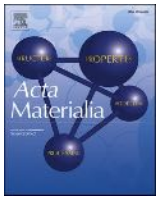
Acta Materialia 290 (2025) 120962

Contents lists available at ScienceDirect

Acta Materialia

journal homepage: www.elsevier.com/locate/actamat



Full length article

A novel training-free approach to efficiently extracting material microstructures via visual large model

Changtai Li [a,b,1], Xu Han [b,1], Chao Yao [c,d,e], Yu Guo [a], Zixin Li [a], Lei Jiang [a], Wei Liu [b], Haiyou Huang [a], Huadong Fu [a,f,*], Xiaojuan Ban [a,b,d,e,f,*]

[a] Beijing Advanced Innovation Center for Materials Genome Engineering, University of Science and Technology Beijing, 100083, Beijing, China
[b] School of Intelligence Science and Technology, University of Science and Technology Beijing, 100083, Beijing, China
[c] School of Computer and Communication Engineering, University of Science and Technology Beijing, 100083, Beijing, China
[d] Beijing Key Laboratory of Big Data Innovation and Application for Skeletal Health Medical Care, 100083, Beijing, China
[e] Key Laboratory of Advanced Materials and Devices for Post-Moore Chips, Ministry of Education, 100083, Beijing, China
[f] Institute of Materials Intelligent Technology, Liaoning Academy of Materials, 110004, Shenyang, China

ARTICLE INFO

Keywords:
Material microstructure
Image recognition
Visual large model
Deep learning

ABSTRACT

The precise quantitative description of material microstructures is essential for deeply exploring the relationship between material composition and property. This significant understanding efficiently enables composition design, process optimization, and property enhancement. Traditionally, the analysis of material microstructures has relied heavily on professional expertise. Even with machine /deep learning (ML/DL)-based analysis methods, substantial expert annotation is required for training, and the trained models often suffer from weak generalizability and poor recognition of new images. This study proposed MatSAM (Materials Segment Anything Model), a novel training-free approach for efficient material microstructure extraction based on the Segment Anything Model (SAM), a type of visual large model (VLM). Integrating region marking and microscopy-adapted points, an automated point-based prompt strategy was developed to achieve accurate and efficient material microstructure recognition. Without any manual annotations, MatSAM precisely identified 11 kinds of metallic material microstructures obtained through various characterization methods. Compared to optimal conventional rule-based methods that do not involve a learning process (non-ML/DL), MatSAM achieved an average relative improvement of 35.4 % in metrics combining the adjusted Rand index (ARI) and Intersection over Union (IoU), outperforming the original SAM by an average of 13.9 %. On four public microstructure segmentation datasets, the IoU of MatSAM showed an average improvement of 7.5 % over corresponding specialist deep models requiring annotations. Meanwhile, MatSAM satisfied the generalization capability of a single model for various microstructures, including grain boundaries, phases, and defects. This approach significantly reduces the labor and computational costs of quantitatively characterizing material microstructures, further accelerating the development of advanced materials.

1. Introduction

Having long been a central pursuit in materials science research, a clear relationship between composition, microstructure, and property is essential for material design, process optimization, and property enhancement [1–5]. Notably, the microstructure of materials acts as a critical link between composition and performance, providing deep insights into material properties [6–8]. Various characterization techniques are employed to obtain images that describe material microstructural information, including optical microscopy (OM), scanning electron microscopy (SEM), transmission electron microscopy (TEM), and X-ray computed tomography (XCT) [9,10].

However, the extraction of key microstructural information—both qualitative and quantitative—from various types of micrographs heavily depends on expert experience and demands huge human and computational resources [11]. An even greater concern is the inconsistency

* Corresponding authors at: Beijing Advanced Innovation Center for Materials Genome Engineering, University of Science and Technology Beijing, 100083, Beijing, China.
E-mail addresses: hdfu@ustb.edu.cn (H. Fu), banxj@ustb.edu.cn (X. Ban).
[1] These authors contributed equally: Changtai Li, Xu Han.

https://doi.org/10.1016/j.actamat.2025.120962
Received 30 September 2024; Received in revised form 7 February 2025; Accepted 20 March 2025
Available online 21 March 2025

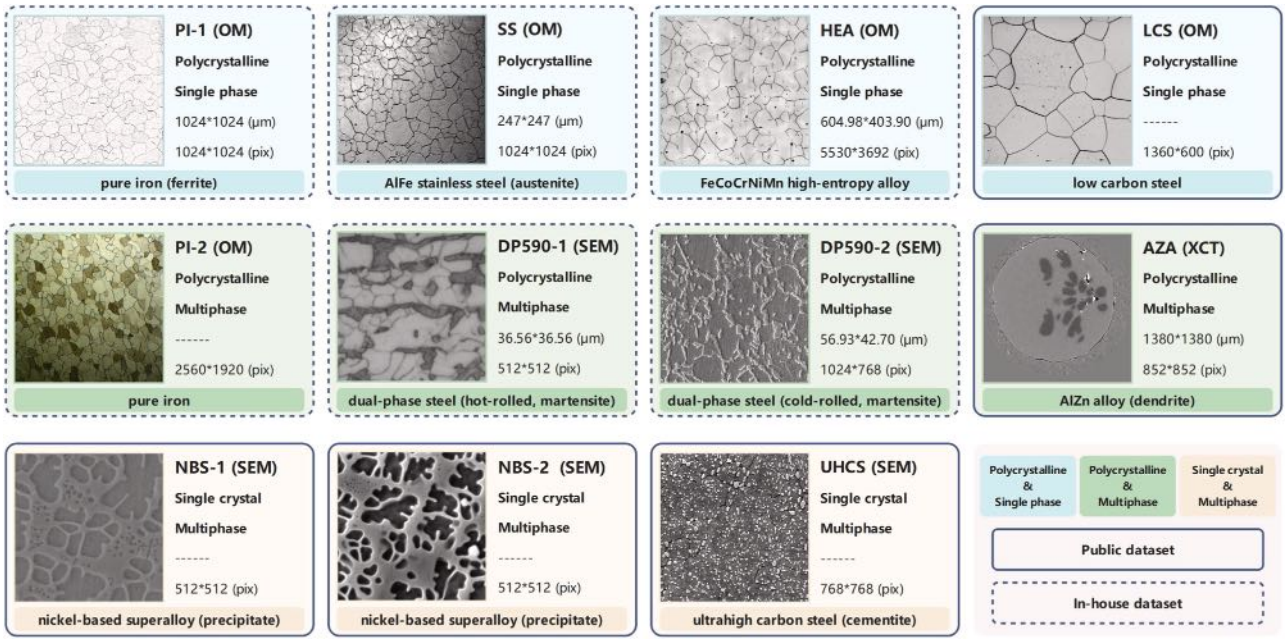


Fig. 1. Microscopy datasets used in this study. Different metallic materials are demonstrated. Based on structural features, the materials are classified into polycrystalline & single-phase, polycrystalline & multiphase, and single crystal & multiphase. The characterizing methods include OM, SEM, and XCT.

when different experts handle the same images, leading to a lack of general comparability. Furthermore, with the advancement of automated characterization equipment and experiment robots [12,13], the volume of micrographs requiring analysis has been growing exponentially [14,15]. Therefore, developing standardized, highly accurate, and robust models for extracting microstructural information from massive images is essential [16]. Such models are expected to significantly accelerate the extraction of microstructures, thereby enabling more efficient optimization of material processes and properties.

In recent years, conventional rule-based and deep learning-based (DL-based) image processing techniques have been increasingly applied in microstructure recognition and analysis [11,17]. Conventional rule-based algorithms, which are non-machine learning or deep learning (non-ML/DL), work well only on high-contrast micrographs with homogeneous compositions [18–20]. Oppositely, DL methods can recognize diverse and complex structures under inconsistent imaging protocols and conditions [6,21,22]. Nonetheless, annotating (or labeling) microstructure regions demands significant expertise and substantial manpower, making it difficult to acquire a large volume of high-quality annotations for model training. Meanwhile, most existing DL approaches struggle to achieve universal and robust microstructure recognition, i.e., weak generalizability across different microscopy datasets [16]. Those specialist models often perform optimally on the datasets they are trained on, restricting their applicability to new images with varying scales, resolutions, and complexities. To reduce the need for annotations and enhance generalizability, pre-training (acquiring robust initial feature representations) on large-scale datasets followed by fine-tuning (fitting the target data domain) has been widely utilized [23] and applied for micrographs [16]. However, these approaches still require annotations to fine-tune the model and hardly handle unseen images. In a nutshell, typical DL methods [6], trained on specific micrographs [24–26] or borrowing pre-training and fine-tuning paradigms [27], still exhibit limited transferability and generalization or inefficiency, making achieving optimal results across different materials, scales, and conditions difficult [28].

More recently, visual large models (VLMs) have been profoundly reshaping the development of image analysis [29]. Segment Anything Model (SAM) has emerged as the first VLM for recognizing any object in any image, as reported in [30]. It was trained on an unprecedented large-scale dataset (SA-1B). Notably, SAM's everything mode automatically generates grid points as a prompt. It identifies all objects without human intervention, which has the potential to greatly assist in efficiently recognizing microstructures of various materials. Unfortunately, despite SAM's strong feature representation ability, it fails to identify microstructures plausibly, mainly due to SAM's inherent bias toward natural images [31–33].

To respond to the above issues and focus on efficiently recognizing material microstructures, we propose MatSAM (Materials Segment Anything Model), a training-free approach built upon SAM, which flexibly and cost-effectively recognizes microstructures in polycrystalline or multiphase micrographs by injecting domain prior knowledge of structure and characterization into the model. In particular, besides the native grid point prompt, MatSAM employs a structure-aware point generation to adaptively provide the prompt to guide the VLM in yielding expected recognition results. Of note, the generation and adaptation of prompt points are automated; when adapting to a batch of new micrographs, users only need to provide the image category (polycrystalline or multiphase for grain boundary or phase extraction), instead of prompt points for each target object.

Further, we evaluated MatSAM's versatility and performance across 11 in-house and public microscopy datasets, including 5 polycrystalline and 6 multiphase datasets characterized by OM, SEM, or XCT. MatSAM, without training and annotations, demonstrated notable improvements over SAM, specialist DL models and rule-based methods, validating the effectiveness of our automated structure-aware prompt strategy. Consequently, these results highlight the potential of leveraging VLM with tailored optimization to reduce human labor in microstructural information extraction, accelerating the development of advanced materials.

2. Materials and methods

2.1. Microscopy datasets

We have collected and curated 11 materials microscopy datasets (see Fig. 1), five public (solid border), and six in-house (dashed border), to explore the recognition capability of MatSAM on different materials and microstructures (distinguished by three colors) under various imaging

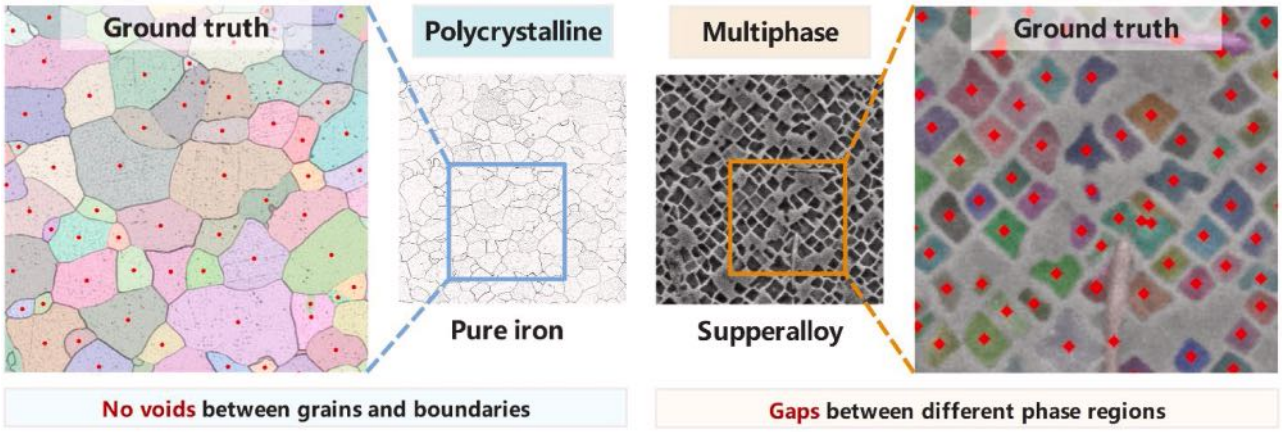


Fig. 2. Illustration of visual differences between polycrystalline and multiphase micrographs.

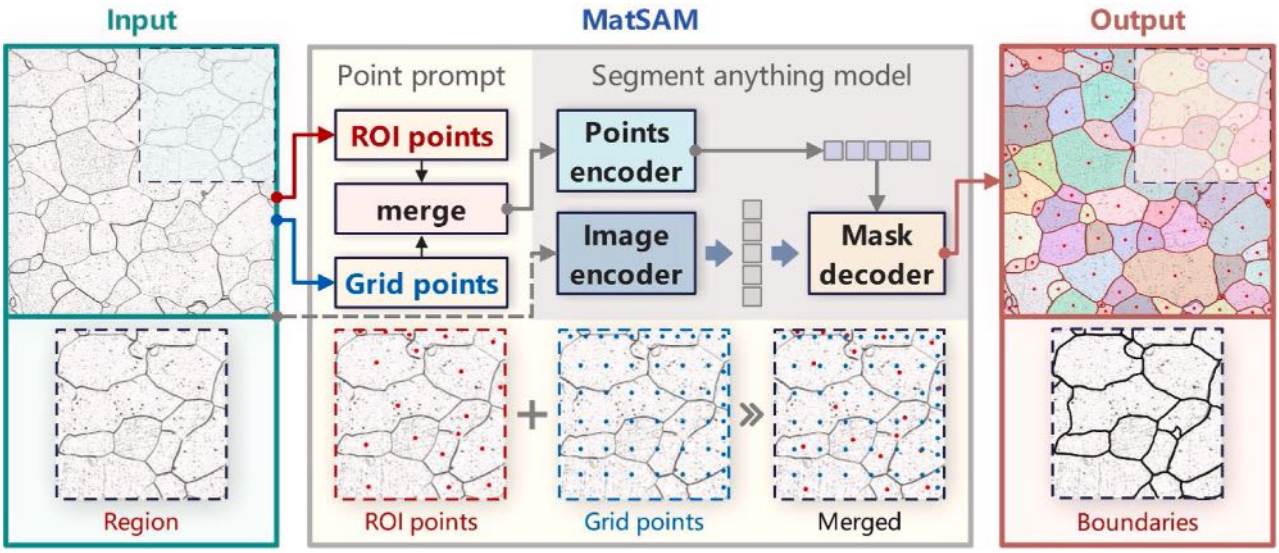


Fig. 3. Schematic of MatSAM for training-free microstructure recognition.

conditions. Each card includes an example image, the names of the material and dataset, the type of microstructure, and the real size and resolution of the micrograph. Polycrystalline & single phase, polycrystalline & multiphase, and single crystal & multiphase have blue, green, and light-yellow backgrounds, respectively. The micrographs in these 11 datasets were acquired using either OM, SEM, or XCT, and all in-house datasets have manual annotations. Specifically, images of pure irons (PI-1/2), stainless steel (SS), high-entropy alloy (HEA), and low carbon steel (LCS) [11] were captured using OM; images of dual-phase steels (DP590–1/2), nickel-based superalloys (NBS-1/2) [16], and ultrahigh carbon steel (UHCS) [11] were obtained using SEM; images of AlZn alloy (AZA) [25] were generated using XCT. To be noted, for public micrograph datasets—AZA [25], NBS-1/2 [16], and UHCS [11]—we used the attached test data. For another public dataset, LCS [11], where labels were not provided, part of the raw images was annotated for evaluation.

We elucidate the characteristics of two common material micrographs, i.e., polycrystalline images and multiphase images, exemplified in Fig. 2. The conspicuous difference between them stems from whether there are voids or gaps between multiple target objects. For a polycrystalline image, boundaries separate grains that adhere closely without voids between them, containing two classes - grain and its boundary with no background. For a multiphase image, gaps usually remain between phase regions that disperse across the whole image, consisting of multiple classes - different types of phases with background. Despite the above diversity, we deem they share a commonality: Regions of Interest (ROIs) can be effectively segmented as long as the prompt points abide by their features and distribution, which the proposed structure-aware prompt point generation of MatSAM can universally provide.

2.2. *Training-free MatSAM framework*

We implemented MatSAM based on SAM [30] which was pre-trained on the largest segmentation dataset to date. SAM used a primary architecture of an image encoder, a prompt encoder, and a segmentation mask decoder. Its emergent zero-shot capability and prompt engineering flexibility have created the conditions for training-free adaptation to various downstream tasks. Based on the pre-trained parameters of SAM and the orchestrated prompt strategy, MatSAM does not require extra training and human labeling; it processes a raw image input through the encoder, tokenizes a prompt, and combines the learned image feature with prompt embedding to output segmentation masks that extract material microstructures like grains, phases, precipitates, cracks, and so on. The schematic of MatSAM is shown in Fig. 3.

MatSAM utilizes SAM's cropping strategy to generate hierarchical crop boxes for multi-scale recognition. Each image is divided into multiple cropped regions of three different sizes: 1/1 (original size), 1/4,

Algorithm 1
Algorithm of structure-aware prompt point generation.

Input: I, c, threshold (preset to 3 pixels)
Output: $\text{Points}_{\text{merge}}$
1: **if** c = 'polycrystalline' **then**
2: $\text{Points}_{\text{ROI}}$ = Centroids (Canny (I))
3: **else if** c = 'multiphase' **then**
4: $\text{Points}_{\text{ROI}}$ = Centroids (OTSU (I))
5: **end if**
6: $\text{Points}_{\text{grid}}$ = Grid (I, $\text{Points}_{\text{ROI}}$)
7: $\text{Points}_{\text{edge}}$ = Edge (I, $\text{Points}_{\text{grid}}$)
8: $\text{Points}_{\text{grid}} \leftarrow \text{Points}_{\text{grid}} + \text{Points}_{\text{edge}}$
9: **for** each $\text{point}_{\text{g}} \in \text{Points}_{\text{grid}}$ **do**
10: **for** each $\text{point}_{\text{R}} \in \text{Points}_{\text{ROI}}$ **do**
11: **if** Distance (point_{g}, point_{R}) < threshold **then**
12: Remove (point_{g})
13: **end if**
14: **end for**
15: **end for**
16: $\text{Points}_{\text{merge}} = \text{Points}_{\text{ROI}} + \text{Points}_{\text{grid}}$
17: **return** $\text{Points}_{\text{merge}}$

and 1/16 of the image, with exact pixel sizes depending on the original resolution. To ensure that objects near the crop boundaries are adequately captured in neighboring crops, overlaps were applied across different crops, with a default overlap ratio of 0.5. The image encoder compresses the input image using a pre-trained Vision Transformer (ViT) [34], handling varying resolutions and outputting embeddings. The prompt encoder receives merged structure-aware prompt points and encodes the position information (mapping the prompt to feature embedding). Along with the prompt embedding, the image embedding is refined through self-attention (SA) and cross-attention (CA) in both prompt-to-image and image-to-prompt directions [35]. The mask decoder then upsamples the fused image embedding and, via a multilayer perceptron (MLP) and linear classifier, generates multiple candidate regions. To manage the resulting ambiguity, like SAM, candidate masks (up to three) are ranked by score, retaining those with the highest confidence, and Non-Maximum Suppression (NMS) is applied first within each cropped frame and then across all frames to produce the final, non-overlapping segmentation masks.

2.3. Automated structure-aware prompt generation

The native point prompt of SAM (everything mode) provided a fixed number of points (e.g., a 32 × 32 grid of equally spaced points over the entire image), which was enough to consider all ROIs in a natural image. Nevertheless, directly using the strategy in microscopy resulted in massive unacceptable false masks, i.e., local mis-segmentation of grain or phase regions and loss of a considerable quantity of objects at the edge of the image. Therefore, improving the prompt strategy was considered to achieve the practical training-free purpose (without manual annotation). Based on this motivation, we constructed a novel prompt strategy that enhanced point generation by tailoring it to microscopy with proper modification. The introduced structure-aware prompt point generation module provided centroid points for each potential target, thus automating the process of generating appropriate points. The workflow of the point generation (**Generation of ROI points** and **Modulation of grid points** below) is described in Algorithm 1, where I is the input microscopic image, c is the category ('polycrystalline' or 'multiphase'), the threshold (preset to 3 pixels) is used to determine the distance between two points are too close, and $\text{Points}_{\text{merge}}$ comprises adjusted ROI and grid points.

Generation of ROI points. Rule-based image segmentation techniques (Canny [18] or OTSU [19]) are employed to pre-segment the visually distinct target regions. For polycrystalline, the edge detection method (Canny) is used to extract the boundary of grains; for multiphase images, the threshold method (OTSU) is employed to extract phase regions. Then, the separated connected domain regions are obtained using the OpenCV library [36], of which the centroid points serve as part of the input to prompt encoder. It should be noted that outlier small regions are removed to eliminate the influence of impurities in the image.

Modulation of grid points. Providing the model with opportune points is the prerequisite to segmenting the micrograph correctly. Thus, the grid points are generated to prevent missing regions that may be overlooked during ROI extraction. Specifically, the potential inaccuracies and incompleteness of pre-segmentation, particularly in regions with subtle grayscale differences, necessitate the addition of grid points to compensate for the omission of the centroid points produced by the conventional method. To reduce redundancy, the added grid points too close to the ROI points (<3 pixels) are removed, as detailed in lines 9–15 of the algorithm.

In addition, a common issue is that the micrograph exhibits uneven quality caused by unstable light or imaging conditions (See SS and PI-2 images in Fig. 1). To address the problem and cover the ROIs near the edge of the image, MatSAM also generates more dense points in the edge regions (corresponding to $\text{Points}_{\text{edge}}$ = Edge(I, $\text{Points}_{\text{grid}}$) in the algorithm), enhancing the segmentation validity of incomplete objects at edges.

Aggregation of points. Fusing the ROI and grid points is straightforward to attain a set of more comprehensive locations. Here, we merge the $\text{Points}_{\text{ROI}}$ and $\text{Points}_{\text{grid}}$ set to form $\text{Points}_{\text{merge}}$, containing all position information. SAM's points encoder is reused to map the coordinate into latent embeddings, concretely,

$$P_{\text{fused}} = \varepsilon_{\text{point}}\left(\text{Points}_{\text{merge}}\right), \tag{1}$$

where P_{fused} is the points embedding of size $N_{\text{points}} \times 256$ (N_{points} is the number of the merged points) and $\varepsilon_{\text{point}}$ represents the points encoder. Hereto, the structure-aware prompt points are obtained through the designed automated strategy, considering most ROIs and imaging factors together.

Table 1
Evaluation protocols of 11 public and in-house microscopy datasets.

Type	Name	Material	Resolution (pixels)	#	Metrics
grain boundary	PI-1	pure iron	1024 × 1024	59	**ARI↑**
	SS	stainless steel	1024 × 1024	44	F1↑
	HEA	high-entropy alloy	1771 × 1061	38	Recall
	PI-2	pure iron	1280 × 960	12	
	LCS [11]	low carbon steel	1360 × 600	16	
phase	NBS-1 [16]	Ni-based superalloy	512 × 512	4	**IoU↑**
					F1↑
	NBS-2 [16]	Ni-based superalloy	512 × 512	5	Recall
	AZA [25]	AlZn alloy	852 × 852	2	
	UHCS [11]	ultrahigh carbon steel	645 × 475	5	
	DP590–1	dual-phase steel	512 × 512	23	
	DP590–2	dual-phase steel	1024 × 768	21	

An upward arrow indicates that a higher value of the metric corresponds to better segmentation performance.

2.4. Evaluation protocols

As for evaluation, conventional rule-based methods, specialist DL models, and the original SAM were selected to compare to MatSAM. For polycrystalline images (PI-1, PI-2, SS, HEA, LCS), grain boundary masks were extracted from the segmented grain regions using a contour-finding approach, enabling the calculation of discrepancies between different methods and human labels. For multiphase images (DP590–1, DP590–2, NBS-1, NBS-2, AZA, UHCS), phase masks are directly segmented for comparisons. Importantly, MatSAM's recognition results are all under the training-free condition.

The corresponding evaluation settings of 11 datasets are given in

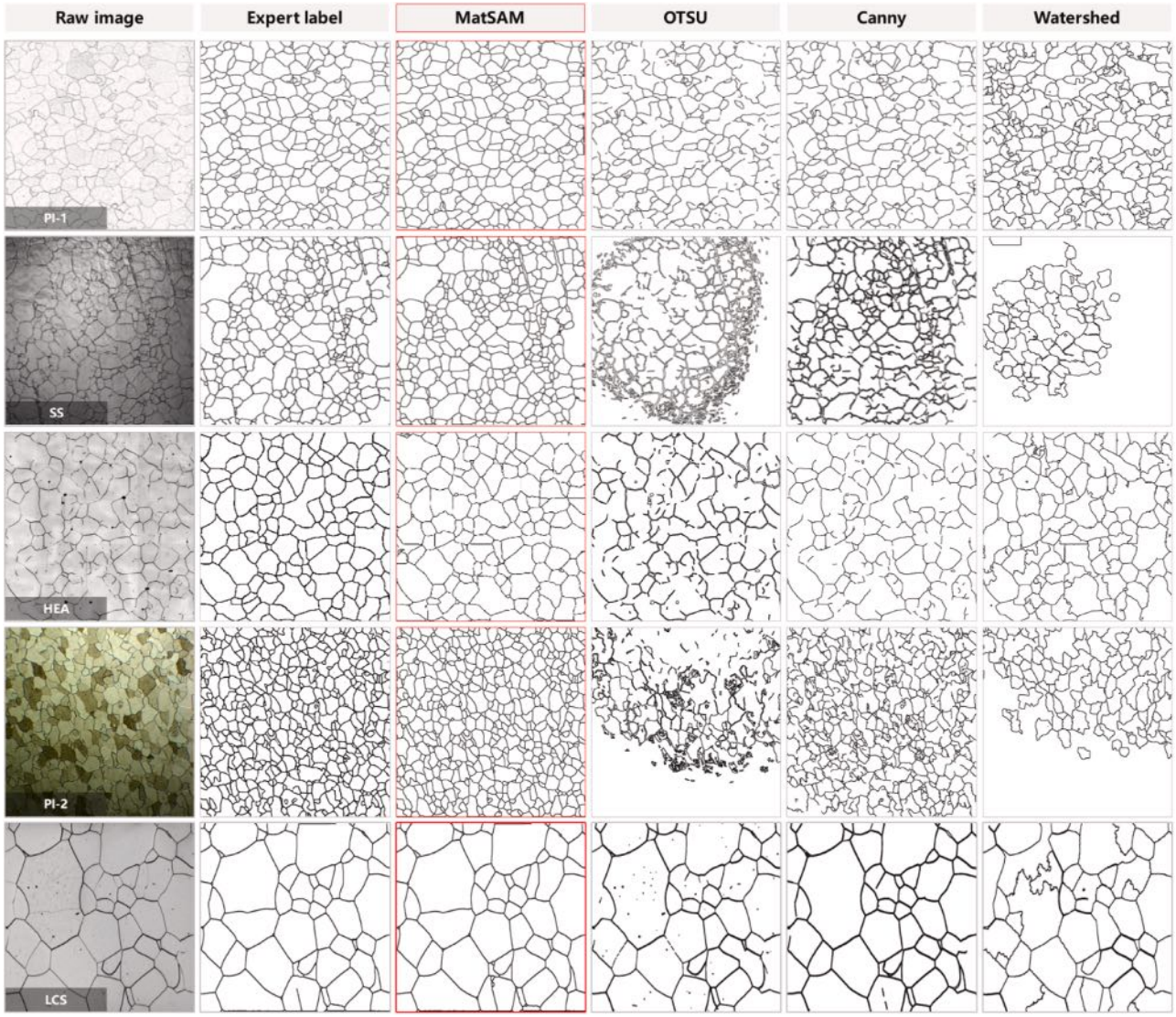


Fig. 4. Visual results of MatSAM and rule-based methods on polycrystalline microscopy.

Table 1, including input resolution, the main type of microstructure, the number of test images, and evaluation metrics. Four metrics are chosen to demonstrate the performance of recognition: adjusted Rand index (ARI, exclusive for polycrystalline images), Intersection over Union (IoU, exclusive for multiphase images), F1 score, and recall. Higher values of ARI, IoU, and F1 score indicate better recognition performance. Note that all numerical results are given in "mean±SD" form.

ARI [37], the corrected-for-chance version of the Rand index (RI), measures the degree to which instances are correctly clustered in a given set, accounting for random chance [38]. The RI, taking values in the range [0,1], is calculated as (TP+TN) / (TP+FP+FN+TN), where TP, TN, FP, and FN are the number of true positives, true negatives, false positives, and false negatives. and the ARI penalizes incorrect classifications more heavily than the RI does, namely,

$$\mathrm{ARI} = (\mathrm{RI} - E[\mathrm{RI}])/(\max(\mathrm{RI}) + E[\mathrm{RI}]), \tag{2}$$

where E[RI] represents the expected value of the RI, the ARI assigns different weights to correct and incorrect classifications, with a range of [−1,1]. Image segmentation can be viewed as clustering pixels into segments. In grain boundary segmentation, a class imbalance arises with fewer pixels in boundary regions compared to grains, while boundaries exhibit complex, irregular, and multi-branch structures. The cluster-based ARI metric evaluates segmentation accuracy by comparing clustering results with ground truth, accounting for random clustering effects to ensure unbiased evaluation. This makes ARI well-suited for handling class imbalance and capturing complex topological structures in grain boundary segmentation [15,39]. A higher value indicates greater similarity between the segmentation results and the ground truth.

IoU [40] or Jaccard index evaluation metric is a commonly used metric for assessing the performance of object detection and image segmentation models, specifically,

$$\mathrm{IoU} = \mathrm{TP}/(\mathrm{TP} + \mathrm{FP} + \mathrm{FN}), \tag{3}$$

which measures the overlap between the predicted region of the model and the ground truth and is used to assess the model's localization accuracy for the target.

We conducted extensive quantitative evaluations by comparing the metrics of different methods (ARI, IoU, F1, and Recall), as well as qualitative evaluations through visual comparisons, including overlaying segmentation masks and error visualizations. Conventional methods–Canny [18], OTSU [19], and Watershed [41] are implemented for polycrystalline images, and OTSU and adaptive thresholding (Adaptive) [20] for multiphase images. Additionally, for public datasets,

Table 2
Performance comparison of MatSAM and other methods on polycrystalline datasets.

Dataset	Metric	Method				
		MatSAM	SAM	Canny	OTSU	Watershed
PI-1	**ARI**	**0.62 (±2.4 %)**	0.48 (±0.8 %)	0.52 (±4.1 %)	0.50 (±3. 8 %)	0.44 (±4. 7 %)
	F1	0.71 (±1.6 %)	0.37 (±0. 5 %)	0.64 (±3.0 %)	0.67 (±2.9 %)	0.56 (±3.2 %)
	Recall	0.71 (±1.5 %)	0.86 (±1. 4 %)	0.86 (±3. 8 %)	0.79 (±3. 8 %)	0.63 (±9.2 %)
SS	**ARI**	**0.56 (±6.6 %)**	0.54 (±1.1 %)	0.34 (±7.2 %)	0.23 (±5. 8 %)	0.24 (±4.3 %)
	F1	0.65 (±5.0 %)	0.66 (±1.7 %)	0.51 (±7.9 %)	0.46 (±5.7 %)	0.41 (±5.6 %)
	Recall	0.59 (±6.0 %)	0.60 (±5. 5 %)	0.41 (±8. 2 %)	0.48 (±7. 9 %)	0.33 (±6.6 %)
HEA	**ARI**	**0.59 (±10.8 %)**	0.56 (±6.0 %)	0.35 (±14. 3 %)	0.37 (±13.3 %)	0.29 (±10. 8 %)
	F1	0.35 (±11.1 %)	0.52 (±11. 5 %)	0.42 (±13.3 %)	0.41 (±11.9 %)	0.39 (±10.4 %)
	Recall	0.24 (±7.8 %)	0.43 (±9. 3 %)	0.33 (±10.4 %)	0.34 (±9.4 %)	0.33 (±8. 7 %)
PI-2	**ARI**	**0.75 (±2.0 %)**	0.69 (±3.3 %)	0.42 (±8. 2 %)	0.19 (±4. 0 %)	0.20 (±4. 6 %)
	F1	0.82 (±1.7 %)	0.80 (±1. 6 %)	0.51 (±7. 9 %)	0.29 (±4.0 %)	0.28 (±5. 2 %)
	Recall	0.83 (±1.8 %)	0.81 (±1.3 %)	0.45 (±9. 0 %)	0.26 (±5. 8 %)	0.24 (±6.5 %)
LCS [11]	**ARI**	**0.96 (±3.7 %)**	0.86 (±9.8 %)	0.73 (±5. 3 %)	0.64 (±4. 3 %)	0.54 (±4. 6 %)
	F1	0.96 (±3.2 %)	0.88 (±0.7 %)	0.77 (±4. 8 %)	0.69 (±4. 3 %)	0.61 (±4. 8 %)
	Recall	0.97 (±2.5 %)	0.88 (±1.8 %)	0.71 (±6.4 %)	0.72 (±7. 7 %)	0.69 (±4. 5 %)

For the primary metric (ARI), bold and underlined values indicate the **best** and second-best results, respectively.

the recognition results of MatSAM are directly compared with those of corresponding specialist DL models. All the validation experiments are conducted on a single Nvidia RTX3090 24 GB GPU using the PyTorch 1.13.1 library [42].

3. Results

3.1. Training-free microstructure recognition

3.1.1. For polycrystalline microscopy datasets

Polycrystalline images typically comprise tightly packed grains, whose boundaries are the target objects of the analysis [43]. As shown in Fig. 4, in some micrographs imaged via OMs, grayscale values vary little between adjacent grains and largely across regions (SS, HEA, and PI-2, see the first column). MatSAM, after receiving structure-aware points, the reasonable prior knowledge of the target regions to the model delivers more accurate and refined recognition results among five datasets (see the third column). On the other hand, threshold methods (OTSU) often suffer from grayscale variations (SS and PI-2, the fourth column). Canny is less affected than threshold methods and can effectively detect most grain boundaries with instinct grayscale differences. However, it is significantly affected by blurred or incomplete grain boundaries (PI-1, SS, and HEA, the fifth column).

Table 2 presents the numerical results of the comparison of MatSAM and other methods, including the original SAM and other rule-based methods. Overall, MatSAM achieves the best ARI results across all five datasets, consistently outperforming SAM, which validates the effectiveness of the automated structure-aware prompt strategy. Additionally, MatSAM significantly outperforms other conventional methods, as expected.

For PI-1, MatSAM achieves an ARI of 0.62 on 59 test images, outperforming SAM (0.48) and Canny (the second best, 0.52). The SS dataset, collected under relatively non-ideal imaging conditions with more blurry regions, scratches, and impurities, poses a significant challenge. MatSAM achieves an ARI of 0.56 on 44 test images, outperforming SAM (0.54) and Canny (0.34). For HEA micrographs, due to the limited field of view, each image is stitched from multiple smaller ones, resulting in visible horizontal or vertical stitching artifacts caused by brightness variations. Additionally, extensive non-closed grain boundaries and protruding cracks make segmentation particularly challenging. MatSAM achieves an ARI of 0.59, higher than SAM (0.56), and the best rule-based method, OTSU (0.37).

On the PI-2 dataset, MatSAM achieves an ARI of 0.75, outperforming SAM (0.69) and Canny (0.42). Uneven lighting, as shown in the fourth row of Fig. 4, leads to fatal false recognition by OTSU and Watershed. For the public LCS dataset, which benefits from good imaging conditions, MatSAM demonstrates the greatest accuracy, achieving an ARI of 0.96 on 16 test images, compared to SAM (0.86) and other recognition techniques by a large margin. From the visual recognition results, MatSAM not only segments almost all grains in the LCS (see the fifth row of Fig. 4) but also accurately identifies the vast majority of incomplete and blurred grain boundaries. Furthermore, despite the large variation in image resolution within this dataset, MatSAM consistently outputs satisfactory results, demonstrating its strong robustness. This robustness is believed to stem from the combination of structure-aware prior knowledge (provided by rule-based methods) and universal feature representations from SAM.

3.1.2. For multiphase microscopy datasets

Micrographs of multiphase microstructures contain one or multiple types of phases (the secondary phase, the precipitate, etc.). Unlike grain structures, the phase regions to be segmented do not typically fill the entire image but are irregularly distributed. Also, different phases have conspicuous differences in morphology, structure, and quantity, which can be used to distinguish them after getting the segmented masks.

Subjective differences between MatSAM and other methods are shown in Fig. 5. Since the threshold-based Adaptive method performed poorly and lacked reference value, it is not included in the figure. To provide a more comprehensive comparison, we have also included the edge detection results from the Canny and Watershed methods. Overall, the recognition results of MatSAM are the most consistent with the expert labels, significantly outperforming other rule-based methods. MatSAM is capable of effectively identifying phase regions of various sizes in images. For the NBS-2, DP590–2, and UHCS micrographs, MatSAM can most comprehensively identify the small regions in the images (various precipitate phases), whereas the OTSU method suffers from under-segmentation due to the minimal grayscale difference between the phase regions and the background.

When it comes to AZA, the round circle boundary is unexpectedly detected by methods such as OTSU, Canny, and Watershed. This discrepancy can be attributed to the nature of these algorithms. Canny and Watershed, as edge detection methods, are highly sensitive to abrupt intensity changes, making them prone to identifying the circle's strongly contrasting boundary as a meaningful edge. Similarly, OTSU, due to the circle's low gray value, likely includes the boundary as part of

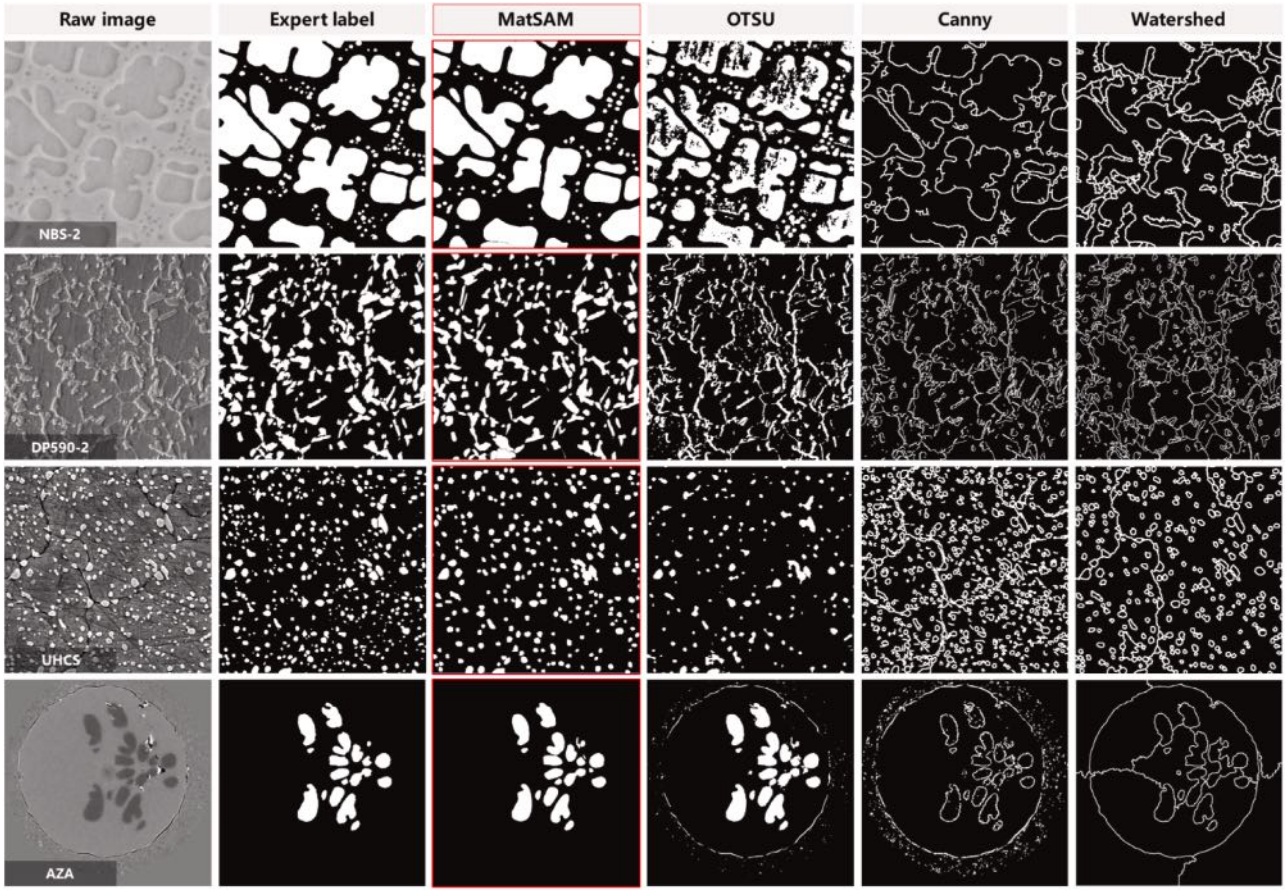


Fig. 5. Visual results of MatSAM and rule-based methods on multiphase microscopy.

the foreground during thresholding. Without post-processing to remove noise, these methods tend to over-segment the image. In contrast, MatSAM focuses on identifying meaningful regions rather than abrupt edges, leading to the suppression of this boundary.

Table 3 shows the numerical segmentation results on multiphase microscopy datasets, comparing MatSAM and other methods. On the two dual-phase steel datasets (DP590–1 and DP590–2), martensite precipitates formed during cold and hot rolling are the targets of interest. For the DP590–1 dataset, it consists of 23 hot-rolled martensitic precipitate images with relatively low definition. The precipitates exhibit low gray values and are contoured by black boundaries, with most appearing more blurred compared to the cold-rolled DP590–2 dataset (see Fig. 1), making them more difficult to distinguish. Among conventional methods, OTSU achieves an IoU of 0.64, while SAM achieves an IoU of 0.75. However, both results are significantly lower than MatSAM's IoU of 0.77, which is 13.0 % higher than OTSU and outperforms SAM by 2.0 %. For the DP590–2 dataset, martensite precipitates appear as light white streaks in the images. The IoU between MatSAM-segmented masks and annotations is 0.82, which is observably higher than the SAM method (0.68) and the OTSU threshold method (0.66).

3.1.3. *For public microscopy datasets*

To validate the zero-shot generalizability of MatSAM across broader scenarios, we performed comparative experiments with conventional approaches and specialist models on 4 materials microscopy test datasets released by previous works, including NBS-1 and NBS-2 (corresponding to Super-2 and Super-4 in the original article) [16], AZA [25], and UHCS [11]. The quantitative results are given in Table 3. LCS [11] is analyzed qualitatively in Fig. 6 due to the lack of released labels. As shown in Table 3, compared to specialist models, MatSAM acquired the best results on three datasets—NBS-1, NBS-2, and UHCS—and ranked second on AZA, slightly lower than the specialist method.

NBS-1 and NBS-2 contain multiple precipitates of significantly different sizes. The tertiary precipitates are particularly small and blurry, making recognition challenging. MatSAM achieves an IoU of 0.91 on NBS-1, outperforming SAM (0.85) and the specialist model (0.77); it also scores 0.82 on NBS-2, 3.0 % higher than the specialist model. Regular matrix and secondary precipitates are well recognized with few phase contour distortions. For smaller and blurred tertiary precipitates with low contrast, MatSAM also exhibits better performance, as shown in the first column of Fig. 6 (yellow arrows highlight regions where MatSAM correctly segments areas that the original method under-segments). When tested on unseen images without fine-tuning, the recognition performance of the specialist model dropped significantly. In contrast, when handling the same diverse unseen images, MatSAM can surpass their results without additional training, underscoring its superior efficiency and adaptability.

AZA comprises two XCT (X-ray computed tomography) images of AlZn dendrite, with the dark precipitate phase in the foreground, located in the center of the circular region to be segmented. For this simple task, dendrites are accurately extracted (see the second column of Fig. 6), achieving an IoU of 0.96, which approaches 0.97 of the corresponding specialist model trained on over 1000 annotated image samples. Besides, for the UHCS dataset, which contains numerous light-colored precipitates (spheroidite particles) uniformly distributed on the top surface of a rough alloy crack, MatSAM attains 0.76, outperforming the specialist model by 5.0 %, with the vast majority of cementite particles correctly identified and segmented. Due to the complex background and large grayscale differences, SAM, OTSU, and Adaptive methods perform poorly on this dataset, with IoU scores of 0.58, 0.56, and 0.46, respectively. Additionally, for polycrystalline images of LCS (the fourth

Table 3
Performance comparison of MatSAM and other methods on in-house and public and multiphase datasets.

Dataset	Metric	Method				
		MatSAM	SAM	OTSU	Adaptive	Specialist method
DP590–1	**IoU**	**0.77 (±8.2 %)**	0.75 (±13.4 %)	0.64 (±9.7 %)	0.48 (±5. 3 %	/
	F1	0.79 (±8.3 %)	0.85 (±9. 2 %)	0.68 (±11.3 %)	0.49 (±8.9 %)	/
	Recall	0.81 (±14.7 %)	0.66 (±26.0 %)	0.91 (±8.6 %)	0.65 (±10.0 %)	/
DP590–2	**IoU**	**0.82 (±3.1 %)**	0.68 (±6.7 %)	0.66 (±1. 3 %)	0.28 (±0. 9 %)	/
	F1	0.84 (±3.2 %)	0.81 (±4. 8 %)	0.66 (±2.0 %)	0.17 (±2.1 %)	/
	Recall	0.80 (±4.9 %)	0.81 (±2. 1 %)	0.58 (±2.7 %)	0.22 (±1.9 %)	/
NBS-1 [16]	**IoU**	**0.91 (±1.2 %)**	0.85 (±7.6 %)	0.75 (±3. 2 %)	0.45 (±1.6 %)	0.77
	F1	0.96 (±0.4 %)	0.91 (±4. 6 %)	0.86 (±3. 6 %)	0.52 (±1.9 %)	–
	Recall	0.94 (±0.4 %)	0.89 (±6.5 %)	0.87 (±8.4 %)	0.38 (±1. 9 %)	–
NBS-2 [16]	**IoU**	**0.82 (±6.3 %)**	0.81 (±6.1 %)	0.67 (±14.1 %)	0.52 (±9.2 %)	0.79
	F1	0.89 (±4.9 %)	0.90 (±3. 8 %)	0.80 (±9. 4 %)	0.62 (±9. 8 %)	–
	Recall	0.84 (±8.8 %)	0.84 (±9.0 %)	0.79 (±7.8 %)	0.53 (±10.8 %)	–
AZA [25]	**IoU**	0.96 (±0.6 %)	0.73 (±6. 8 %)	0.89 (±1.2 %)	0.61 (±2. 3 %)	**0.97**
	F1	0.96 (±0.5 %)	0.84 (±4. 5 %)	0.88 (±1. 7 %)	0.48 (±5.8 %)	–
	Recall	0.96 (±0.9 %)	0.96 (±0. 5 %)	0.97 (±0. 1 %)	0.76 (±3. 2 %)	–
UHCS [11]	**IoU**	**0.76 (±0.7 %)**	0.58 (±8. 2 %)	0.56 (±3. 1 %)	0.46 (±3. 5 %)	0.71
	F1	0.75 (±0.9 %)	0.73 (±6. 6 %)	0.72 (±2.5 %)	0.43 (±5. 6 %)	–
	Recall	0.78 (±2.7 %)	0.77 (±3.3 %)	0.59 (±3. 2 %)	0.93 (±1.3 %)	0.88

For the primary metric (IoU), bold and underlined values indicate the **best** and second-best results, respectively.

column of Fig. 6), MatSAM demonstrates a comparable recognition performance to the original method without human intervention.

3.2. Microstructure recognition in complex scenarios

While validating MatSAM's capabilities, we observed that even in challenging scenarios involving complex, variable, and difficult-to-interpret micrographs, MatSAM consistently provided reasonable outputs. This effectiveness is illustrated through examples from polycrystalline and multiphase images (See Fig. 7).

The common issue for grains in the images is that incompleteness or non-closing occurs on the boundary, as marked on the left in Fig. 7, often due to uneven or incomplete etching during the preparation process. Most untailored approaches underperform in the situation, causing under-segmentation and further affecting statistical fidelity. By contrast, MatSAM mitigates this problem, as exemplified in (b); the grain boundaries in the region indicated by the arrow are nearly invisible, yet MatSAM can still infer the missing parts. We believe this is due to SAM's extensive pre-training on a vast number of images, which has enabled it to effectively capture the characteristics and patterns of such imaging. Combined with the optimized point prompt in MatSAM, it can detect that a corresponding boundary should exist in this region based on the pixel distribution surrounding the area. Also, in (d), some boundaries are hard to tell at the grain interfaces. Nevertheless, MatSAM captures and reasonably recognizes the subtle changes near the boundaries, demonstrating the flexibility of MatSAM when handling this pervasive and troublesome challenge.

For multiphase images from (e) to (h), with phase regions containing tiny holes (e), target phase boundaries intermingled with surrounding areas (f), precipitate sizes differing amongst large, medium, and small (g), and small-sized particles scattered throughout the image (h), MatSAM ideally identifies the difficult areas. The designed structure-aware prompt strategy enables MatSAM to recognize target regions of varying sizes without labeling accurately. This capability is highly beneficial for automated phase detection and quantitative analysis.

4. Discussion

4.1. Effectiveness of MatSAM

SAM [30] excels at extracting general features due to extensive pre-training but exhibits suboptimal performance on micrographs, where grayscale variations dominate. Besides, SAM's foreground bias, which stems from pre-training on sparse distinct targets within natural scene images [31], does not align with the imaging features of microscopy. While SAM performs well with human-provided prompts, its efficiency diminishes for large-scale tasks. MatSAM addresses these challenges and transfers the general capability of SAM to microscopy scenarios by automating prompt point generation. Using rule-based methods and optimizing grid points to guide the mask decoder effectively, our approach thereby better handles various micrographs better.

To evaluate the effectiveness of the automated structure-aware prompt strategy, we conducted comparative validation experiments against the original SAM method on representative SEM and XCT images from DP590–2 and AZA datasets, respectively. Three common types of noise frequently encountered in microscopic images—salt-and-pepper (S&P) noise, speckle noise, and Gaussian noise—were selected [44,45] and introduced into the images using the random noise generation function from the sci-kit image [46] library. The robustness of MatSAM compared to SAM was further investigated under varying levels of these noise types. Fig. 8 presents the comparative results of two microscopic images with varying levels of S&P noise processed by each method (details for the other two types of noise are provided in Supplementary Note 1). The noise levels were adjusted by the variance (Var. = 0.01, 0.05, 0.1, 0.2).

Overall, MatSAM demonstrates better recognition accuracy (both subjective and objective) on images with varying levels of noise compared to SAM, identifying more target regions. Additionally, the introduction of noise exacerbates the uncertainty of both methods in recognizing prominent regions within the images. The boundaries of the target regions become increasingly blurred and more easily confused with the background, which has a more detrimental impact on microscopic images—characterized by inherently lower contrast compared to natural scene images. MatSAM's IoU on DP590–2 images decreases from 0.822 to 0.549 (SAM drops from 0.768 to 0.485), while on AZA images, the IoU declines from 0.946 to 0.815 (SAM drops from 0.903 to 0.776). As noise intensity increases, under-segmentation gradually becomes the

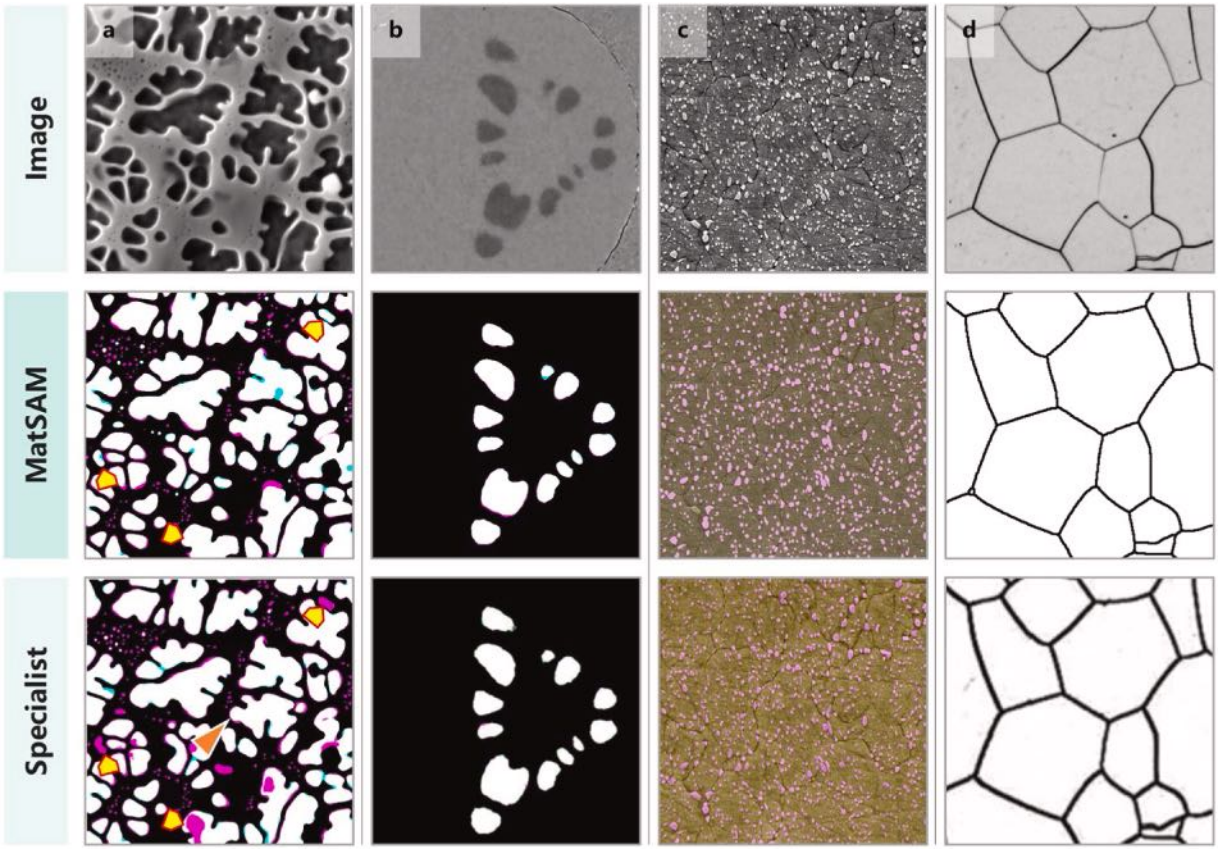


Fig. 6. Visual results of MatSAM and other specialist models. The first row shows open-sourced images: **(a) NBS-2** [16], **(b) AZA** [25], **(c) UHCS** [11], **and (d) LCS** [11]. The second row displays the results of MatSAM, while the third row presents the results from data-specific models (the mask images are cropped from the raw figures in their respective articles). We adopt the color styles from each article: for the results of **(a)** and **(b)**, pink and blue represent under- and over-segmentation, respectively; for **(c)**, pink areas are the masks.

dominant type of error, leading to a significant decline in accuracy for both methods. These results highlight the challenge of noise in microscopic image segmentation and demonstrate MatSAM's relative robustness in maintaining higher accuracy compared to SAM.

4.2. *Efficiency of MatSAM*

Given MatSAM's effective segmentation capability without the need for training, we believe it, as an out-of-the-box tool, has the potential to significantly accelerate the process in real-world quantitative characterization of material microstructures. We qualitatively compared several common modes conducted by researchers for microstructure extraction tasks, elaborated in Fig. 9. These include "fully manual annotation (Mode 1)," "rule-based methods with manual correction (Mode 2)," "learning-based methods with manual correction (Mode 3)," and "visual large models with manual correction (Mode 4)."

Mode 1: this mode has the longest time cost, and due to the subjective influence of annotators, the final annotation accuracy may be affected detrimentally. **Mode 2**: while the runtime for general rule-based methods is minimal, significant time is often required to adjust the algorithm's hyperparameters for new images to accommodate their visual characteristics. The complexity of the images also necessitates manual intervention for corrections, which can be time-consuming. **Mode 3**: researchers typically annotate a batch of data, followed by training with DL methods. The trained model is then applied to the full dataset to generate predictions, which are followed by manual adjustments to finalize the outputs. Compared to the first two modes, this approach significantly reduces time costs while improving accuracy, both during model development and the manual correction phase. This is currently the most widely adopted approach, though it still demands substantial time and resources. **Mode 4**: by using an improved version of a general-purpose visual model, such as MatSAM, researchers can perform full inference on all images without adjusting method parameters. While MatSAM's inference time per image is longer than that of classical deep learning methods and traditional methods, it requires no additional effort for model training or annotation. The segmentation results are almost immediately usable without extensive modification, allowing researchers to obtain high-quality recognition results in a significantly shorter time with minimal effort.

For normal micrographs, segmentation masks generated by MatSAM can meet the accuracy requirement for subsequent descriptive statistics. MatSAM can replace conventional or DL-based methods as a segmenter in challenging scenarios at inconsistent imaging conditions and various structures to give a more reliable initial result, e.g., continuous three-dimensional (3D)-slice images acquired by serial sectioning techniques [47–49]. Then, the false-segmented regions are corrected for model training through interactive correction involving human intervention. Empirically, it usually takes about 45 min to pre-segment and manually correct a microscopic image with approximately 300 grains using conventional methods. This process is highly susceptible to the operator's subjective biases. In contrast, using MatSAM can reduce the delineation plus correction time to within 5 min. The essential reason is that the initial results from MatSAM are already precise enough to identify and segment most grains. The operator only needs to correct the over- or under-segmentation regions caused by distinct defects or image blurring, less susceptible to visual fatigue and distraction.

In general, MatSAM's efficiency can be attributed to the following two factors: First, because SAM has already extensively learned

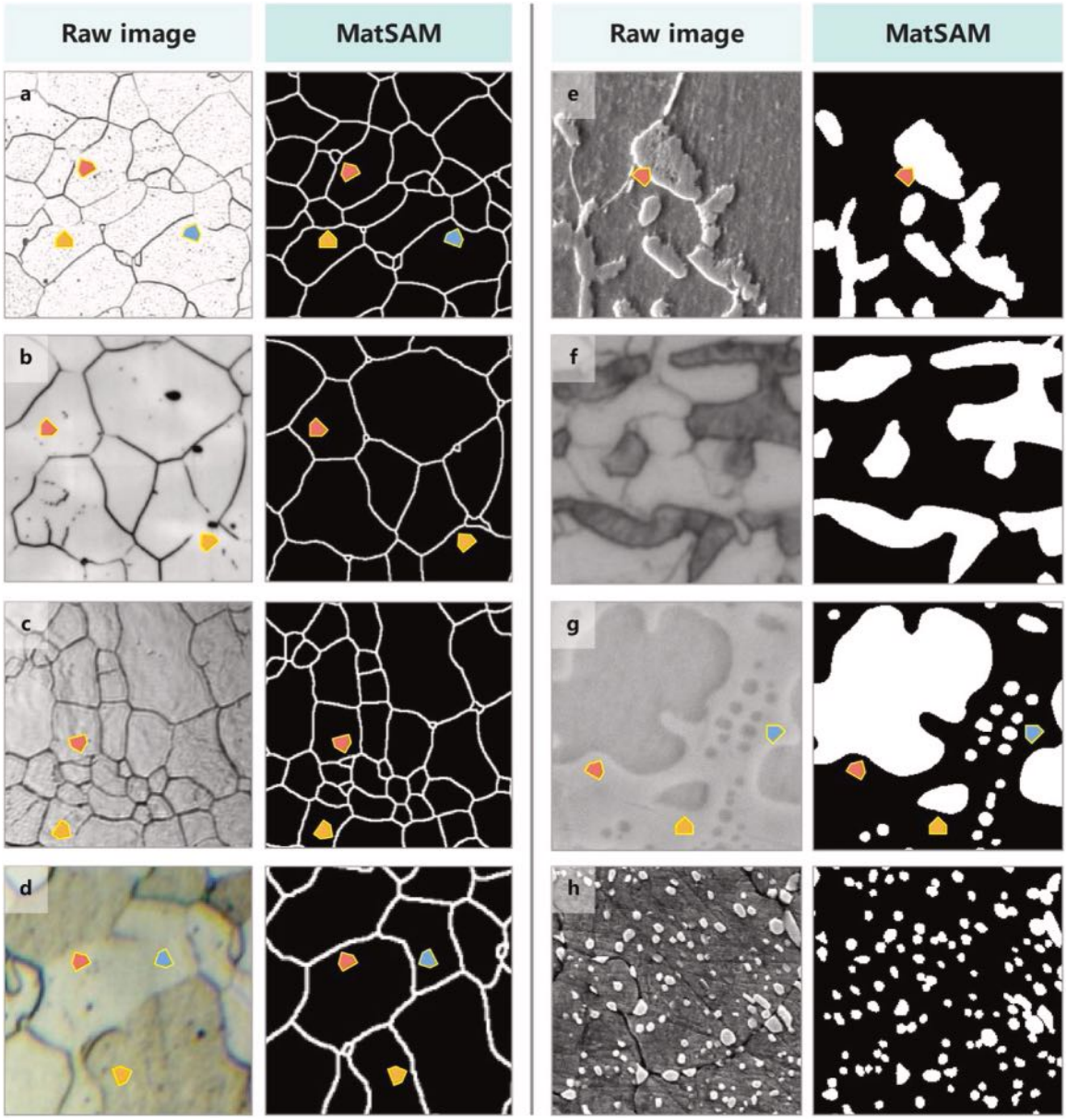


Fig. 7. Illustration of MatSAM segmenting microstructures under different morphologies. The left is polycrystalline images, and the right is multiphase images. **(a) PI-1, (b) HEA, (c) SS, (d) PI-2, (e) DP590–2, (f) DP590–1, (g) NBS-1, and (h) UHCS**. Arrows point to noteworthy areas (a pair of arrows shares the same color).

universal image features, there is no need for additional feature learning or selection when applied to material micrographs. Second, MatSAM leverages conventional methods to direct the model's attention to ROIs. Since this step only requires a rough estimation of potential regions and does not demand high precision, there is no need to modify the model for specific image types. Therefore, MatSAM can serve as a powerful and fast tool for microscopy analysis, accelerating quantitative characterization.

4.3. *Extensibility of MatSAM*

MatSAM's extensibility or generalizability enables accurate and robust recognition of various microstructures at different scales and morphologies within a single micrograph. To elucidate that, we conducted extended experiments on six additional micrographs, extracting corresponding microstructures, i.e., NBS-3 (γ' phase), NBS-3 (TCP phase), IN939–1 (melted pool boundary), IN939–2 (cracks), C22000 and C21000 brasses (twin grain) [50]. These image samples are not provided with corresponding annotations. Fig. 10, which demonstrates MatSAM's capability to identify different phases and key microstructures. See Supplementary Note 2 for results on a wider variety of images.

For micrographs of NBS-3, as shown in (a) and (b) of the figure, γ' (small light grey areas) and TCP phases (fine needle-like) are the key objects of the analysis. Additionally, impurities, which are deeply dark in color and distinctly larger in area, are also present. MatSAM accurately recognizes all phases, as illustrated in the Overlay columns. For micrographs of superalloy IN939–1 and IN939–2, shown in (c) and (d) of the figure, the melted pool boundary and crack are the main focus. The boundaries of melted pools and crack regions are reasonably recognized and clearly separated. For micrographs of brasses, commercial bronze (C22000) and gilding metal (C21000) shown in the (e) and (f) of the image, the twin grains—characterized by straight or planar boundaries and contrasting regions—are the primary components of interest.

MatSAM, prompted by structure-aware points, segments nearly all the mentioned phases. Following the recognition process, each type of microstructure can finally be extracted by filtering all generated masks

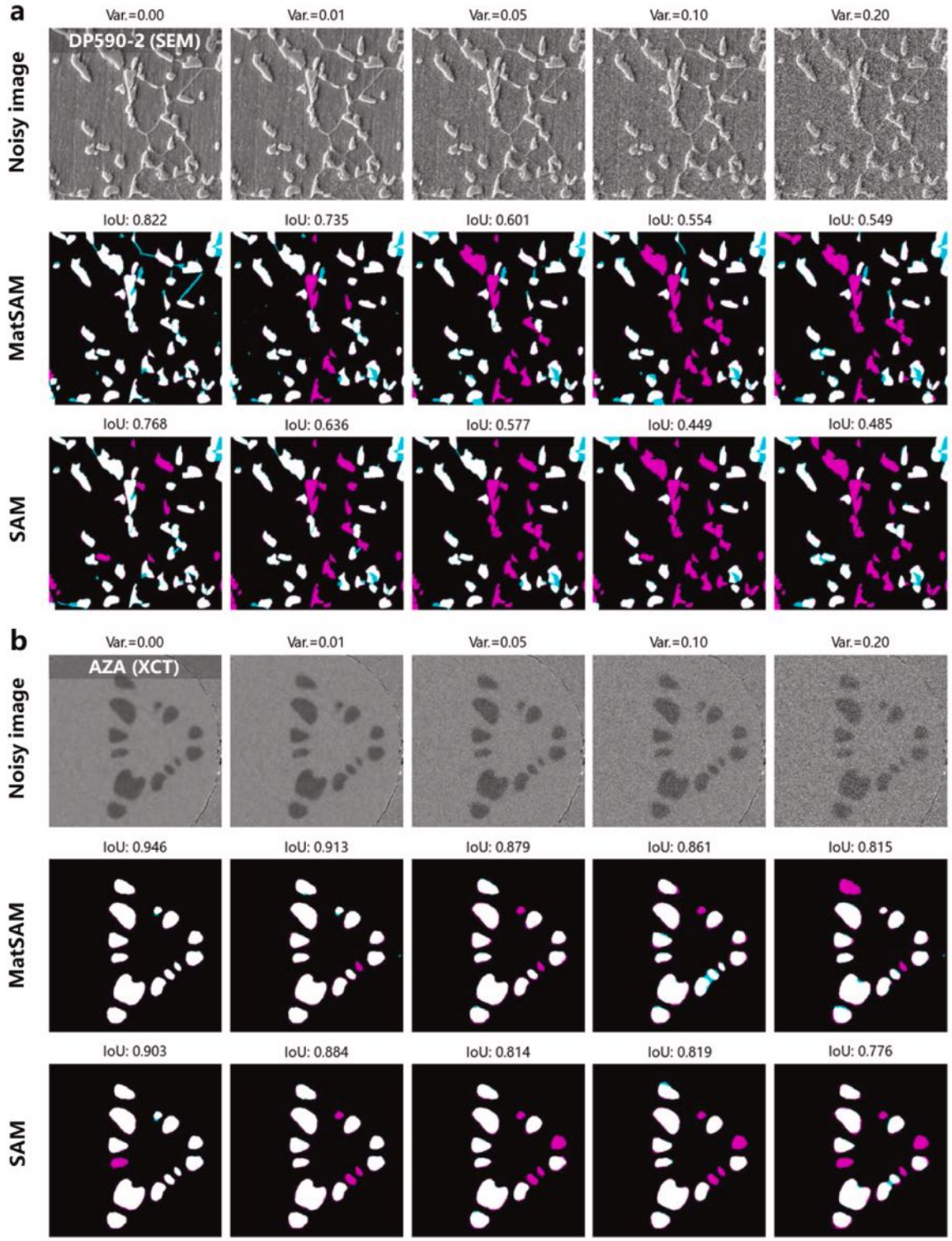


Fig. 8. Visual results of MatSAM and SAM on micrographs with different levels of S&P noise. Pink and blue regions represent under- and over-segmentation, respectively. **(a) Results visualization on an SEM image from the dual-phase steel dataset (DP590–2). (b) Results visualization on an XCT image from AlZn alloy dataset (AZA).**

based on their morphological features, such as area, color distribution, orientation, and shape characteristics. As shown in the Key structure columns of the figure, the binary mask results of the corresponding ROIs are precisely separated. Consequently, researchers can launch autonomous parameter calculations and further statistical analyses without concerning themselves with the annotation and training processes.

4.4. Limitations of MatSAM

Despite the significant advancements achieved by MatSAM in material microstructure extraction tasks, several limitations persist that merit further exploration. First, MatSAM exhibits limitations in handling highly noised micrographs with smaller scales, particularly in

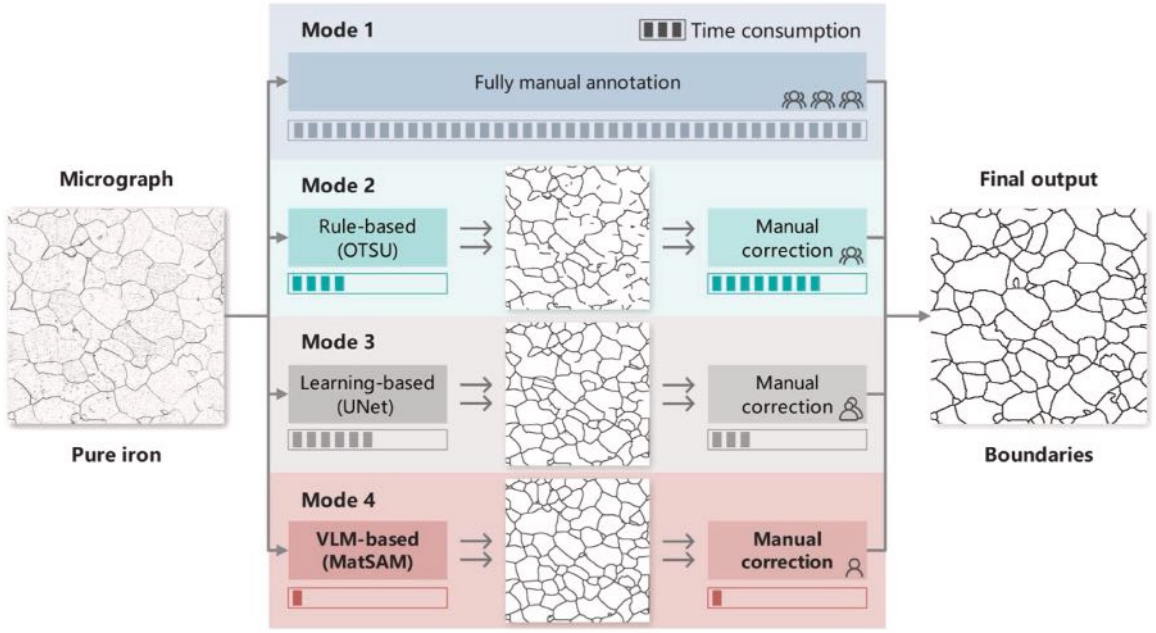


Fig. 9. Comparison of four modes of microstructure extraction.

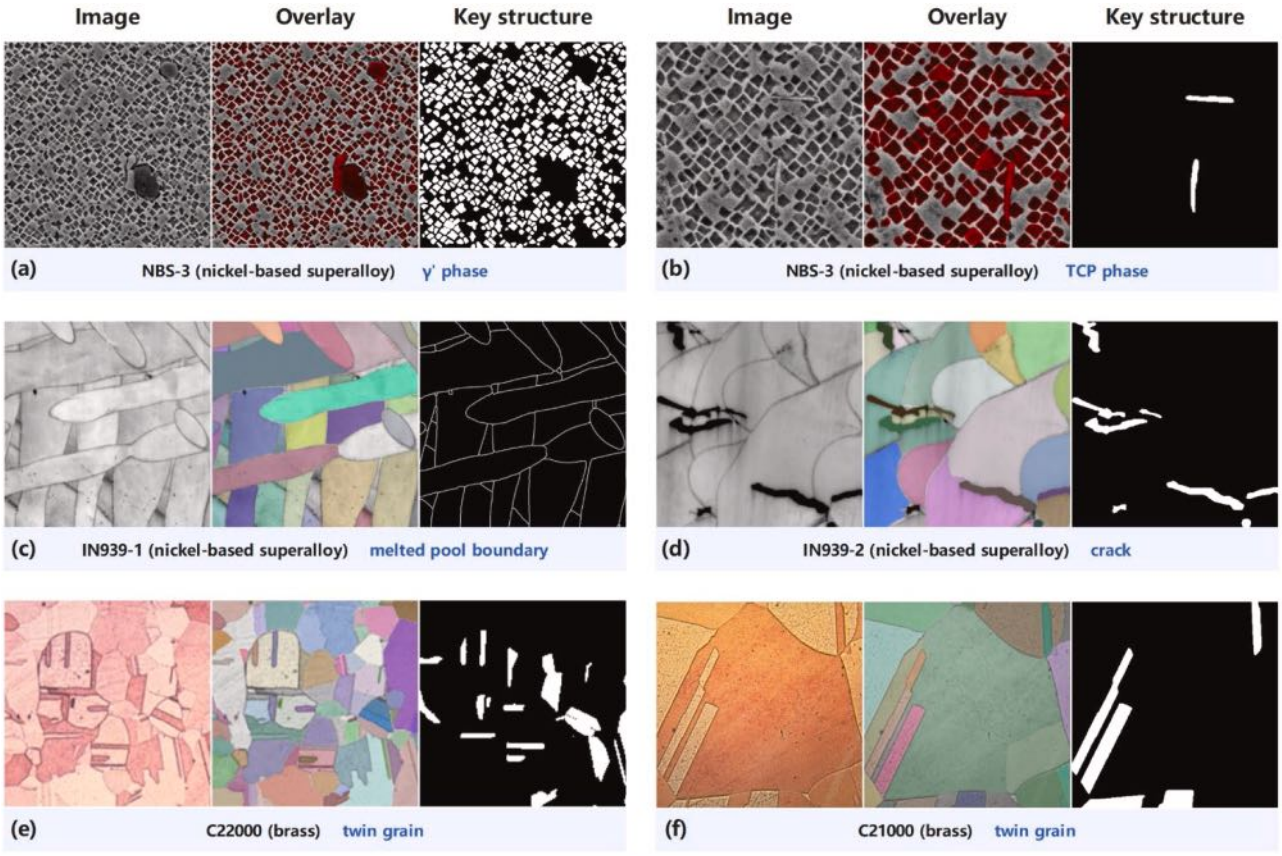


Fig. 10. Examples of extracting different microstructures. (a) NBS-3 (γ' phase), (b) NBS-3 (TCP phase), (c) IN939–1 (melted pool boundary), (d) IN939–2 (cracks), (e) C22000 brass (twin grain), and (f) C21000 brass (twin grain). On each side, the first column displays raw images; the second column shows segmentation masks overlaying the raw images for better visualization; the third column presents binary images highlighting different phases (foreground).

challenging scenarios with inconspicuous targets, substantial noise, and fine atomic details. For the example in (a) of Fig. 11, on HRTEM images of gold nanoparticles, the model fails to comprehensively extract multiple particles, likely due to the size disparity between small atomic features and the overall region, as well as variations in atomic orientations within individual particles. These factors reduce mask confidence and lead to severe under-segmentation, highlighting MatSAM's difficulty in effectively handling fine-grained and noisy datasets.

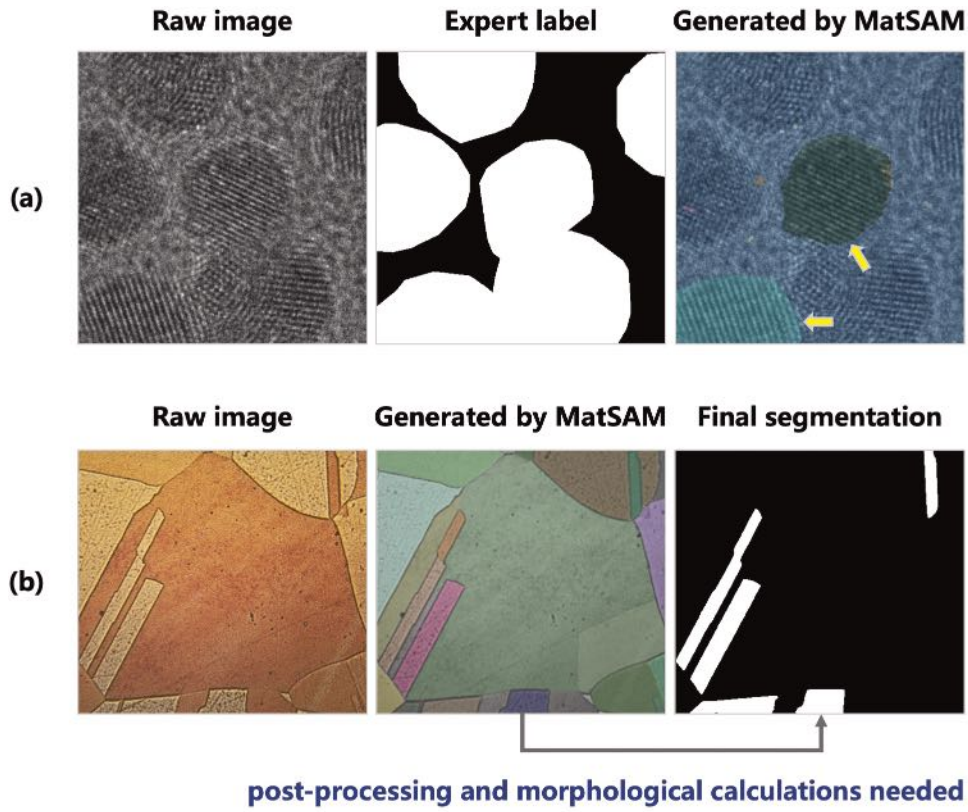


Fig. 11. Demonstrations of limitations of MatSAM. (a) the failure example on the HRTEM image of gold nanoparticles. (b) the process of filtering target twin grain regions on the brass micrograph.

Table 4
Inference time and corresponding segmentation performance of MatSAM and other methods on public LCS and NBS-1 datasets.

Dataset	Method	Metric	Segmentation performance	Inference time per image (ms)
LCS (1360 × 600) (Polycrystalline)	Watershed	ARI	0.54	6.12
	OTSU		0.64	**4.54**
	Canny		0.73	16.48
	SAM		0.86	1850.35
	MatSAM		**0.96**	1926.35
NBS-1 (512 × 512) (Multiphase)	OTSU	IoU	0.75	**0.53**
	Adaptive		0.45	0.58
	SAM		0.85	1575.24
	MatSAM		**0.91**	1812.67

Bold values indicate the **best** results.

Second, even though MatSAM performs well in binary foreground-background segmentation tasks, delivering accurate and efficient results for identifying regions most relevant to the analysis, it should be acknowledged that for images with relatively complex semantic components, further post-processing is often required to extract specific targets. As shown in (b) of Fig. 11, in segmenting brass micrographs, parameters like circularity, contour geometry, area, and grayscale range were calculated for each mask. By setting tailored filtering conditions, elongated twin grains are effectively distinguished. This reliance on additional semantic filtering underscores a limitation of MatSAM in addressing more intricate segmentation demands.

Lastly, while MatSAM achieves impressive segmentation accuracy, it comes at a higher computational cost, as shown in Table 4. On the LCS dataset, it achieves an ARI score of 0.96 with an inference time of 1926.35 ms per image, compared to OTSU's 4.54 ms. Similarly, on the NBS-1 dataset, our method obtains the highest IoU score of 0.91 but requires 1812.67 ms. Although its high accuracy significantly reduces the need for manual corrections, the trade-off in inference speed presents challenges for real-time applications, emphasizing the need for further optimization.

5. Conclusion

In this study, a novel training-free approach, MatSAM, was proposed to efficiently extract material microstructures based on SAM. Without the need for manual annotations, it precisely identified 11 kinds of material microstructures. Combining two key metrics, ARI (for polycrystalline images) and IoU (for multiphase images), MatSAM achieved an average relative improvement of 35.4 % over the best-performing conventional rule-based methods and 13.9 % over the original SAM. On four public microstructure segmentation datasets of various materials—nickel-based superalloys (NBS-1/2), ultrahigh carbon steel (UHCS), and AlZn alloy (AZA)—the average IoU attained by MatSAM was 7.5 % higher than that of the corresponding specialist DL models. The proposed MatSAM tackles the challenge of high time and labor costs in the quantitative characterization of microstructures. Meanwhile, it has obtained improved generalization capability, enabling a single model to be reused for different microstructures, including grain boundaries, phases, and defects. The methodology and empirical results of this study deliver a new perspective on addressing key challenges in large-scale, automated microstructural information extraction.

Data availability

The test data of publicly released datasets can be found in the referenced articles or our code repository: https://github.com/USTB-AI3DVIP/matsam. All the raw data of in-house datasets are stored at the University of Science and Technology Beijing and are available upon request from the corresponding author. The source code of this work is available in this repository: https://github.com/

USTB-AI3DVIP/matsam.

CRediT authorship contribution statement

Changtai Li: Writing – original draft, Methodology, Formal analysis, Conceptualization. **Xu Han:** Writing – original draft, Validation, Investigation, Data curation. **Chao Yao:** Writing – review & editing, Validation. **Yu Guo:** Writing – review & editing, Conceptualization. **Zixin Li:** Resources, Data curation. **Lei Jiang:** Writing – review & editing. **Wei Liu:** Resources, Data curation. **Haiyou Huang:** Writing – review & editing, Data curation. **Huadong Fu:** Writing – review & editing, Supervision, Resources, Conceptualization. **Xiaojuan Ban:** Writing – review & editing, Supervision, Project administration.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work is supported by the National Major Science and Technology Projects of China (2024ZD0608100), the National Natural Science Foundation of China (52425409, U22A2022, 62332017, 62303043), and Xiaomi Young Scholars Program.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.actamat.2025.120962.

References

[1] M. Koyama, Z. Zhang, M. Wang, D. Ponge, D. Raabe, K. Tsuzaki, H. Noguchi, C. C. Tasan, Bone-like crack resistance in hierarchical metastable nanolaminate steels, Science 355 (2017) 1055–1057, https://doi.org/10.1126/science.aal2766.

[2] D. Ren, C. Wang, X. Wei, Q. Lai, W. Xu, Building a quantitative composition-microstructure-property relationship of dual-phase steels via multimodal data mining, Acta. Mater. 252 (2023) 118954, https://doi.org/10.1016/j.actamat.2023.118954.

[3] H. Zhang, H. Fu, S. Zhu, W. Yong, J. Xie, Machine learning assisted composition effective design for precipitation strengthened copper alloys, Acta Mater. 215 (2021) 117118, https://doi.org/10.1016/j.actamat.2021.117118.

[4] A. Khosravani, A. Cecen, S.R. Kalidindi, Development of high throughput assays for establishing process-structure-property linkages in multiphase polycrystalline metals: application to dual-phase steels, Acta. Mater. 123 (2017) 55–69, https://doi.org/10.1016/j.actamat.2016.10.033.

[5] L. Jiang, H. Fu, H. Zhang, J. Xie, Physical mechanism interpretation of polycrystalline metals' yield strength via a data-driven method: a novel hall–petch relationship, Acta. Mater. 231 (2022) 117868, https://doi.org/10.1016/j.actamat.2022.117868.

[6] A.R. Durmaz, M. Müller, B. Lei, A. Thomas, D. Britz, E.A. Holm, C. Eberl, F. Mücklich, P. Gumbsch, A deep learning approach for complex microstructure inference, Nat. Commun. 12 (2021) 6272, https://doi.org/10.1038/s41467-021-26565-5.

[7] Z.-L. Wang, Y. Adachi, Property prediction and properties-to-microstructure inverse analysis of steels by a machine-learning approach, Mater. Sci. Eng. A 744 (2019) 661–670, https://doi.org/10.1016/j.msea.2018.12.049.

[8] X.C. Li, J.X. Zhao, J.H. Cong, R.D.K. Misra, X.M. Wang, X.L. Wang, C.J. Shang, Machine learning guided automatic recognition of crystal boundaries in bainitic/martensitic alloy and relationship between boundary types and ductile-to-brittle transition behavior, J. Mater. Sci. Technol. 84 (2021) 49–58, https://doi.org/10.1016/j.jmst.2020.12.024.

[9] M.D. Uchic, M.A. Groeber, D.M. Dimiduk, J.P. Simmons, 3D microstructural characterization of nickel superalloys via serial-sectioning using a dual beam FIB-SEM, Scr. Mater. 55 (2006) 23–28, https://doi.org/10.1016/j.scriptamat.2006.02.039.

[10] Z. Feng, R. Fu, C. Lin, G. Wu, T. Huang, L. Zhang, X. Huang, TEM-based dislocation tomography: challenges and opportunities, Curr. Opin. Solid State Mater. Sci. 24 (2020) 100833, https://doi.org/10.1016/j.cossms.2020.100833.

[11] J. Na, S.-J. Kim, H. Kim, S.-H. Kang, S. Lee, A unified microstructure segmentation approach via human-in-the-loop machine learning, Acta Mater. 255 (2023) 119086, https://doi.org/10.1016/j.actamat.2023.119086.

[12] N.J. Szymanski, B. Rendy, Y. Fei, R.E. Kumar, T. He, D. Milsted, M.J. McDermott, M. Gallant, E.D. Cubuk, A. Merchant, H. Kim, A. Jain, C.J. Bartel, K. Persson, Y. Zeng, G. Ceder, An autonomous laboratory for the accelerated synthesis of novel materials, Nature 624 (2023) 86–91, https://doi.org/10.1038/s41586-023-06734-w.

[13] D.J. Rowenhorst, L. Nguyen, A.D. Murphy-Leonard, R.W. Fonda, Characterization of microstructure in additively manufactured 316L using automated serial sectioning, Curr. Opin. Solid State Mater. Sci. 24 (2020) 100819, https://doi.org/10.1016/j.cossms.2020.100819.

[14] M. Wei, L. Zhao, L. Jiang, L. Yang, H. Wang, Development of high-throughput rapid heat-treatment and characterization process, Acta Mater. 261 (2023) 119365, https://doi.org/10.1016/j.actamat.2023.119365.

[15] B. Ma, X. Wei, C. Liu, X. Ban, H. Huang, H. Wang, W. Xue, S. Wu, M. Gao, Q. Shen, M. Mukeshimana, A.O. Abuassba, H. Shen, Y. Su, Data augmentation in microscopic images for material data mining, Npj Comput. Mater. 6 (2020) 1–9, https://doi.org/10.1038/s41524-020-00392-6.

[16] J. Stuckner, B. Harder, T.M. Smith, Microstructure segmentation with deep learning encoders pre-trained on a large microscopy dataset, Npj Comput. Mater. 8 (2022) 200, https://doi.org/10.1038/s41524-022-00878-5.

[17] K. Choudhary, B. DeCost, C. Chen, A. Jain, F. Tavazza, R. Cohn, C.W. Park, A. Choudhary, A. Agrawal, S.J.L. Billinge, E. Holm, S.P. Ong, C. Wolverton, Recent advances and applications of deep learning methods in materials science, Npj Comput. Mater. 8 (2022) 1–26, https://doi.org/10.1038/s41524-022-00734-6.

[18] J. Canny, A computational approach to edge detection, IEEE Trans. Pattern Anal. Mach. Intell. PAMI-8 (1986) 679–698, https://doi.org/10.1109/TPAMI.1986.4767851.

[19] N. Otsu, A threshold selection method from gray-level histograms, IEEE Trans. Syst. Man Cybern. 9 (1979) 62–66, https://doi.org/10.1109/TSMC.1979.4310076.

[20] P. Roy, S. Dutta, N. Dey, G. Dey, S. Chakraborty, R. Ray, Adaptive thresholding: a comparative study, in: 2014 International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT), 2014, pp. 1182–1186, https://doi.org/10.1109/ICCICCT.2014.6993140.

[21] Y. Dong, C. Su, P. Qiao, L. Sun, Microstructural crack segmentation of three-dimensional concrete images based on deep convolutional neural networks, Constr. Build. Mater. 253 (2020) 119185, https://doi.org/10.1016/j.conbuildmat.2020.119185.

[22] Y. Han, R. Li, B. Wang, L. Ruan, Q. Chen, A pseudo-labeling based weakly supervised segmentation method for few-shot texture images, Expert Syst. Appl. 238 (2024) 122110, https://doi.org/10.1016/j.eswa.2023.122110.

[23] K. He, R. Girshick, P. Dollar, Rethinking ImageNet pre-training, in: 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea (South), IEEE, 2019, pp. 4917–4926, https://doi.org/10.1109/ICCV.2019.00502.

[24] J. Yeom, T. Stan, S. Hong, P.W. Voorhees, Segmentation of experimental datasets via convolutional neural networks trained on phase field simulations, Acta Mater. 214 (2021) 116990, https://doi.org/10.1016/j.actamat.2021.116990.

[25] T. Stan, Z.T. Thompson, P.W. Voorhees, Optimizing convolutional neural networks to perform semantic segmentation on large materials imaging datasets: x-ray tomography and serial sectioning, Mater. Charact. 160 (2020) 110119, https://doi.org/10.1016/j.matchar.2020.110119.

[26] Z. Cao, Q. Liu, Q. Liu, X. Yu, J.J. Kruzic, X. Li, A machine learning method to quantitatively predict alpha phase morphology in additively manufactured Ti-6Al-4V, Npj Comput. Mater. 9 (2023) 1–15, https://doi.org/10.1038/s41524-023-01152-y.

[27] Q. Luo, E.A. Holm, C. Wang, A transfer learning approach for improved classification of carbon nanomaterials from TEM images, Nanoscale Adv 3 (2021) 206–213, https://doi.org/10.1039/D0NA00634C.

[28] J. Na, J. Lee, S.-H. Kang, S.-J. Kim, S. Lee, Label-free grain segmentation for optical microscopy images via unsupervised image-to-image translation, Mater. Charact. 206 (2023) 113410, https://doi.org/10.1016/j.matchar.2023.113410.

[29] M. Awais, M. Naseer, S. Khan, R.M. Anwer, H. Cholakkal, M. Shah, M.-H. Yang, F. S. Khan, Foundation models defining a new era in vision: a survey and outlook, IEEE Trans. Pattern Anal. Mach. Intell. (2025) 1–20, https://doi.org/10.1109/TPAMI.2024.3506283.

[30] A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A.C. Berg, W.-Y. Lo, P. Dollar, R. Girshick, Segment anything, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2023, pp. 4015–4026. https://openaccess.thecvf.com/content/ICCV2023/html/Kirillov_Segment_Anything_ICCV_2023_paper.html. accessed November 26, 2024.

[31] W. Ji, J. Li, Q. Bi, T. Liu, W. Li, L. Cheng, Segment anything is not always perfect: an investigation of SAM on different real-world applications, Mach. Intell. Res. 21 (2024) 617–630, https://doi.org/10.1007/s11633-023-1385-0.

[32] J. Ma, Y. He, F. Li, L. Han, C. You, B. Wang, Segment anything in medical images, Nat. Commun. 15 (2024) 654, https://doi.org/10.1038/s41467-024-44824-z.

[33] X. Ma, Q. Wu, X. Zhao, X. Zhang, M.-O. Pun, B. Huang, SAM-assisted remote sensing imagery semantic segmentation with object and boundary constraints, IEEE Trans. Geosci. Remote Sens. 62 (2024) 1–16, https://doi.org/10.1109/TGRS.2024.3443420.

[34] S. Khan, M. Naseer, M. Hayat, S.W. Zamir, F.S. Khan, M. Shah, Transformers in Vision: a survey, ACM Comput. Surv. 54 (2022), https://doi.org/10.1145/3505244, 200:1-200:41.

[35] M.-H. Guo, T.-X. Xu, J.-J. Liu, Z.-N. Liu, P.-T. Jiang, T.-J. Mu, S.-H. Zhang, R. R. Martin, M.-M. Cheng, S.-M. Hu, Attention mechanisms in computer vision: a survey, Comput. Vis. Media 8 (2022) 331–368, https://doi.org/10.1007/s41095-022-0271-y.

[36] G. Bradski, The OpenCV Library, Dr. Dobb's J.: Software Tools Profess. Prog. 25 (2000) 120–123.

[37] D. Steinley, Properties of the Hubert-Arable Adjusted Rand Index, Psychol. Methods 9 (2004) 386–396, https://doi.org/10.1037/1082-989X.9.3.386.

[38] J.M. Santos, M. Embrechts, On the use of the adjusted Rand index as a metric for evaluating supervised classification, in: Artificial Neural Networks – ICANN 2009, Berlin, Heidelberg, Springer, 2009, pp. 175–184, https://doi.org/10.1007/978-3-642-04277-5_18.

[39] S.A. Chowdhury, M.F.N. Taufique, J. Wang, M. Masden, M. Wenzlick, R. Devanathan, A.L. Schemer-Kohrn, K.S. Kappagantula, Automated grain boundary (GB) segmentation and microstructural analysis in 347H stainless steel using deep learning and multimodal microscopy, Integr. Mater. Manuf. Innov. 13 (2024) 244–256, https://doi.org/10.1007/s40192-023-00305-7.

[40] M. Levandowsky, D. Winter, Distance between sets, Nature 234 (1971) 34–35, https://doi.org/10.1038/234034a0.

[41] F. Meyer, S. Beucher, Morphological segmentation, J. Vis. Commun. Image Represent. 1 (1990) 21–46, https://doi.org/10.1016/1047-3203(90)90014-M.

[42] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, S. Chintala, PyTorch: an imperative style, high-performance deep learning library, in: Advances in Neural Information Processing Systems, 2019.

[43] H. Kakinuma, S. Ajito, T. Hojo, M. Koyama, E. Akiyama, In situ visualization of misorientation-dependent hydrogen diffusion at grain boundaries of pure polycrystalline Ni using a hydrogen video imaging system, Acta Mater. 263 (2024) 119536, https://doi.org/10.1016/j.actamat.2023.119536.

[44] S. Mohan, R. Manzorro, J.L. Vincent, B. Tang, D.Y. Sheth, E.P. Simoncelli, D. S. Matteson, P.A. Crozier, C. Fernandez-Granda, Deep denoising for scientific discovery: a case study in electron microscopy, IEEE Trans. Comput. Imaging 8 (2022) 585–597, https://doi.org/10.1109/TCI.2022.3176536.

[45] C. Lu, K. Chen, H. Qiu, X. Chen, G. Chen, X. Qi, H. Jiang, Diffusion-based deep learning method for augmenting ultrastructural imaging and volume electron microscopy, Nat. Commun. 15 (2024) 4677, https://doi.org/10.1038/s41467-024-49125-z.

[46] S. van der Walt, J.L. Schönberger, J. Nunez-Iglesias, F. Boulogne, J.D. Warner, N. Yager, E. Gouillart, T. Yu, scikit-image: image processing in Python, PeerJ 2 (2014) e453, https://doi.org/10.7717/peerj.453.

[47] M. Groeber, S. Ghosh, M.D. Uchic, D.M. Dimiduk, A framework for automated analysis and simulation of 3D polycrystalline microstructures.: part 1: statistical characterization, Acta Mater. 56 (2008) 1257–1273, https://doi.org/10.1016/j.actamat.2007.11.041.

[48] D.J. Rowenhorst, A.C. Lewis, G. Spanos, Three-dimensional analysis of grain topology and interface curvature in a β-titanium alloy, Acta Mater. 58 (2010) 5511–5519, https://doi.org/10.1016/j.actamat.2010.06.030.

[49] C. Li, R. Jiang, H. Wang, W. Xue, Y. Guo, X. Ban, DeepMMP: efficient 3D perception of microstructures from serial section microscopic images, Comput. Mater. Sci. 235 (2024) 112826, https://doi.org/10.1016/j.commatsci.2024.112826.

[50] Resources: standards & properties - copper & copper alloy microstructures: brasses, (n.d.). https://www.copper.org/resources/properties/microstructure/brasses.html (accessed January 17, 2025).