



图像分割评估方法在显微图像分析中的应用

马博渊 姜淑芳 尹豆 申昊锴 班晓娟 黄海友 王浩 薛维华 封华

Image segmentation metric and its application in the analysis of microscopic image

MA Bo-yuan, JIANG Shu-fang, YIN Dou, SHEN Hao-kai, BAN Xiao-juan, HUANG Hai-you, WANG Hao, XUE Wei-hua, FENG Hua

引用本文:

马博渊, 姜淑芳, 尹豆, 申昊锴, 班晓娟, 黄海友, 王浩, 薛维华, 封华. 图像分割评估方法在显微图像分析中的应用[J]. *工程科学学报*, 2021, 43(1): 137-149. doi: 10.13374/j.issn2095-9389.2020.05.28.002

MA Bo-yuan, JIANG Shu-fang, YIN Dou, SHEN Hao-kai, BAN Xiao-juan, HUANG Hai-you, WANG Hao, XUE Wei-hua, FENG Hua. Image segmentation metric and its application in the analysis of microscopic image[J]. *Chinese Journal of Engineering*, 2021, 43(1): 137-149. doi: 10.13374/j.issn2095-9389.2020.05.28.002

在线阅读 View online: <https://doi.org/10.13374/j.issn2095-9389.2020.05.28.002>

您可能感兴趣的其他文章

Articles you may be interested in

基于监督学习的前列腺MR/TRUS图像分割和配准方法

Prostate MR/TRUS image segmentation and registration methods based on supervised learning
工程科学学报. 2020, 42(10): 1362 <https://doi.org/10.13374/j.issn2095-9389.2019.10.10.006>

基于自动多种子区域生长的遥感影像面向对象分割方法

Object-oriented remote sensing image segmentation based on automatic multiseed region growing algorithm
工程科学学报. 2017, 39(11): 1735 <https://doi.org/10.13374/j.issn2095-9389.2017.11.017>

基于改进CV模型的金相图像分割

Segmentation of metallographic images based on improved CV model
工程科学学报. 2017, 39(12): 1866 <https://doi.org/10.13374/j.issn2095-9389.2017.12.013>

一种改进的非刚性图像配准算法

An improved non-rigid image registration approach
工程科学学报. 2019, 41(7): 955 <https://doi.org/10.13374/j.issn2095-9389.2019.07.015>

一种提高图像对比度和视觉质量的新型区域背光算法

A novel local dimming algorithm to improve image contrast and visual quality
工程科学学报. 2017, 39(12): 1888 <https://doi.org/10.13374/j.issn2095-9389.2017.12.016>

序列图像运动自适应V1-MT光流估计算法

Bio-inspired motion-adaptive estimation algorithm of sequence image
工程科学学报. 2017, 39(8): 1238 <https://doi.org/10.13374/j.issn2095-9389.2017.08.014>

图像分割评估方法在显微图像分析中的应用

马博渊^{1,2,3,4}, 姜淑芳⁵, 尹 豆³, 申昊锴⁶, 班晓娟^{1,2,3,4}, 黄海友^{1,7,8}✉,
王 浩^{1,9}, 薛维华^{9,10}, 封 华³

1) 北京科技大学北京材料基因工程高精尖创新中心, 北京 100083 2) 北京科技大学材料领域知识工程北京市重点实验室, 北京 100083
3) 北京科技大学计算机与通信工程学院, 北京 100083 4) 北京科技大学人工智能研究院, 北京 100083 5) 解放军总医院第一医学中心
妇产科, 北京 100853 6) 中国石油大学(北京)信息科学与工程学院, 北京 102249 7) 北京科技大学新材料技术研究院, 北京 100083
8) 北京科技大学顺德研究生院, 佛山 528300 9) 北京科技大学材料科学与工程学院, 北京 100083 10) 辽宁工程技术大学材料科学与工
程学院, 阜新 123099

✉通信作者, E-mail: huanghy@mater.ustb.edu.cn

摘 要 图像分割是计算机视觉领域中的重要分支, 旨在将图像分成若干个特定的、具有独特性质的区域. 随着计算机硬件计算能力的提高和计算方法的进步, 大量基于不同理论的图像分割算法获得了长足的发展. 因而选择合适的评估方法对分割结果的准确性和适用性进行综合评估, 从而选择最优分割算法, 成为图像分割研究中的必要环节. 在综述 14 种图像分割评估指标的基础上, 将其分成基于像素的评估方法、基于类内重合度的评估方法、基于边界的评估方法、基于聚类的评估方法和基于实例的评估方法五大类. 在材料显微图像分析的应用背景下, 通过实验讨论了不同分割方法和不同典型噪声在不同评估方法中的表现. 最终, 讨论了各种评估方法的优势和适用性.

关键词 计算机视觉; 图像分割; 图像处理; 评估方法; 材料显微图像

分类号 TP3

Image segmentation metric and its application in the analysis of microscopic image

MA Bo-yuan^{1,2,3,4}, JIANG Shu-fang⁵, YIN Dou³, SHEN Hao-kai⁶, BAN Xiao-juan^{1,2,3,4}, HUANG Hai-you^{1,7,8}✉, WANG Hao^{1,9},
XUE Wei-hua^{9,10}, FENG Hua³

1) Beijing Advanced Innovation Center for Materials Genome Engineering, University of Science and Technology Beijing, Beijing 100083, China
2) Beijing Key Laboratory of Knowledge Engineering for Materials Science, University of Science and Technology Beijing, Beijing 100083, China
3) School of Computer and Communication Engineering, University of Science and Technology Beijing, Beijing 100083, China
4) Institute of Artificial Intelligence, University of Science and Technology Beijing, Beijing 100083, China
5) Department of Obstetrics and Gynecology, General Hospital of PLA, Beijing 100853, China
6) College of Information Science and Engineering, China University of Petroleum Beijing, Beijing 102249, China
7) Institute for Advanced Materials and Technology, University of Science and Technology Beijing, Beijing 100083, China
8) Shunde Graduate School, University of Science and Technology Beijing, Foshan 528300, China
9) School of Materials Science and Engineering, University of Science and Technology Beijing, Beijing 100083, China
10) School of Materials Science and Engineering, Liaoning Technical University, Fuxin 123099, China

✉ Corresponding author, E-mail: huanghy@mater.ustb.edu.cn

ABSTRACT Material microstructure data are an important type of data in building intrinsic relationships between compositions, structures, processes, and properties, which are fundamental to material design. Therefore, the quantitative analysis of microstructures is

收稿日期: 2020–05–28

基金项目: 海南省财政科技计划资助项目(ZDYF2019009); 北京科技大学顺德研究生院科技创新专项资金资助项目(BK19BE030)

essential for effective control of the material properties and performances of metals or alloys in various industrial applications. Microscopic images are often used to understand the important structures of a material, which are related to certain properties of interest. One of the key steps during material design process is the extraction of useful information from images through microscopic image processing using computational algorithms and tools. For example, image segmentation, which is a task that divides the image into several specific and unique regions, can detect and separate each microstructure to quantitatively analyze its size and shape distribution. This technique is commonly used in extracting significant information from microscopic images in material structure characterization field. With great improvement in computing power and methods, a large number of image segmentation methods based on different theories have made great progress, especially deep learning-based image segmentation method. Therefore selecting an appropriate evaluation method to assess the accuracy and applicability of segmentation results to properly select the optimal segmentation methods and their indications on the direction of future improvement is necessary. In this work, 14 evaluation metrics of image segmentation were summarized and discussed. The metrics were divided into five categories: pixel, intra class coincidence, edge, clustering, and instance based. In the application of material microscopic image analysis, we collected two classical datasets (Al-La alloy and polycrystalline images) to conduct quantitative experiment. The performance of different segmentation methods and different typical noises in different evaluation metrics were then compared and discussed. Finally, we discussed the advantages and applicability of various evaluation metrics in the field of microscopic image processing.

KEY WORDS computer vision; image segmentation; image processing; segmentation evaluation metrics; material microscopic image

计算机视觉是一门“教”会计算机如何去“看”世界的科学,旨在使用计算机及相关设备对生物视觉进行模拟,进而代替眼睛对数字图像中的目标进行处理、识别、分析和理解^[1]。图像分割是从图像中提取感兴趣区域的方法,是计算机视觉的重要组成部分^[2]。在实际应用中,对图像分割结果的正确评估是分割方法完善的重要研究手段^[3]。

在材料科学领域,随着材料信息学的发展,大量先进的信息学技术在材料领域获得应用,并展现出蓬勃的生机和快速发展的态势。材料图像,特别是材料显微图像,是材料组织信息的重要载体。使用图像分割等计算机视觉方法可提取材料图像中的组织信息,是高效建立材料成分-工艺-组织-性能关系的重要环节。

本文在材料显微图像分析的应用背景下,重

点综述了图像分割的评估方法,并讨论了不同评估方法的适用性。

1 图像分割的定义及其在材料显微图像分析中的应用

图像分割是将图像分成若干个特定的、具有独特性质区域的技术^[4],是计算机视觉的重要分支。分割结果可进一步为图像识别和图像描述等图像理解相关任务提供信息支撑。

目前,随着多学科间的深入交叉和融合,科学工作者开始尝试将以图像分割为代表的计算机视觉技术运用到表面缺陷检测、金相组织分析、神经细胞拓扑分析等材料 and 生物医学领域^[5-6]。如图 1 所示为材料显微图像分割流程图,针对原始的材料显微图像,采用图像分割方法 (Image segmenta-

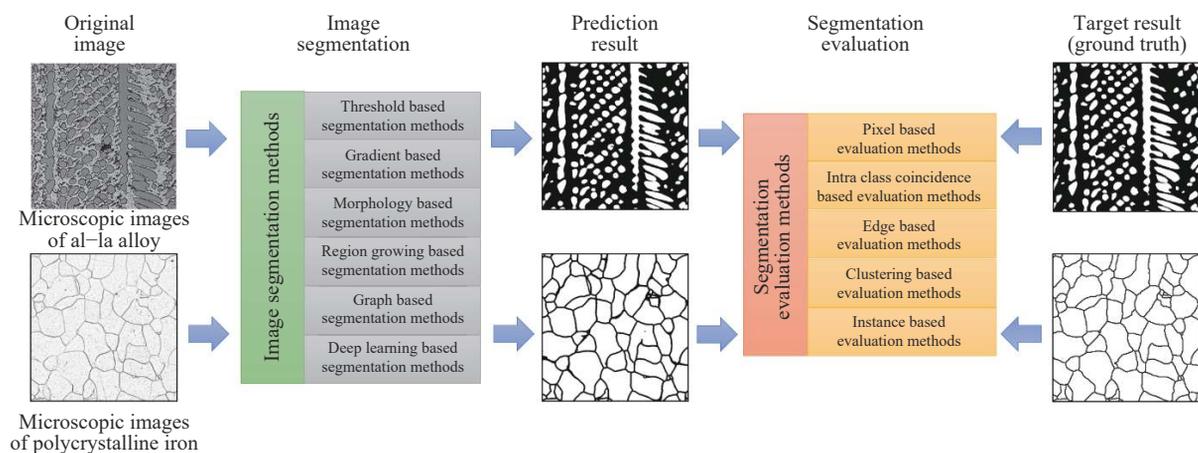


图 1 材料显微图像分割流程示意

Fig.1 Flow chart of material microscopic image segmentation

tion methods) 提取图像中科研人员感兴趣的组织结构区域^[7], 如图 1 中的白色区域对应的材料组织(枝晶和晶粒)。通过定量分析材料内部的微观组织特征(如尺寸和形状分布等), 可进一步解构材料微观组织特征与材料成分、制备工艺、宏观性能之间的内禀关系^[8], 进而辅助材料性能的优化和新材料的研发^[9]。

随着计算机硬件以及以深度学习^[10]为代表的机器学习技术的进步, 多种基于不同理论的图像分割方法如雨后春笋般涌现, 如图 1 中基于阈值的分割方法(Threshold based segmentation methods)^[11]、基于梯度的分割方法(Gradient based segmentation methods)^[12]、基于形态学的分割方法(Morphology based segmentation methods)^[13]、基于区域生长的分割方法(Region growing based segmentation methods)^[14]、基于聚类的分割方法(Clustering based segmentation methods)^[15]、基于图论的分割方法(Graph based segmentation methods)^[16]和基于深度学习的分割方法(Deep learning-based segmentation methods)^[17-18]等, 上述方法被广泛应用于材料显微图像分析任务^[19-21]。为了进一步理解各种图像分割方法的适用性以及改进分割质量, 本文对不同方法的分割结果进行了系统地评估。虽然人的视觉判断是最直接的判断方法, 但是人的主观不确定性会导致算法评估难以量化。因此, 采用自动的客观的评估方法成为图像分割结果评估的主流方式。当前存在多种基于不同理论的图像分割客观评估方法(Segmentation evaluation methods), 每种方法均从特定的角度评估分割结果的优劣, 可应用于不同目标的图像分割任务中。

本文综述了 14 种基于不同理论的图像分割客观评估指标, 将其分成五大类方法分别展开叙述(图 1): 基于像素的评估方法(Pixel based evaluation methods)、基于类内重合度的评估方法(Intra class coincidence based evaluation methods)、基于边界的评估方法(Edge based evaluation methods)、基于聚类的评估方法(Clustering based evaluation methods)和基于实例的评估方法(Instance based evaluation methods)。并在材料显微图像分析应用中, 选择两种典型材料显微图像分割任务作为分析目标, 对比采用不同分割方法和不同典型噪声下各种评估指标的性能, 进而对不同评估方法的适用性展开讨论。

2 图像分割评估方法

通过现有文献整理和研究, 本文将图像分割

评估指标分为基于像素的评估方法、基于类内重合度的评估方法、基于边缘的评估方法、基于聚类的评估方法和基于实例的评估方法等五个类别。

2.1 基于像素的评估方法

基于像素的评估方法主要包括像素准确率(Pixel accuracy)和平均准确率(Mean accuracy)^[17]两个评估指标。

像素准确率统计了预测正确的像素占总像素的比例^[22], 其定义为:

$$\text{Pixel accuracy} = \frac{\sum_{i=0}^k p_{ii}}{\sum_{i=0}^k \sum_{j=0}^k p_{ij}} \quad (1)$$

在图像分割任务中, 假定有 $k+1$ 个类别(“+1”指包括背景类), 通常使用 p_{ij} 表示真值结果为类 i 而预测结果为类 j 的像素点总数, 同理, p_{ii} 表示真值结果和预测结果均为类 i 的像素点总数, 即预测正确的像素点数目。

平均准确率(Mean accuracy): 是对像素准确率的改进, 分别计算每个类内被正确分类像素数的比例, 进而对所有类求平均值。该方法可缓解当分割任务中不同类别包含像素数目差异较大时, 准确率被包含像素数目较多的类别所主导的问题, 其定义为:

$$\text{Mean accuracy} = \frac{1}{k+1} \sum_{i=0}^k \frac{p_{ii}}{\sum_{j=0}^k p_{ij}} \quad (2)$$

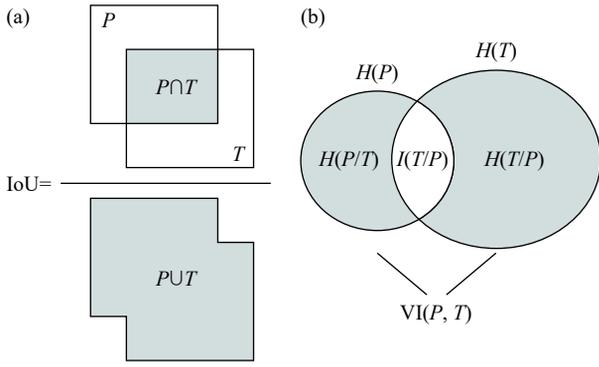
2.2 基于类内重合度的评估方法

基于类内重合度的评估方法主要包括平均交并比(Mean intersection over union, MIoU)^[17], 频权交并比(Frequency weighted intersection over union, FWIoU)^[17]和 Dice 系数(Dice score)^[23]三个评估指标。

平均交并比 MIoU 在计算每个类的交并比的基础上求均值, 可反映预测结果和真值结果的重合度。IoU(或称为 Jacarrd index)的计算方式如图 2(a)所示, 其中 P 代表预测结果(Predicted segmentation), T 代表真值结果(也称目标结果, Target truth), 其思路是计算真值结果和预测结果区域集合的交集和并集之比, 即公式(3)所示。

$$\text{IoU} = \frac{|P \cap T|}{|P \cup T|} \quad (3)$$

具体地, 采用上述描述方法, MIoU 可定义为^[17]:



Note: $VI(P, T)$ means variation of information; $I(T, P)$ denotes mutual information; $H(P/T)$ and $H(T/P)$ denote conditional entropy.

图 2 评估指标示意图。(a)IoU 指标示意图; (b)VI 指标示意图

Fig.2 Schematics of evaluation metrics: (a) IoU metric diagram; (b) VI metric diagram

$$MIoU = \frac{1}{k+1} \sum_{i=0}^k \frac{P_{ii}}{\sum_{j=0}^k P_{ij} + \sum_{j=0}^k P_{ji} - P_{ii}} \quad (4)$$

其中, p_{ji} 表示真值结果为类 j 而预测结果为类 i 的像素点总数。

频权交并比 FWIoU 是对原始的 MIoU 的改进, 即根据每一类出现的频率对各个类的 IoU 进行加权求和, 定义为^[17]:

$$FWIoU = \frac{1}{\sum_{i=0}^k \sum_{j=0}^k P_{ij}} \sum_{i=0}^k \frac{\sum_{j=0}^k P_{ij} P_{ii}}{\sum_{j=0}^k P_{ij} + \sum_{j=0}^k P_{ji} - P_{ii}} \quad (5)$$

Dice 系数 (Dice score) 是图像分割结果衡量的常用指标, 其计算公式为^[23]:

$$Dice = \frac{2|P \cap T|}{|P| + |T|} \quad (6)$$

其中, $|P \cap T|$ 代表预测结果 P 和真值结果 T 之间的重合像素数量, $|P|$ 表示预测结果 P 中的像素数量, 同理 $|T|$ 表示真值结果 T 中的像素数量, 通常 P 、 T 集合仅统计前景区域的像素。Dice score 实质上是两个样本间的重叠度量 (0~1), 其中该指标为 1 代表完全重叠, 表示预测结果较好。Dice score 相比于 IoU 的区别在于分子和分母间均增加 $|P \cap T|$ 的数目。

2.3 基于边界的评估方法

边界检测也称边缘检测, 是指将图像分成边缘区域和非边缘区域。由于在边缘检测任务中, 边缘区域往往仅占据少量像素区域, 并且绘制真值边缘时容易产生误差, 致使预测边缘和真实边缘在像素上通常无法一一对应, 因此该类评估方法主要通过计算预测边缘像素和真实边缘像素间的

距离来反映边缘检测效果。

基于边界的评估方法主要包括 FoM (Figure of merit)^[24]、Completeness^[25]、Correctness^[25] 和 Quality^[25] 4 个评估指标。

FoM (Figure of merit) 计算公式为:

$$FoM = \frac{1}{\max(|T|, |P|)} \sum_{i=1}^{|P|} \frac{1}{\xi d_i^2} \quad (7)$$

其中, 通过人工检视标注得到的真值边缘像素点个数为 $|T|$; 通过检测算法检测的边缘像素点个数为 $|P|$; 检测算法得到的第 i 个边缘像素点和离它最近的真值边缘像素点间的欧氏距离为 d_i ; 常量系数 ξ 一般为 0.1, 表示错位边缘惩罚项。FoM 值越大表明检测出的边缘越靠近人工标注的真值边缘, 即该边缘检测算法的检测精度越高。

Completeness、Correctness 和 Quality 评估方法在基于欧氏距离的基础上重新定义了匹配的概念。其匹配的计算公式为:

$$\mu_B(A) = \{a \in A | \exists b \in B, d(a, b) < \theta\} \quad (8)$$

设两个像素集合 A 和 B , 其中 a 和 b 分别为两个集合中的两个元素, $d(a, b)$ 表示两个元素之间的欧氏距离。 $\mu_B(A)$ 指集合 A 中满足与距集合 B 某一元素的距离小于阈值 θ 的元素的子集。根据上述定义, 分别计算预测边缘像素集合 P 和真值边缘像素集合 T 的 $\mu_T(P)$ 和 $\mu_P(T)$, $\mu_T(P)$ 代表预测结果中与真值结果相匹配的元素的子集, $\mu_P(T)$ 代表真值结果中与预测结果相匹配的元素的子集。如图 3 所示。

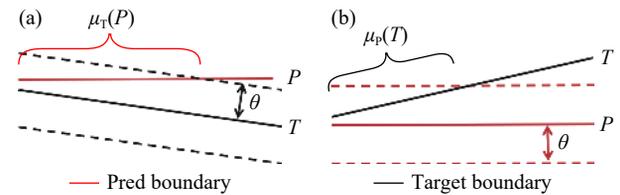


图 3 边缘检测匹配示意图。(a) 令预测边界与真值边界进行匹配; (b) 令真值边界与预测边界进行匹配

Fig.3 Matching schematics of edge detection: (a) matching ground truth with prediction skeleton; (b) matching prediction with ground truth skeleton

在此基础上, Completeness 的计算公式为:

$$Completeness = \frac{|\mu_T(P)|}{|P|} \quad (9)$$

其中, $|P|$ 为预测边缘像素集合 P 中的边缘像素个数。同理, $|\mu_T(P)|$, $|\mu_P(T)|$ 和 $|T|$ 分别代表各自集合中的边缘像素个数。

Correctness 的计算公式为:

$$Correctness = \frac{|\mu_P(T)|}{|T|} \quad (10)$$

Quality 的计算公式为:

$$\text{Quality} = \frac{|\mu_T(P)|}{|P| - |\mu_P(T)| + |T|} \quad (11)$$

2.4 基于聚类的评估方法

图像分割可以看作是一种聚类任务, 将图像中的所有像素分割成 n 个分区或段, 通过对比真值图和预测图中的分区结果可对图像分割任务进行评估. 在基于聚类的评估方法中类的定义和图像类别中类的定义不同, 聚类中每个类具体指图像中得到每个分区, 在图像分割任务中, 某个图像类别可以有多个分区.

基于聚类的评估方法主要包括: 兰德系数 (Rand index, RI)^[26-27]、调整兰德系数 (Adjusted rand index, ARI)^[28] 和信息变化量 (Variation of information, VI)^[29] 3 个评估指标.

给定一个含有 n 个元素的集合 S 的两种划分, S 划分为 r 个子集: $T = \{T_1, T_2, \dots, T_r\}$ 和 S 划分为 s 个子集: $P = \{P_1, P_2, \dots, P_s\}$. 在图像分割任务中, S 代表像素集合, T 代表真值分割结果, P 代表预测分割结果. 通常, 采用联通区域方法 (Connected component) 分别将 T 和 P 划分成不同的分区 (子集).

进一步, 定义:

a 为 S 集合中在 T 中属于同一子集, 在 P 中也属于同一子集的元素对数; b 为 S 集合中在 T 中属于不同子集, 在 P 中也属于不同子集的元素对数; c 为 S 集合中在 T 中属于同一子集, 但在 P 中属于不同子集的元素对数; d 为 S 集合中在 T 中属于不同子集, 但在 P 中属于同一子集的元素对数;

则 RI 定义为:

$$\text{RI} = \frac{a+b}{a+b+c+d} = \frac{a+b}{\binom{n}{2}} = \frac{a+b}{n(n-1)/2} \quad (12)$$

直观地, $a+b$ 可被认为是 T 和 P 之间一致的元素对个数, $c+d$ 可被认为是 T 和 P 之间不一致的元素对个数.

虽然 RI 的取值范围为 $[0,1]$, 但该指标不能保证在类别标签是随机分配的情况下其值接近 0. 由于 RI 基于像素点对数目进行评估, 而在材料显微图像分割应用中, 大多数像素点对不相连, 即公式 (12) 中的 b 非常高, 导致不同分割方法的评估结果均接近 1, 造成区分度较低. 因此, ARI 在 RI 的基础上进行了修正, 令 $n_{ij} = |T_i \cap P_j|$ 来表示 T_i 与 P_j 之间交集的元素, 基于聚类任务的列联表如表 1 所示. 其计算公式如式 (13) 所示.

表 1 基于聚类任务的列联表

Table 1 Contingency table					
Union	P_1	P_2	...	P_s	Sums
T_1	n_{11}	n_{12}	...	n_{1s}	a_1
T_2	n_{21}	n_{22}	...	n_{2s}	a_2
...
T_r	n_{r1}	n_{r2}	...	n_{rs}	a_r
Sums	b_1	b_2	...	b_s	

$$\text{ARI} = \frac{\sum_{i=1}^r \sum_{j=1}^s \binom{n_{ij}}{2} - \left[\sum_{i=1}^r \binom{a_i}{2} \sum_{j=1}^s \binom{b_j}{2} \right] / \binom{n}{2}}{\frac{1}{2} \left[\sum_{i=1}^r \binom{a_i}{2} + \sum_{j=1}^s \binom{b_j}{2} \right] - \left[\sum_{i=1}^r \binom{a_i}{2} \sum_{j=1}^s \binom{b_j}{2} \right] / \binom{n}{2}} \quad (13)$$

ARI 的取值范围为 $[-1,1]$, 值越大, 代表预测结果中的分区结果更接近与真值结果中的分区结果.

VI 公式如 (14) 所示, 将预测结果 P 和真值结果 T 看成两种数据分布, 利用信息熵计算两者之间的分布变化. 核心是计算预测和真实标注图之间的条件熵, 如图 2(b) 所示.

$$\text{VI}(P, T) = H(P|T) + H(T|P) \quad (14)$$

其中, $H(P|T)$ 代表过分割错误, 过分割即错误地将一个区域分割成两个区域, 也被称为拆分错误 (Split error, SE). $H(T|P)$ 代表欠分割错误, 即将两个区域错误地合并成一个区域, 也被称为合并错误 (Merge error, ME). VI 代表总错误指标, 为 ME 和 SE 的和, 值越小, 代表分割结果中存在的错误越少, 即分割结果越准确. 并且, 随着误差的变化, VI 呈现线性变化, 而 RI 呈现二次方的变化, 因此, VI 更适用于作为距离指标反映误差的变化^[30].

2.5 基于实例的评估方法

基于实例的评估方法主要包括平均均值精确率 (Mean average precision, MAP)^[31-32] 和基数差 (Cardinality difference, CD)^[33] 两个评估指标.

MAP 是图像分割和物体检测任务中的经典评估指标, 与基于聚类的评估方法类似, 该方法首先使用联通区域方法分别将真值分割结果和预测分割结果分成多个分区, 进而在每个预测分区上逐个与所有的真值分割结果的分区计算 IoU, IoU 的阈值范围从 0.5 到 0.95, 步长为 0.05, 即在阈值 t 为 (0.5, 0.55, 0.6, 0.65, 0.7, 0.75, 0.8, 0.85, 0.9, 0.95) 情形下分别统计匹配的数目. 在 MAP 计算过程中, 匹配指某一预测结果中的分区与真值结果中

的分区 IoU 大于特定的阈值. 其中 True positive (TP) 表示真实和预测结果中正确匹配的分区对数量, False positive (FP) 表示预测分割结果中未正确匹配的分区数量, False negative (FN) 表示真值分割结果中未正确匹配的分区数量.

在每个阈值 t 上, 根据预测分割的分区结果与真值分割的分区结果比较产生的 TP、FN、和 FP 的数量计算出一个精度值, 并除以阈值个数获得均值精确率. 在此基础上, 对所有图像 n 计算各自均值精确率再求平均, 获得最终平均均值精确率, 公式如下所示:

$$MAP = \frac{1}{n \times |t|} \sum_t \frac{TP(t)}{TP(t) + FP(t) + FN(t)} \quad (15)$$

CD 是 Waggoner 等引入的一种比较分割结果中分区个数的评估指标^[33], 其定义为:

$$CD = Num^T - Num^P \quad (16)$$

式中, Num^T 代表真实结果中的实例个数, Num^P 代表预测结果中的实例个数, 如果 CD 是正数, 表示欠分割, 如果 CD 是负数, 表示过分割.

综上, 各评估指标的简要信息如表 2 所示, 其中 \uparrow 代表值越大性能越高, \downarrow 代表值越低性能越高, N 代表图像中像素个数.

表 2 各指标的简要概括

Table 2 Brief description of different evaluation methods

Properties	Pixel based evaluation methods		Intra class coincidence based evaluation methods			Edge based evaluation methods		
	Pixel accuracy	Mean accuracy	MIoU	FWIoU	Dice score	Figure of merit	Completeness	Correctness
Value range	[0, 1]	[0, 1]	[0, 1]	[0, 1]	[0, 1]	[0, 1]	[0, 1]	[0, 1]
tendency	\uparrow	\uparrow	\uparrow	\uparrow	\uparrow	\uparrow	\uparrow	\uparrow

Properties	Edge based evaluation methods		Clustering based evaluation methods			Instance based evaluation methods		
	Quality	RI	ARI	VI		CD	MAP	
				ME	SE	VI		
Value range	[0, 1]	[0, 1]	[-1, 1]	$[0, \log_2 N]$	$[0, \log_2 N]$	$[0, \log_2 N]$	$(-\infty, +\infty)$	[0, 1]
tendency	\uparrow	\uparrow	\uparrow	\downarrow	\downarrow	\downarrow	Closer to 0, the better	\uparrow

3 图像分割评估方法对比实验

本章以两种典型材料显微图像分割任务为应用背景, 定量地衡量图像分割评估方法在不同分割算法和不同噪声下的性能. 本文所有代码及示例图像可开源获取, 详见如下网址: <https://github.com/Keep-Passion/ImageSegMetrics>.

3.1 不同图像分割算法下的评估结果

为了降低图像分割算法对评估方法的影响, 本文选择了 6 种最为经典的图像分割算法进行比较: 基于阈值的图像分割算法 OTSU^[11], 基于梯度的边缘提取算法 Canny^[34], 基于形态学的图像分割算法 Watershed^[35], 基于聚类的图像分割方法 K-means^[36], 基于图论的图像分割算法 Random Walker^[37] 和基于深度学习的图像分割算法 Unet^[18].

本文选择多晶纯铁 (Polycrystalline iron) 晶粒组织图像和铝镧合金 (Al-La alloy) 枝晶组织图像作为目标数据集, 数据集参数如表 3 所示. 两数据集示例如图 1 所示, 其中多晶纯铁晶粒的前景晶粒区域像素个数远高于背景边界区域的像素个数, 属于类别不平衡的图像分割任务. 铝镧合金枝

晶图像的前景组织区域像素个数与背景区域的像素个数相当, 属于类别平衡的图像分割任务. 在实践过程中, 类别不平衡的图像分割任务更加困难.

表 3 材料显微图像数据集参数

Table 3 Description of two material micrographic image datasets

ID	Microstructure	Image size	Image number
1	Polycrystalline iron	1024×1024	296
2	Al-La alloy	1024×1024	50

由于本文选择基于深度学习的图像分割模型 Unet 作为不同评估指标的基线模型, 而 Unet 需要样本训练模型才可进一步使用. 因此将数据集按照 7 : 1 : 2 的方式划分为训练集、验证集和测试集. 为保证实验的公平性, 所有图像分割算法均在测试集上进行评估. 对于 Unet 模型, 模型在训练集上训练, 并在每轮迭代中使用验证集验证, 选择验证集损失最低的网络参数在测试集上进行评估. 其中, Unet 网络训练参数为: Batch size 为 10, 初始学习率为 1×10^{-4} , 在 Adam 优化器下训练 50 轮. 所有实验均在装备一台 Nvidia Tesla V100 显卡的

工作站上测试。

多晶纯铁晶粒组织图像在不同分割算法下的各评估结果如表 4 所示,其可视化示例如图 4 上两

行所示。铝镧合金枝晶组织图像在不同分割算法下的各评估结果如表 5 所示,其可视化示例如图 4 下两行所示。

表 4 多晶纯铁晶粒组织图像不同分割算法下评估结果

Table 4 Evaluation results under different segmentation algorithms for polycrystalline iron image

Segmentation algorithm	Pixel based evaluation methods		Intra class coincidence based evaluation methods			Edge based evaluation methods		
	Pixel accuracy	Mean accuracy	MIoU	FWIoU	Dice score	Figure of merit	Completeness	Correctness
OTSU	0.9443	0.7800	0.7226	0.8979	0.9696	0.6593	0.8298	0.9146
Canny	0.9145	0.6364	0.5811	0.8468	0.9540	0.4085	0.7007	0.9156
Watershed	0.9017	0.5613	0.5109	0.8236	0.9476	0.2009	0.4516	0.6537
K-means	0.5739	0.5469	0.4331	0.5307	0.5771	0.4906	0.8598	0.5796
Random walker	0.9447	0.7925	0.7293	0.8994	0.9697	0.6963	0.8445	0.9059
Unet	0.9311	0.9423	0.7510	0.8898	0.9605	0.8933	0.9784	0.8562

Segmentation algorithm	Edge based evaluation methods		Clustering based evaluation methods				Instance based evaluation methods	
	Quality	RI	ARI	VI		CD	MAP	
				ME	SE			
OTSU	0.7361	0.7187	0.0582	4.3330	0.0218	4.3548	178.8333	0.1593
Canny	0.5482	0.1356	-0.0053	7.2893	0.0084	7.2978	199.0000	0.0333
Watershed	0.2164	0.9444	0.1552	1.7498	0.3439	2.0937	29.5667	0.1592
K-means	0.4751	0.4687	0.0273	3.8777	0.5202	4.3979	-234.8333	0.0932
Random walker	0.7477	0.7651	0.1087	3.8414	0.0262	3.8676	164.1667	0.1789
Unet	0.8450	0.9819	0.5917	0.0504	0.0722	0.1226	-33.1333	0.5465

由于多晶纯铁晶粒组织图像分割任务的目标是提取并分离每个晶粒,因此,该任务可被解读为晶粒分割或晶界提取两种不同的描述。因此,对于后者,可采用基于边界的评估方法对晶界提取的准确性进行评估。

由于基于深度学习的机器学习算法可在训练过程中自动学习到数据中适用于目标任务的强有效特征,因此其分割性能远高于传统图像处理方法。由图 4 直观比较,基于深度学习的图像分割模型 Unet 在两种数据集上的可视化效果均显著超过了其余经典算法。但在多晶纯铁晶粒组织图像的客观评估中(表 4),Unet 和传统方法(除 Unet 外的其他方法)在基于像素和基于类内重合度的评估方法的数值指标上相差较小。在铝镧合金枝晶组织图像的客观评估中(表 5),传统方法和 Unet 在基于像素的评估方法的数值指标上相差较小。基于像素的评估方法逐像素地比较预测分割结果和真值分割结果的区别,而基于类内重合度的评估方法逐类别的比较两者之前的差异,均未考虑材料显微组织的拓扑特性,因此,本文认为在材料显微图像分析及定量表征应用中,该评估方法适用

于目标为统计物相组成比例等的表征任务,但不适用于需要分析单个组织特征的尺寸或形状统计等的表征任务。

基于边界的评估方法仅适用于基于边缘提取的图像分割任务,其泛用性较低。同时晶界提取的目标在于分离每个晶粒,基于边界的评估方法并未考虑晶界不闭合的情况。因此,本文认为在材料显微图像分析应用中,基于边界的评估方法可适用于缺陷(如裂纹)检测^[25],但不适用组织结构定量表征任务。

基于聚类和基于实例的评估方法由于在执行过程中需要先使用联通区域方法对分割结果进行分区操作,考虑了显微组织结构的拓扑特性,因此本文认为这两类方法更适用于材料显微图像分析任务中统计尺寸和形状分布等微观组织定量表征的目标。

在基于聚类的评估方法中,RI 被证明在最差情况下也难以达到最低值,造成各方法间差距过小,如多晶纯铁晶粒图像分割结果(表 4)中的 Watershed(RI=0.9444)和 Unet(RI=0.9819)。ARI 是对 RI 指标的改进,各方法间该指标的差距较大。

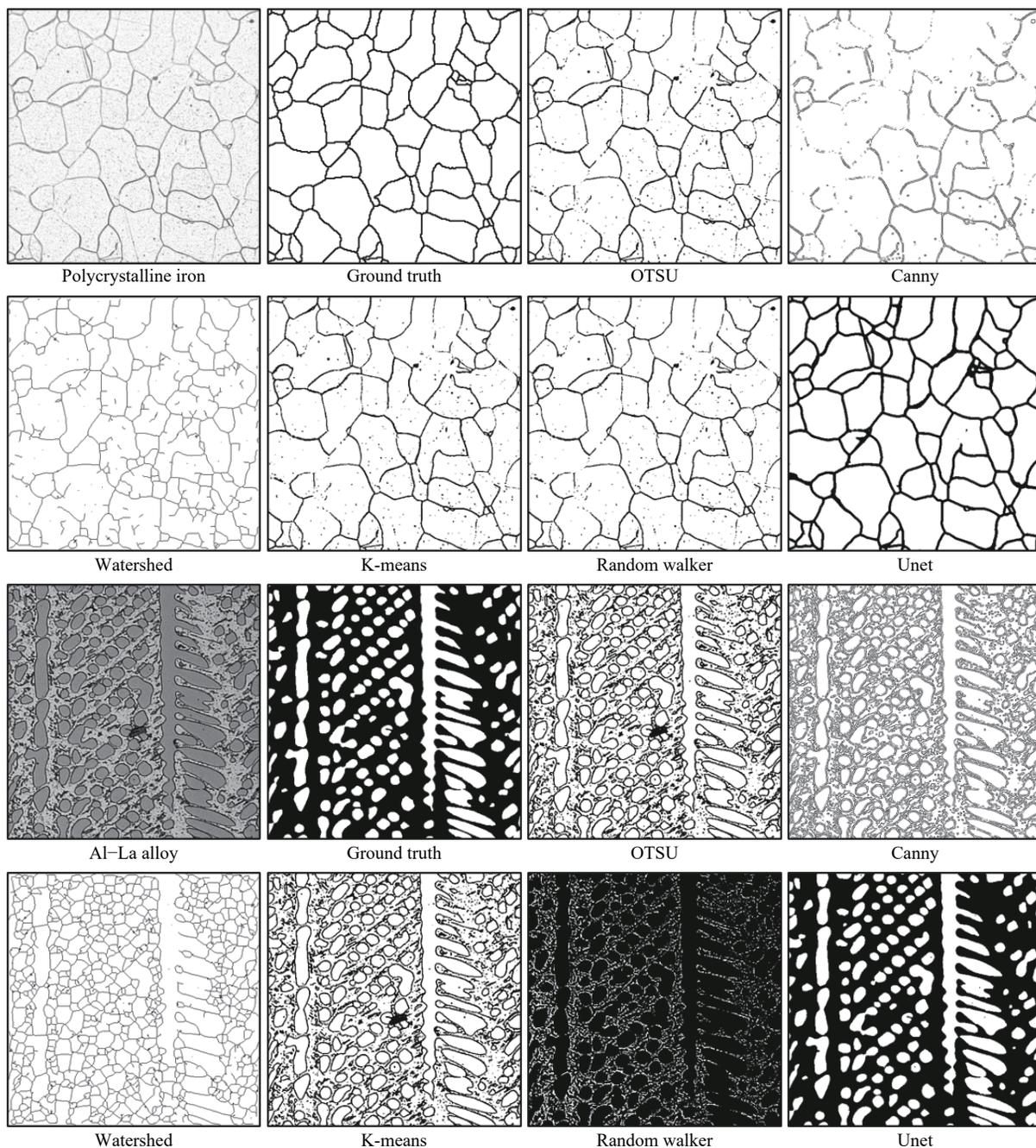


图 4 多晶纯铁晶粒组织及铝镧合金枝晶组织图像在不同分割算法结果的可视化对比

Fig.4 Visualization results of different segmentation methods for polycrystalline iron and Al-La alloy microscopic image

VI 指标进一步分为合并错误 (Merge Error, ME) 和分离错误 (Split Error, SE), 可更加全面的分析算法的分割结果.

在基于实例的评估算法中, CD 指标易受到噪声影响, 且该指标不存在上下限, 造成各方法的评估结果差距过大, 难以直观评估分割结果, 如在铝镧合金枝晶图像分割任务中, CD 指标从 -10 到 -10918 产生变化, 因此本文认为该指标不适合材料显微图像分析任务. MAP 的值域范围是 [0,1], 且在不同交并比阈值下综合评估分割结果, 性能

较为准确, 但也同样容易受到微小噪声的影响.

3.2 引入不同噪声下评估方法效果

为进一步加深对图像分割评估方法的理解, 本文在图像中增加固定数目的典型噪声, 以展示各图像评估方法的鲁棒性.

本文从多晶纯铁晶粒图像和铝镧合金枝晶图像中分别选择 500×500 像素尺寸的图像, 并通过控制变量, 随机引入三种在材料显微图像分析任务中常见的噪声^[38], 每种噪声各具有 500 像素的改变量: 随机斑点噪声、划痕噪声和边界模糊或消失现象.

表5 铝镧合金枝晶组织图像不同分割算法下评估结果

Table 5 Evaluation of different segmentation results for Al-La microscopic image

Segmentation algorithm	Pixel based evaluation methods		Intra class coincidence based evaluation methods			Clustering based evaluation methods
	Pixel accuracy	Mean accuracy	MIoU	FWIoU	Dice score	RI
OTSU	0.6263	0.7025	0.4538	0.4441	0.6573	0.6981
Canny	0.5259	0.6126	0.3497	0.3315	0.5890	0.4780
Watershed	0.4199	0.5426	0.2373	0.1974	0.5557	0.5902
K-means	0.5078	0.5098	0.3287	0.3482	0.3434	0.5210
Random walker	0.5110	0.4027	0.2559	0.3249	0.0024	0.3552
Unet	0.9850	0.9854	0.9684	0.9706	0.9796	0.9810

Segmentation algorithm	Clustering based evaluation methods				Instance based evaluation methods	
	ARI	VI			CD	MAP
		ME	SE	VI		
OTSU	0.3000	0.3075	0.0135	0.3210	-1680.9000	0.2170
Canny	-0.0237	6.0105	0.0054	6.0158	-1467.1000	0.0459
Watershed	0.0022	0.5689	0.2377	0.8066	-1014.4000	0.0131
K-means	0.0064	0.8567	0.2247	1.0814	-2216.2000	0.1102
Random walker	-0.1720	0.0000	2.2680	2.2680	-10918.3000	0.0000
Unet	0.9604	0.0191	0.0182	0.0373	-10.7000	0.8453

(1) 随机斑点噪声 (Random noises): 在制备过程中, 需要在材料表面涂抹锈蚀剂, 才可在抛光过程中磨掉材料表面区域, 进而获得材料的微观形貌. 若采用系列截面法逐层观察材料组织形貌, 则需要频繁的使用锈蚀剂, 在抛光过程中, 锈蚀剂不可避免地残留在微观组织表面, 形成随机斑点噪声, 如图5中黑色箭头所示.

(2) 划痕噪声 (Scratch noises): 如前所述, 在制备过程中, 需要对材料表面进行研磨. 该制备方法不可避免地在显微组织表面留下划痕噪声. 划痕噪声通常为黑色直线. 由于其与多晶纯铁晶粒图像中晶界区域在像素值上相似, 容易被误认为是晶界从而会影响分割结果, 如图5中红色箭头所示.

(3) 边界模糊或消失现象 (Blurred or missing bound-

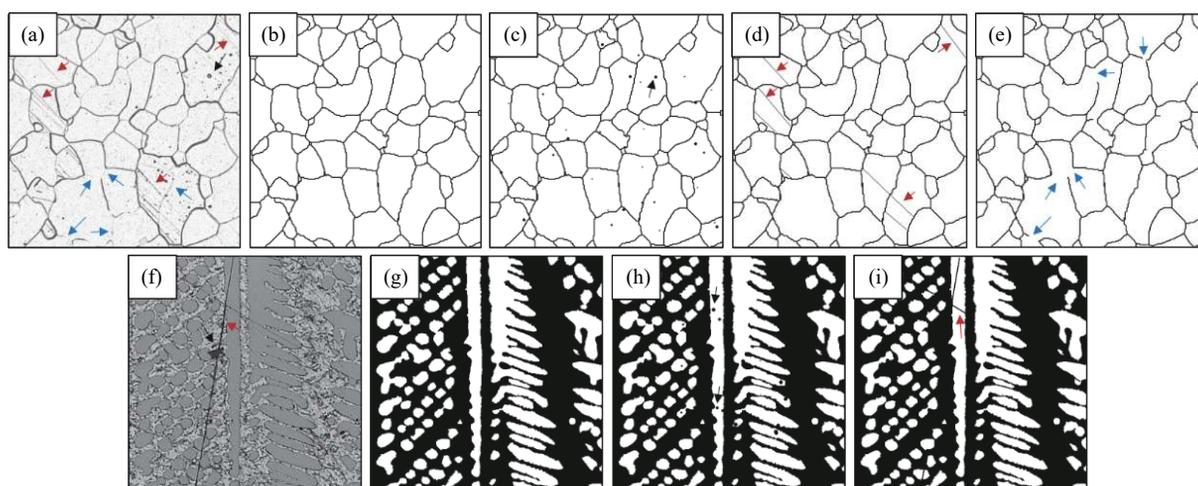


图5 两种图像数据引入不同种类噪声的结果. (a)多晶纯铁晶粒图像; (b)图(a)的真值结果; (c)在(b)中随机引入 500 像素的噪声点; (d)在(b)中引入 500 像素的划痕噪声; (e)在(b)中引入 500 像素的消失晶界噪声; (f)铝镧合金枝晶组织图像; (g)图(f)的真值结果; (h)在(g)中随机引入 500 像素的噪声点; (i)在(g)中引入 500 像素的划痕噪声

Fig.5 Two microscopic images with different noises: (a) polycrystalline iron; (b) ground truth of (a); (c) random noises with 500 pixels in (b); (d) scratch noises with 500 pixels in (b); (e) missing boundaries with 500 pixels in (b); (f) Al la alloy; (g) ground truth of (f); (h) random noises with 500 pixels in (g); (i) scratch noises with 500 pixels in (g)

daries): 在制备过程中, 除锈蚀和研磨外, 需使用一定比例的浸蚀液对材料表面进行处理, 才可浸蚀出晶界或相界(后文以晶界为例), 为防止形成双晶界或粗晶界缺陷, 浸蚀时间不宜过长, 因此容易出现浸蚀不完全现象. 具体地表现为晶界区域没有完全浸蚀, 致使出现晶界模糊或晶界消失现象. 由于晶界消失后的表现特征与晶粒内部区域相似, 容易被误认为是晶粒区域从而影响分割结果, 如图 5 中蓝色箭头所示.

在本实验中, 只有斑点噪声属于随机分布, 而划痕噪声和缺失晶界噪声是根据原图的纹理信息和传统方法的错误识别结果在标注图中相应增加而形成. 通过固定噪声数目, 可以直观地展示不同评估方法的区别. 比如对于基于像素的评估方法并不考虑分割结果的拓扑变化, 相同噪声数目但不同噪声类型的评估结果基本相同. 对于基于实例和基于聚类的评估方法考虑了显微组织的拓扑结构, 相同噪声数目但不同噪声类型的评估结果具有很大差异.

随机斑点噪声由于面积较小且较为分散, 不影响显微组织拓扑结构, 并且可通过设置最小连通区域的面积阈值进而消除该噪声. 虽然 500 个像素只占一幅 500×500 尺寸图像的 $1/500$, 但是划痕噪声和模糊或消失晶界噪声改变了分割结果的拓扑结构: 划痕噪声会造成分割结果中的分区错误(即将一个晶粒错误地分离成两个晶粒), 严重影响显微组织尺寸及形状统计. 模糊或消失晶界同样会影响多晶纯铁晶粒图像中的微观组织表征

(即将两个晶粒错误地合并成一个晶粒). 因此, 本文期望在划痕噪声和模糊或消失晶界噪声下, 评估指标出现明显的退化, 以正确反映图像分割结果.

多晶纯铁晶粒图像在不同噪声下的各指标评估结果如表 6 所示, 铝镧合金枝晶图像在不同噪声下的各指标评估结果如表 7 所示. 两表中每个数据项括号中的数值表示变化的幅值. 基于像素、基于类内重合度和基于边界的评估方法在三种噪声影响下变化较小(变化幅值均在百分位或千分位), 且性能均较高, 不符合本文的要求. 而基于实例和基于聚类的评估指标的变化幅度在十分位以上, 相较于基于像素, 基于类内重合度和基于边界的评估方法更能有代表性地反映分割结果的退化情况. 基于实例的评估方法中, RI 的变化幅度较小, ARI 虽然在模糊或消失晶界处变化较大, 但在划痕噪声情况变化幅度较小. VI 可通过合并错误(ME)和分离错误(SE)综合评估分割结果. 在划痕噪声影响下, 其分离错误出现了大幅提高; 在模糊和缺失晶界噪声影响下, 其合并错误出现了大幅提高; MAP 在多晶纯铁晶粒图像中在上述两种噪声影响下出现大幅度变化, 变化幅度与 VI 相当, 但在铝镧合金枝晶图像中, 在模糊或消失晶界噪声下其变化幅度相对 VI 指标较低.

4 讨论

本文综述了 14 种图像分割的评估指标, 并将其分为五大类: 基于像素点的评估方法、基于类内重合度的评估方法、基于边界的评估方法、基于

表 6 多晶纯铁晶粒图像在不同噪声下各评估方法的结果

Table 6 Results of different evaluation methods for polycrystalline iron image under different noises

Noise type	Pixel based evaluation methods		Intra class coincidence based evaluation methods			Edge based evaluation methods		
	Pixel accuracy	Mean accuracy	MIoU	FWIoU	Dice score	Figure of merit	Completeness	Correctness
Random noises	0.9980 (-0.0020)	0.9989 (-0.0011)	0.9833 (-0.0167)	0.9961 (-0.0039)	0.9989 (-0.0011)	0.9790 (-0.0011)	1.0000 (-0.0000)	0.9737 (-0.0263)
Scratch noises	0.9980 (-0.0020)	0.9989 (-0.0011)	0.9833 (-0.0167)	0.9961 (-0.0039)	0.9989 (-0.0011)	0.9739 (-0.0261)	1.0000 (-0.0000)	0.9702 (-0.0298)
Missing boundaries	0.9964 (-0.0036)	0.9713 (-0.0287)	0.9694 (-0.0306)	0.9929 (-0.0071)	0.9981 (-0.0019)	0.9426 (-0.0574)	0.9465 (-0.0535)	1.0000 (-0.0298)
Noise type	Edge based evaluation methods		Clustering based evaluation methods				Instance-based evaluation methods	
	Quality	RI	ARI	VI		VI	CD	MAP
			ME	SE				
Random noises	0.9737 (-0.0263)	0.9997 (-0.0003)	0.9940 (-0.0060)	0.0000 (+0.0000)	0.0002 (+0.0002)	0.0002 (+0.0002)	-3.0000 (-3.0000)	0.9629 (-0.0371)
Scratch noises	0.9702 (-0.0298)	0.9964 (-0.0036)	0.9322 (-0.0678)	0.0000 (+0.0000)	0.2045 (+0.2045)	0.2045 (+0.2045)	-14.0000 (-14.0000)	0.7795 (-0.2205)
Missing boundaries	0.9463 (-0.0537)	0.9844 (-0.0156)	0.7743 (-0.2257)	0.4990 (+0.4990)	0.0000 (+0.0000)	0.4990 (+0.4990)	12.0000 (+12.0000)	0.7606 (-0.2394)

表 7 铝镧合金枝晶图像在不同噪声下各评估方法的结果

Table 7 Results of different evaluation methods for Al La alloy under different noises

Noise type	Pixel based evaluation methods		Intra class coincidence based evaluation methods			Clustering based evaluation methods
	Pixel accuracy	Mean accuracy	MIoU	FWIoU	Dice score	RI
Random noises	0.9980(-0.0020)	0.9974(-0.0026)	0.9958(-0.0042)	0.9960(-0.0040)	0.9974(-0.0026)	0.9974(-0.0026)
Scratch noises	0.9980(-0.0020)	0.9974(-0.0026)	0.9958(-0.0042)	0.9960(-0.0040)	0.9974(-0.0026)	0.9966(-0.0034)

Noise type	Clustering based evaluation methods			Instance based evaluation methods		
	ARI	VI		CD	MAP	
		ME	SE			VI
Random noises	0.9945(-0.0055)	0.0000(+0.0000)	0.0010(+0.0010)	0.0010(+0.0010)	-4.0000(-4.0000)	0.9513(-0.0487)
Scratch noises	0.9929(-0.0071)	0.0000(+0.0000)	0.1004(+0.1004)	0.1004(+0.1004)	-2.0000(-2.0000)	0.9627(-0.0373)

聚类的评估方法和基于实例的评估方法. 并在材料显微图像分析应用中, 通过实验对比了两种数据集在不同分割方法和不同常见噪声类型下各评估指标的评估结果.

由于基于像素的评估方法逐像素地比较预测分割结果和真值分割结果的区别, 而基于类内重合度的方法逐类别地比较两者之间的差异, 均未考虑材料显微组织的拓扑特性, 因此, 本文认为在材料显微图像分析及定量表征应用中, 该评估方法适用于图像分割的目标为统计物相组成比例的表征任务, 但不适用于需要分析单个组织特征的尺寸或形状统计等的表征任务.

基于边界的评估方法仅适用于基于边缘提取的图像分割任务, 其泛用性较低. 同时针对多晶纯铁晶粒分割任务, 晶界提取的目标在于分离每个晶粒, 基于边界的评估方法并未考虑晶界不闭合的情况. 因此, 本文认为在材料显微图像分析应用中, 基于边界的评估方法可适用于表面缺陷(如裂纹)检测, 但不适用组织结构定量表征任务.

基于聚类 and 基于实例的评估方法由于在执行过程中需要先使用联通区域方法对分割结果进行分区, 考虑了显微组织结构的拓扑特性, 因此本文认为这两类方法更适用于材料显微图像分析任务中统计尺寸和形状分布等微观组织定量表征的目标. 同时在引入典型噪声后, 当组织拓扑结构发生巨大变化时, VI 和 MAP 其变化幅度较大, 本文认为上述两个指标更适用于材料显微图像分割任务. 并且, VI 可进一步细分为合并错误和分离错误, 可更加全面地评估分割结果.

5 总结与展望

随着硬件计算能力和计算机算法的发展, 图像分割领域取得了巨大突破, 大量基于不同理论的优

秀的图像分割方法争相问世. 因此, 在应用过程中, 准确地评估各种分割方法的准确性和适用性是科研工作中的必要环节, 同时理解和掌握各种评估方法是改进图像分割方法的途径之一. 本文首先按照不同类别(基于像素的评估方法、基于类内重合度的评估方法、基于边界的评估方法、基于聚类的评估方法和基于实例的评估方法)详细介绍了各评估方法的计算方式, 并在材料显微图像分析的应用背景下, 进行实验讨论了不同分割方法和不同噪声对不同评估方法的影响. 最后探讨了每种方法的优势和适用性. 本文认为在材料显微图像分析及定量表征应用中, 基于像素和基于类内重合度的评估方法适用于图像分割的目标为统计物相组成比例的表征任务, 但不适用于需要分析单个组织特征的尺寸或形状统计等的表征任务; 基于边界的评估方法可适用于表面缺陷(如裂纹)检测任务; 基于聚类和基于实例的评估方法更适用于材料显微图像分割任务. 并且, VI 可进一步细分为合并错误和分离错误, 可更加全面的评估分割结果.

同时, 由于材料显微图像分割任务的目标是获得微观组织的定量表征, 如面积或形状分布等. 在未来的研究中, 可更进一步从材料学目标出发, 如选择 KL 散度(Kullback-Leibler divergence), JS 散度(Jensen-Shannon divergence)或 Wasserstein 距离等方法评估预测结果的面积分布和真实结果的面积分布的相似性.

致谢

本文的计算工作得到了北京材料基因工程高精尖创新中心材料计算平台的支持.

参 考 文 献

[1] Szeliski R. *Computer Vision: Algorithms and Applications*. Springer Science & Business Media, 2011

- [2] Zhang Y Z. *Image Engineering*. 4th Ed. Beijing: Tsinghua University Press, 2018
(章毓晋. 图像工程. 4版. 北京: 清华大学出版社, 2018)
- [3] Xu R. An overview of image segmentation technique and performance evaluation. *J Ningbo Univ Technol*, 2011, 23(3): 76
(徐瑞. 图像分割方法及性能评价综述. *宁波工程学院学报*, 2011, 23(3): 76)
- [4] Gonzalez R C, Woods R E. *Digital Image Processing*. 3rd Ed. Upper Saddle River: Prentice Hall Press, 2008
- [5] Wang J, Xu X L, Liang K Y, et al. Thermal conductivity enhancement of porous shape-stabilized composite phase change material for thermal energy storage applications: a review. *Chin J Eng*, 2020, 42(1): 26
(王静静, 徐小亮, 梁凯彦, 等. 多孔基定形复合相变材料传热性能提升研究进展. *工程科学学报*, 2020, 42(1): 26)
- [6] Cong M, Wu T, Liu D, et al. Prostate MR/TRUS image segmentation and registration methods based on supervised learning. *Chin J Eng*, 2020, 42(10): 1362
(丛明, 吴童, 刘冬, 等. 基于监督学习的前列腺MR/TRUS图像分割和配准方法. *工程科学学报*, 2020, 42(10): 1362)
- [7] Song X Y. Progress on the multi-disciplinary relationship of stereology, image analysis and computational materials science. *Chin J Stereol Image Anal*, 2008, 13(4): 280
(宋晓艳. 体视学, 图像分析与计算材料学之间的关系及进展. *中国体视学与图像分析*, 2008, 13(4): 280)
- [8] Rajan K. Materials informatics: The materials “gene” and big data. *Ann Rev Mater Res*, 2015, 45: 153
- [9] Butler K T, Davies D W, Cartwright H, et al. Machine learning for molecular and materials science. *Nature*, 2018, 559(7715): 547
- [10] LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*, 2015, 521(7553): 436
- [11] Otsu N. A threshold selection method from gray-level histograms. *IEEE Trans Syst Man Cybern*, 1979, 9(1): 62
- [12] Lakshmi S, Sankaranarayanan D V. A study of edge detection techniques for segmentation computing approaches. *Int J Comput Appl*, 2010, 1(1): 35
- [13] Roerdink J B T M, Meijster A. The watershed transform: definitions, algorithms and parallelization strategies. *Fundam Inform*, 2000, 41(1-2): 187
- [14] Adams R, Bischof L. Seeded region growing. *IEEE Trans Pattern Anal Mach Intell*, 1994, 16(6): 641
- [15] Jain A K. Data clustering: 50 years beyond K-means. *Pattern Recognit Lett*, 2010, 31(8): 651
- [16] Ma B Y, Ban X J, Su Y, et al. Fast-FineCut: Grain boundary detection in microscopic images considering 3D information. *Micron*, 2019, 116: 5
- [17] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation // *Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition*. Boston, 2015: 3431
- [18] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation // *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, Cham, 2015: 234
- [19] Li W, Field K G, Morgan D. Automated defect analysis in electron microscopic images. *npj Comput Mater*, 2018, 4: 36
- [20] Azimi S M, Britz D, Engstler M, et al. Advanced steel microstructural classification by deep learning methods. *Sci Rep*, 2018, 8: 2128
- [21] Ma B Y, Ban X J, Huang H Y, et al. Deep learning-based image segmentation for Al-La alloy microscopic images. *Symmetry*, 2018, 10(4): 107
- [22] Powers D M. Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. *J Mach Learn Technol*, 2011, 2(1): 37
- [23] Dice L R. Measures of the amount of ecologic association between species. *Ecology*, 1945, 26(3): 297
- [24] Abdou I E, Pratt W K. Quantitative design and evaluation of enhancement/thresholding edge detectors. *Proc IEEE*, 1979, 67(5): 753
- [25] Mosinska A, Marquez-Neila P, Koziński M, et al. Beyond the pixel-wise loss for topology-aware delineation // *Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City, 2018: 3136
- [26] Yeung K Y, Ruzzo W L. Details of the adjusted rand index and clustering algorithms, supplement to the paper an empirical study on principal component analysis for clustering gene expression data. *Bioinformatics*, 2001, 17(9): 763
- [27] Rand W M. Objective criteria for the evaluation of clustering methods. *J Am Stat Assoc*, 1971, 66(336): 846
- [28] Hubert L, Arabie P. Comparing partitions. *J Classific*, 1985, 2(1): 193
- [29] Meilă M. Comparing clusterings — an information based distance. *J Multivar Anal*, 2007, 98(5): 873
- [30] Nunez-Iglesias J, Kennedy R, Parag T, et al. Machine learning of hierarchical clustering to segment 2D and 3D images. *PLoS ONE*, 2013, 8(8): e71715
- [31] Booz Allen Hamilton. Data Science Bowl[J/OL]. *Kaggle* (2018-05-21)[2020-12-02]. <https://www.kaggle.com/c/data-science-bowl-2018/overview/evaluation>
- [32] Lin T Y, Maire M, Belongie S, et al. Microsoft COCO: Common Objects in Context // *European Conference on Computer Vision*. Zurich, 2014: 740
- [33] Waggoner J, Zhou Y J, Simmons J, et al. 3D materials image

- segmentation by 2D propagation: a graph-cut approach considering homomorphism. *IEEE Trans Image Process*, 2013, 22(12): 5282
- [34] Canny J. A computational approach to edge detection. *IEEE Trans Pattern Anal Mach Intell*, 1986, 8(6): 679
- [35] Vincent L, Soille P. Watersheds in digital spaces: an efficient algorithm based on immersion simulations. *IEEE Trans Pattern Anal Mach Intell*, 1991, 13(6): 583
- [36] Hartigan J, Wong M. Algorithm AS 136: A K-means clustering algorithm. *Appl Stat*, 1979, 28(1): 100
- [37] Grady L. Random walks for image segmentation. *IEEE Trans Pattern Anal Mach Intell*, 2006, 28(11): 1768
- [38] Xue W H. *Three-Dimensional Modeling and Quantitative Characterization of Grain Structure*[Dissertation]. Beijing: University of Science & Technology Beijing, 2017
(薛维华. 晶粒组织的三维模型构建与定量表征研究[学位论文]. 北京: 北京科技大学, 2017)