2022, 38(1): 46-52.
- [2] Yu W, Xue Y, Rob K, Yu D, Evgeniya B, Kang X, Pietro F, Zheng D, Liu P, Chen H, Shi H, Liu C, Zhao J. Automated diatom searching in the digital scanning electron microscopy images of drowning cases using the deep neural networks. *Int J Leg Med*, 2021, 135(2): 497-508.
- [3] 李峥, 杨佳雯, 陈治秋, 史小杰, 吴天真, 陈峰, 陈鹏. 硅藻检验推测溺亡位置 1 例. 法医学杂志, 2020, 36(6): 878-880.
- [4] Zhou Y, Zhang J, Huang J, Deng K, Zhang J, Qin Z, Wang Z, Zhang X, Tuo Y, Chen L, Chen Y, Huang P. Digital whole-slide image analysis for automated diatom test in forensic cases of drowning using a convolutional neural network algorithm. *Forensic Sci Int*, 2019, 302: 109922.
- [5] Li Z, Liu X, Yu Y, Huang H, Li X, Ji Q, Li K, Yu Y, Li D, Mao Z, Pu Y, Chen P, Chen F. Barcoding for diatoms in the Yangtze River from the morphological observation and 18S rDNA polymorphic analysis. *Forensic Sci Int*, 2019, 297: 81-89.
-