ARTICLE   OPEN

Rapid and flexible segmentation of electron microscopy data using few-shot machine learning

Sarah Akers [ID][1], Elizabeth Kautz[2], Andrea Trevino-Gavito[1,3], Matthew Olszta[2], Bethany E. Matthews[2], Le Wang[4], Yingge Du[4] and Steven R. Spurgeon [ID][2✉]

Automatic segmentation of key microstructural features in atomic-scale electron microscope images is critical to improved understanding of structure–property relationships in many important materials and chemical systems. However, the present paradigm involves time-intensive manual analysis that is inherently biased, error-prone, and unable to accommodate the large volumes of data produced by modern instrumentation. While more automated approaches have been proposed, many are not robust to a high variety of data, and do not generalize well to diverse microstructural features and material systems. Here, we present a flexible, semi-supervised few-shot machine learning approach for segmentation of scanning transmission electron microscopy images of three oxide material systems: (1) epitaxial heterostructures of $SrTiO_3/Ge$, (2) $La_{0.8}Sr_{0.2}FeO_3$ thin films, and (3) MoO_3 nanoparticles. We demonstrate that the few-shot learning method is more robust against noise, more reconfigurable, and requires less data than conventional image analysis methods. This approach can enable rapid image classification and microstructural feature mapping needed for emerging high-throughput characterization and autonomous microscope platforms.

INTRODUCTION

Material microstructures govern the functionality of many important technologies, including catalysts, energy storage devices, and emerging quantum computing architectures. Scanning transmission electron microscopy (STEM) has long served as a foundational tool to study microstructures because of its ability to simultaneously resolve structure, chemistry, and defects at atomic-scale resolution for a range of materials classes[1–3]. STEM has helped elucidate the nature of microstructural features ranging from complex dislocation networks to secondary phases and point defects, leading to refined structure–property models[2,4,5]. Traditionally, STEM images have been analyzed by a domain expert manually or semi-automatically, utilizing a priori knowledge of the system to identify microstructural features of interest. While this approach is suitable for measuring a limited number of microstructural features in small data volumes, it is impractical for samples possessing high density, rare, or noisy features[6,7]. Moreover, manual and semi-automatic approaches are difficult to scale to include multiple data modalities and cannot be performed at high speed, hindering our ability to perform in situ, complementary or correlative studies harnessing the full potential of modern instruments[8]. At a more fundamental level, variability in how such measurements are conducted and a lack of standardized approaches contributes to the broader issue of reproducibility in experimentation[9]. Though these limitations apply to all materials classes, they are particularly pronounced for complex oxides, whose properties are heavily influenced by even trace amounts of unwanted defects[10–12]. Hence, there is an urgent need to develop approaches to characterize microstructural features with greater accuracy, speed, and statistical rigor than is possible with existing methodologies.

A central challenge in quantitatively describing microscopy image data (i.e., micrographs) is the wide variety of possible microstructural features and data modalities. The same instrument that is used to examine interfaces at atomic-resolution one day may be used to examine nanoparticle morphologies or grain boundaries the next. A common goal in any study employing electron microscopy, in particular STEM, is to extract quantitative and semantically-meaningful microstructural descriptors that can be linked to underlying physical models[13,14]. For example, estimating the area fraction of a specific phase or abundance of a feature through image segmentation is an important part of understanding synthesis products and phase transformation kinetics[15–19]. Although several image segmentation methods exist (e.g., Otsu[20], the watershed algorithm[21], k-means clustering[22]), these are often not easily generalizable to different material systems, image types, and may require significant tailored image preprocessing.

Machine learning (ML) methods, specifically convolutional neural networks (CNNs), have recently been adopted for the recognition and characterization of microstructural data across length scales[23–26]. Classification tasks have been performed to either assign a label to an entire image that represents a material or microstructure class (e.g., dendritic, equiaxed, etc.)[26–29], or to assign a label to each pixel in the image so that they are classified into discrete categories[25,30–32]. The latter classification type, categorization of pixels in an image to identify local features (e.g., line defects, phases, crystal structures), is referred to as segmentation. Still, many challenges remain in the practical application of segmentation methods, such as the large data set size required for training and the difficulty of developing methods that are generalizable to a wide variety of data. In deep-learning and computer vision approaches, like a CNN, learning models of image categories has typically required large databases of labeled training examples[33], such as the large image data set available through the ImageNet database[34,35]. However, recent research has

demonstrated effective and lightweight learning techniques that require relatively few labeled training examples for the purpose of automatic segmentation and denoising[36,37]. The ability to analyze data sets on the basis of limited training data, as often encountered in microscopy[38,39], is an important frontier in materials and data science. Unfortunately, even one labeled example oftentimes includes tedious manual annotation of thousands or even millions of pixels. The architecture developed by Pelt and Sethian[37], while much less complex than existing networks and requiring fewer training points, still uses eight manually annotated images of $512 \times 512 \times 512$ cubic pixel tomographic reconstructions. Additionally, the classification or segmentation task is also dependent upon these labels, i.e., new labels must be constructed in order to change the classification/segmentation output.

Motivated by the ability of humans, and especially children, to learn novel visual concepts with sufficient previous knowledge[40] one-shot or few-shot approaches allow human-level performance with fewer and less intensively labeled images (i.e., shots) and little to no training[41,42], but there are limited studies on such methods in the materials science domain[36]. While many characterization tools may provide just a few data points, a single electron micrograph (and potentially additional imaging/spectral channels) may encompass many microstructural features of interest. The one-shot or few-shot learning concept also has significant implications for the study of transient or unstable materials, as well as those where limited samples are available for analysis due to long lead-time experimentation (such as corrosion or neutron irradiation studies). In other cases, there exists data from previous studies that may be very limited or poorly understood, for which advanced data analysis methods could be applied[43].

In this work, we present a rapid and flexible approach to recognition and segmentation of STEM images using few-shot machine learning. Three oxide materials systems were selected for model development (epitaxial heterostructures of $SrTiO_3$ (STO)/Ge, $La_{0.8}Sr_{0.2}FeO_3$ (LSFO) thin films, and MoO_3 nanoparticles) due to the range of microstructural features they possess, and their importance in semiconductor, spintronic, and catalysis applications[44,45]. While the three systems selected are all oxides imaged in the STEM-HAADF mode, these data encompass a wide range of morphologies, length scales, and microstructural features. It should also be mentioned that the features of interest in these systems and data sets (interfaces, nanoparticles) are commonly analyzed in many other material systems, such as metals and alloys. See Supplementary Note 1 for example applicability of this method to a broader data set. We demonstrate that with only 5–8 sub-images (referred to here as chips) that represent examples of a specific microstructural feature (e.g., a crystal motif or particular particle morphology), our model yields segmentation results comparable to those produced by a domain expert for all systems studied here. The successful image mapping can be attributed to the low noise sensitivity and high learning capability of few-shot machine learning in comparison to other segmentation methods (e.g., Otsu thresholding, watershed, k-means clustering, etc.). The few-shot approach rapidly identifies varying microstructural features across STEM data streams, which can inform real-time image data collection and analysis. More broadly, our findings underscore the power of image-driven machine learning to enable
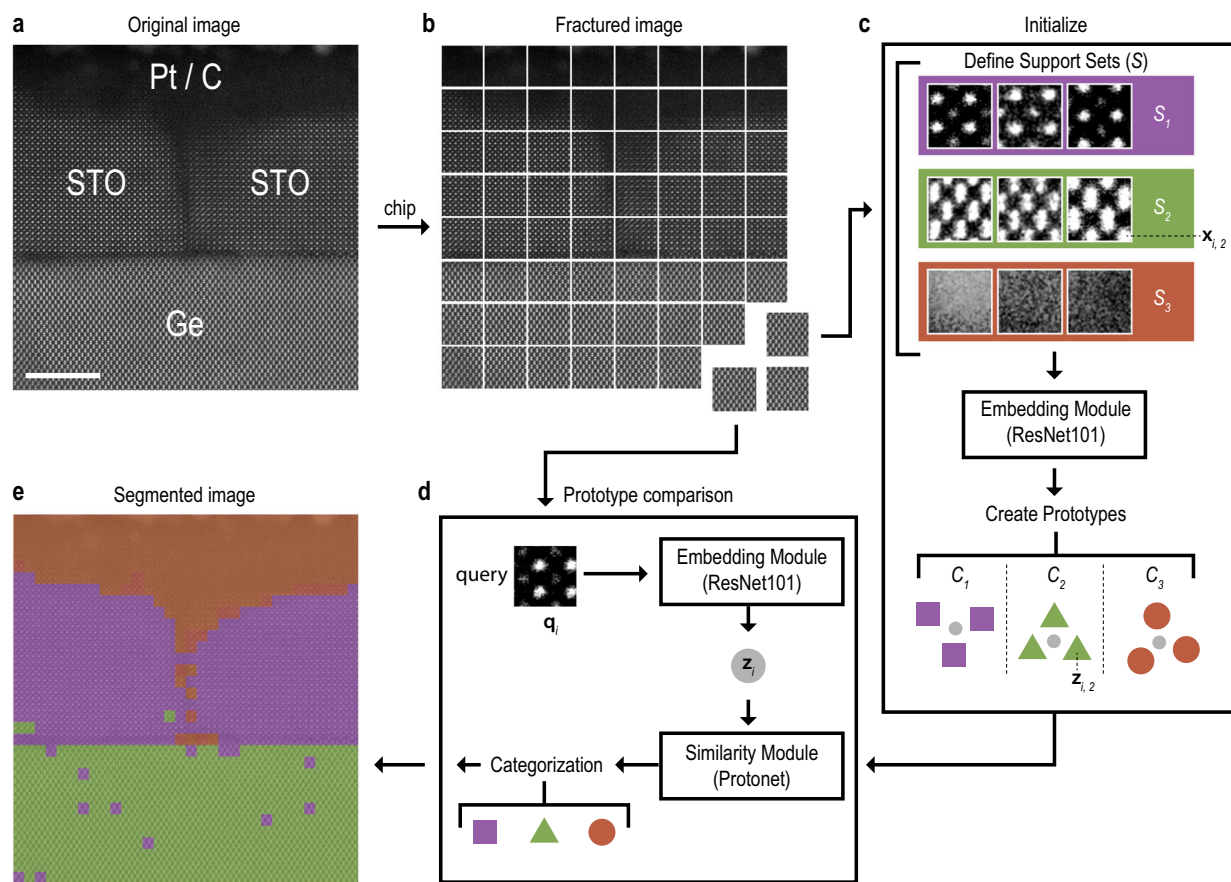
Fig. 1 Few-shot model architecture. The raw STO/Ge image (**a**) is broken into several smaller chips (**b**) and a few user defined chips are used to represent desired segmentation classes in the support set (**c**). Each chip then acts as a query and is compared against a prototype (**d**), defined by the support set, and categorized according to the minimum Euclidean distance between the query and each prototype, yielding the segmented image (**e**). Scale bar = 5 nm.

Table 1. Image preprocessing parameters listed by material system and task with respective libraries/methods for implementation.

Task	Method/Library	Material system		
		LSFO	STO/Ge	MoO$_3$
Image reading	cv2.imread	flag = 0 (grayscale)	flag = 0 (grayscale)	flag = 0 (grayscale)
Equalization	CLAHE	clipLimit = 2.0, tileGridSize = (8, 8)	clipLimit = 2.0, tileGridSize = (8, 8)	clipLimit = 2.0, tileGridSize = (8, 8)
Chipping	image_slicer	tiles = 500	tiles = 1000	tiles = 256
Resizing (pixels)	ResNet	256 × 256	256 × 256	256 × 256
LoG	skimage	max_sigma = 30, num_sigma = 8, threshold = 0.1	N/A	N/A

improved microstructural characterization for materials discovery and design.

RESULTS AND DISCUSSION

A deep-learning approach known as few-shot learning was developed for superpixel semantic segmentation of STEM images, i.e., the classification of superpixels in a STEM image. The premise of this few-shot learning model is to use very few labeled examples (<10) per class for the model to identify regions of an image that correspond to each class. This superpixel segmentation approach additionally allows us to leverage the repeating patterns typical in material microstructures and use sub-images for informing the few-shot network, bypassing the need for even a single fully labeled training image. The general approach to image segmentation using few-shot learning is schematically described in Fig. 1. This methodology involves breaking an input image into a grid of sub-images (referred herein as chips), model initialization, inference, and output of a segmented micrograph. The process of chipping relies on domain-specific knowledge of the materials microstructure, as indicated in the annotations in Fig. 1a. Most computer vision segmentation techniques provide pixel level classification, referred to as semantic segmentation. While technically more precise in terms of granularity, these methods can be more error prone, e.g., in noisy images or in instances of artifacts[46]. In cases where pixel to pixel variability ranges widely, a chip may better capture microstructural features of interest as a whole versus specific pixels. Additionally, recent developments in few-shot texture segmentation[47] may provide feasible routes to semantic segmentation for STEM images.

Preprocessing

To separate and measure distinct phases that have varying contrast in the STEM images, preprocessing of original image data was required. A histogram equalization (HE) technique designed to enhance local image qualities without introducing global artifacts termed contrast limited adaptive HE (CLAHE)[48,49] was selected for use in this work. The details of the CLAHE implementation are described in Table 1. CLAHE was first performed on original images and then the processed image was sectioned into a set of smaller sub-images, as shown in Fig. 1b. The chip size varied between 95 × 95 pixels and 32 × 32 pixels, however all chips are resized to 256 × 256 in the ResNet101 embedding module. The variable size allowed for each chip to be large enough to capture a microstructural motif and small enough to provide granularity between adjoining spatial regions, as shown in Fig. 1. The final preprocessing step is an enhancement technique[50] that marks the position and size of atomic columns using a Laplacian of Gaussians (LoG) blob detection routine[51]. This step was used on the LSFO system to enhance the extremely subtle differences between classes.

Model architecture

The few-shot model inputs the preprocessed STEM image, typically with high resolution on the order of 3000 × 3000 pixels, that has been broken down into a series of smaller chips, x_{ik}, typically not larger than 100 by 100 pixels. A handful of these chips are used as examples, or a support set, to define each of one or several classes. While most image applications for few-shot learning use disparate x_{ik} to define a support set for each class (S_k)[52–56], here S_k was created by breaking the original image into a grid of smaller sub-images (Fig. 1b). A subset of chips were labeled for each class. The set of N labeled examples for $k = 1, .., K$ classes makes up the support set defined by: $S = \{(x_1, y_1), ... (x_N, y_N)\}$, where x_i represents an image i and y_i is the corresponding true class label (Fig. 1c).

A Prototypical Network[57] was selected in this work, given its lightweight design and simplicity. This model is based on the premise that each S_k may be represented by a single *prototype*, c_k. To compute c_k, each x_{ik} is passed through an embedding function f_ϕ, which maps a D-dimensional image into an M-dimensional representation through learnable parameters ϕ. The transformed chip, or $f\phi(x_{ik}) = z_{ik}$, then creates the prototype for class k as the mean vector of the embedded support points c_k, as follows:

$$c_k = \frac{1}{N_{S_k}} \sum_{(z_i, y_i) \in S_k} z_i \tag{1}$$

After class prototypes are created, an untrained Prototypical Network classifies a new data point, or *query* q_i, by first transforming the query through the embedding function and then calculating a distance, e.g. Euclidean distance, between the embedded query vector and each of the class prototype vectors (Fig. 1d). After the distances are computed, a softmax normalizes the distance into class probabilities, where the class with the highest probability becomes the label for the query[57]. The final output of the model, for each q_i, is the respective class label (Fig. 1e).

Model Inference

In order to quantify phase fractions in a STEM image (which can range from nm to μm in spatial dimension) each chip is used as a query point, q_i, so that the entire set of query points, Q, makes up the full image. The size of Q is directly proportional to the size of each chip and the size of the full image, as shown in Table 2. All q_i first go through the embedding function and distances to each prototype are computed using the selected distance function. The network then produces a distribution over each of the K classes by computing a softmax over the distances and assigns a class label according to the highest normalized value[57]. In the current implementation of the Prototypical Network, query chips must be assigned to one of the pre-selected class prototypes. To account for unknown features, the user is advised to add an additional support class.

Table 2. Few-shot specific implementation information including model parameters, image information, and computing device used for the images shown in Fig. 2.

Parameter	Material system		
	LSFO	STO/Ge	MoO$_3$
Support classes	2	3	3
Batch size	8	8	16
Image size (pixels)	2048 × 2048	3042 × 3044	512 × 512
Total chips	500	1000	256
Chip size (pixels)	34 × 35	95 × 95	32 × 32
Encoding	ResNet101	ResNet101	ResNet101
	pretrained = True	pretrained = True	pretrained = True
Distance metric	Euclidean	Euclidean	Euclidean
Similarity module	Protonet	Protonet	Protonet
Normalization	Softmax	Softmax	Softmax
Training	None	None	None
Computing	device = CPU	device = CPU	device = CPU

The "pretrained" flag indicates that ResNet weights trained on ImageNet are used in encoding.

The model-specific implementation and parameters are given in Table 2. While the selection of model parameters is often tedious, specific model parameters in the few-shot context are generally straightforward, since it is often possible to leverage pretrained models for the embedding architecture. Here, a residual network with 101 layers, ResNet101[58], was used as the embedding architecture. ResNet was specifically selected owing to its success in several related image recognition tasks[58]. Any number of networks may be used for encoding, including more lightweight architectures such as smaller MS-D-Nets[37], U-Nets[59], or highly specialized networks such as DefectSegNet[31]. ResNet is a popular model that is widely available in several programming languages with model weights available, making learned knowledge easily transferable to STEM-specific tasks. Model weights for ResNet101 are available from PyTorch[60] pytorch/vision v0.6.0, as trained on the image database ImageNet[61]. We note that other networks optimized for microscope images may yield performance gains in training and inference, as well as permit greater batch sizes for improved regularization. Additionally, the Euclidean distance metric was used, since this metric generally performs well across a wide variety of benchmark data sets and classification tasks[57]. These pretrained models come with specified parameters and trained model weights. However, any embedding architecture may be used, especially those well-suited for segmentation[62].

The similarity module can be any few-shot or meta-learning architecture as well; however, Protonets are generally simple and easy to implement. Parameters not necessarily specific to the models—namely chip size and batch size—should take the size of each distinct micrograph into consideration in addition to computational memory capacity. A chip should generally encompass a single micrograph and may take trial and error depending on the size of the full image and magnification. The batch size is simply the number of chips to evaluate at once. Generally, a machine with at least 16 GB of RAM and 2.7 GHz of processing power can reasonably compute model predictions at a rate of about 1 chip per 0.5 s, with a batch size of 100 chips measuring 64 × 64 pixels. The compute time naturally depends on processing power in addition to the chip size and the number of parameters in the embedding module. In the case of training a few-shot model, rather than simple inference as shown here, at least one

GPU is necessary and may take several days to reach convergence given a sizeable database, such as a typical image database like ImageNet[61] contains 14 million images. The scope of this manuscript will only discuss the former, using an untrained few-shot model and pure inference to make judgments about an image.

Classification

The segmentation output of few-shot classification using the Prototypical architecture for three oxide systems is shown in Fig. 2. The model output is a superpixel classification, i.e., every pixel that belongs to a chip receives the same label and corresponding color, much in the same way other computer vision applications approach segmentation[63]. Here, the support set classes define the set of possible output labels. The percentage of chips belonging to each class, shown in Fig. 2 (right), can be scaled from percentages to area using pixel scale conversions for a total area estimate for each distinct micrograph.

The STO/Ge system presents a particular challenge for most image analysis techniques in that the contrast varies irregularly across the whole image. The sample also contains multiple interfaces and is representative of typical thin film-substrate imaging data. The selected LSFO image shows a secondary phase in the perovskite-structured matrix, and the secondary phase appears to have a gradient from top to bottom, which drastically diminishes the very subtle differences between the two micrographs. Separation of the two interpenetrating microstructural domains is necessary to understand the synthesis process and resulting properties, such as electrical conductivity. While preprocessing can adjust for some of these irregularities, traditional threshold-based segmentation techniques such as Otsu's Method[20] and watershedding[21] are not robust enough for a consistent solution and even adaptive methods can fail against a gradient and certainly when applied generally across multiple images. Another approach to phase discovery has been successfully demonstrated based on the use of a sliding FFT[18]. While powerful, this approach requires periodicity in the image and is less amenable to lower magnification images of the kind shown Fig. 2c. In this case, a convolution of contrast, edge, and sampling parameters makes straightforward interpretation of the FFT difficult. In addition, this approach is less well-suited to classifying amorphous materials, which may possess correlated but not periodic order that is difficult to quantify locally. The few-shot technique is not completely immune to these irregularities as seen in Fig. 2a, where the segmented micrograph contains a handful of misclassified chips. These misclassifications reveal some sensitivity to support set selection, which is an important topic for additional study. In general, we find that classification based on support sets containing canonical features (well-resolved, uniform) outperforms classification based on support sets containing outlier (noisy, incomplete) features. In the LSFO system, few-shot is slightly more inconsistent in identifying the (green) microstructural features, as shown in Fig. 2b. However, these issues may be corrected with a post hoc spatial smoothing. For example, chips completely surrounded by one label within some radius would weight the class probability, or adjustments in the chips that define the support set. Despite some irregularities, the few-shot method is much more robust to noise within the full input image when compared against other segmentation techniques. Additionally, this few-shot method is easily generalizable to several different material systems, since a single support set defined by one image can be applied without adjustment to multiple images of the same type for an unmatched time savings in the analysis of image series. This feature is particularly important in the case of large area mapping, as shown for the MoO$_3$ nanoparticles, where it is necessary to collect image montages to survey the wide variety of possible particle
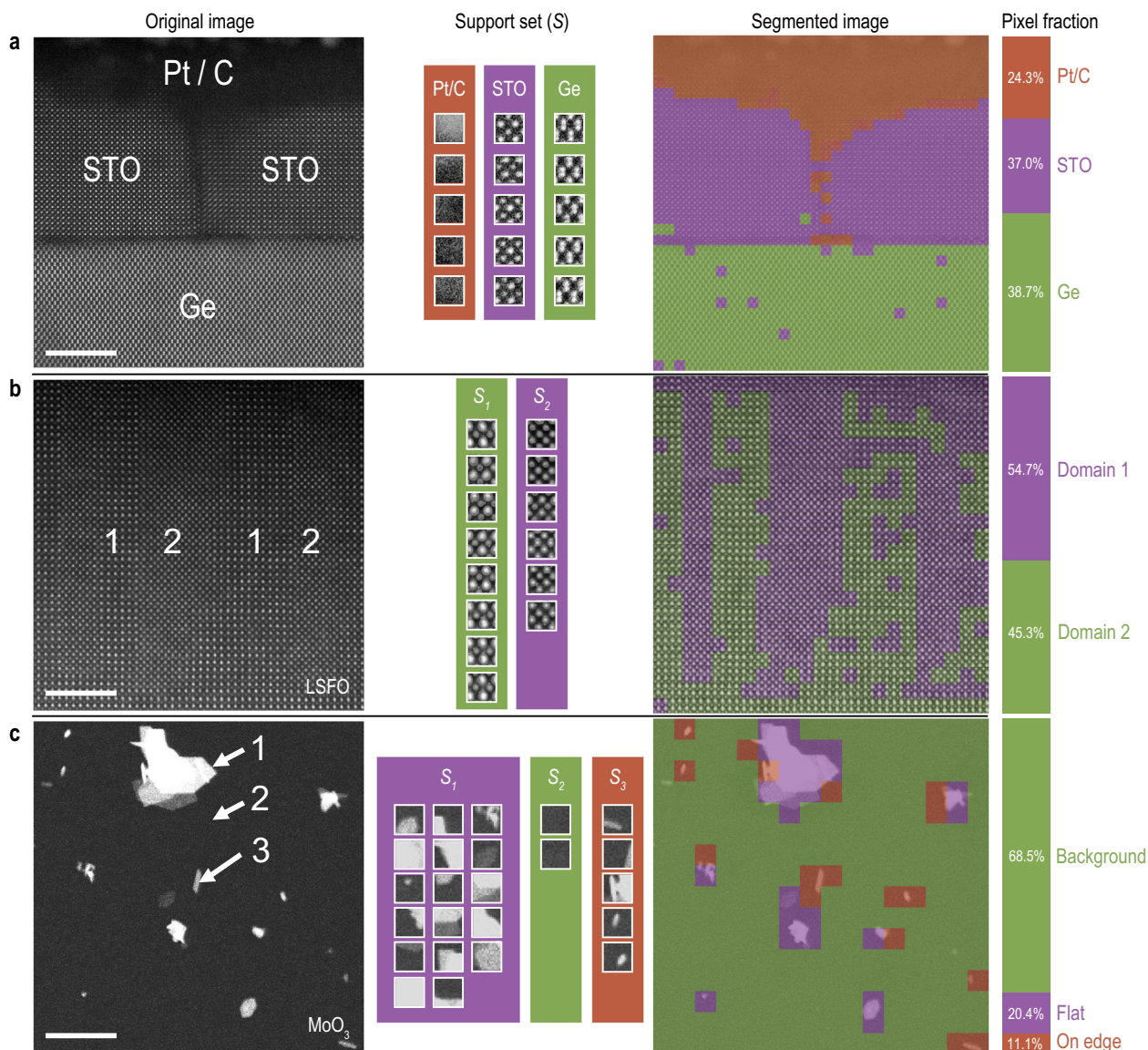
Fig. 2 Few-shot segmentation results. a–c Analysis of three materials systems: STO/Ge, LSFO, and MoO$_3$. Raw images (left), respective support sets (center), and output of few-shot segmentation (right) are shown. The estimated proportion of each support set class is shown in the inset bar chart for each system. Scale bars = 5 nm (**a–b**) and 5 μm (**c**).
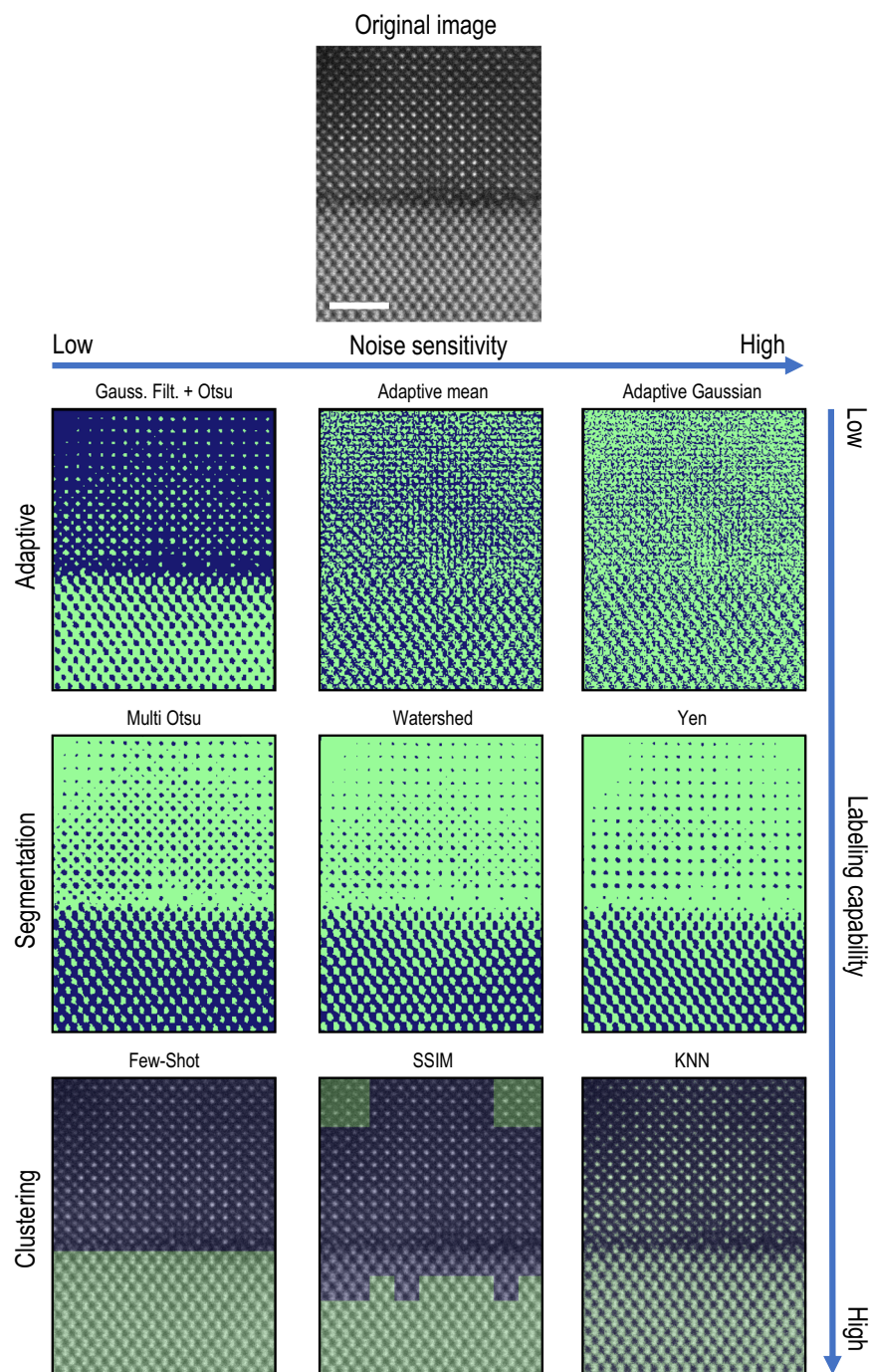
morphologies. Here again the few-shot method successfully distinguishes several nanoparticle orientations from the carbon support background, with minimal instances of inaccurate labeling. We note the ability of the few-shot approach to accommodate the visual complexity of S_1 seen in Fig. 2c, with a range of shapes, contrast, and sizes defining this flat category. While S_1 here is defined with several more chips than the others, the model is able to reasonably perform a segmentation task impossible using contrast based methods alone. Overall, we find that the model generalizes well to different material systems containing varying microstructural features.

Comparison to other methods

Initially, several image analysis techniques were explored in an effort to quantify microstructral features of interest in specific micrographs, i.e., segmentation. It was immediately obvious that no single segmentation method would perform well in the absence of preprocessing steps, such as contrast adjustments, smoothing, and sharpening. Ideally, the aim of preprocessing in

these analyses is to globally minimize artificial contrast textures and locally emphasize object edges, a critical noise reduction step for most segmentation routines[64]. Given that preprocessing and segmentation are often inseparable[65], we examine comparable segmentation methods in the context of both segmentation and preprocessing together. In an effort to compare the few-shot approach with more widely-used segmentation methods, an example image from the STO/Ge system was analyzed using techniques with varying noise sensitivity and segmentation capabilities, with results shown in Fig. 3.

The simplest approach to segmentation falls under a family of thresholding techniques shown in the first row of Fig. 3. The three methods shown in the top row are designed to separate pixels in an image into two or more classes, based on an intensity threshold. The threshold in these methods is determined using information about the distribution of pixel intensities either globally (top row left) or locally using a neighborhood of pixels (top row center and right). The neighborhood methods are commonly more sensitive to noise, while Otsu's more global

Original image

Low　　　　　Noise sensitivity　　　　　High

Gauss. Filt. + Otsu　　　Adaptive mean　　　Adaptive Gaussian

Adaptive

Low

Multi Otsu　　　　Watershed　　　　Yen

Segmentation

Labeling capability

Few-Shot　　　　SSIM　　　　KNN

Clustering

High

Fig. 3 Comparison of analysis techniques. An image from the STO/Ge system (top) is analyzed with a suite of image processing techniques with varying noise sensitivity and labeling capabilities. The thresholding techniques (top) typically separate background from foreground in an image. The strict segmentation techniques (middle) have the ability to separate the image further into multiple classes, though the classes are defined solely on pixel intensity. The clustering techniques (bottom) separate an image into classes based calculated image properties such as a centroid (KNN), structural similarity threshold (SSIM), or a prototype (few-shot). Scale bar = 2 nm.

technique appears to separate foreground pixels (light) from background (dark) relatively well.

Moving beyond simple thresholding, we begin to look towards separating pixels into classes other than background and foreground. The segmentation methods shown in Fig. 3 typically have the ability to separate intensities into multiple classes again defined by the distribution of pixel intensities in the image. Two classes are specified for these routines in order to demonstrate the premise that, ideally, the image could be segmented according to the two distinct micrographs. These approaches also typically involve blurring filters and/or morphological operations[66] in order to remove pixels that are not a part of a larger group or shape. While shape edges are more defined in the middle row of Fig. 3 than in the top row, we note that the resulting segmentation still appears to be background/foreground and misses the distinction between micrograph structures. One obvious limitation of a direct implementation of these methods is that the resulting classes will always be based on intensity and not on the size or shape of the

underlying micrographs. It may be possible to layer these methods with a shape detection routine where shapes of approximately the same size may be clustered into the same class. However, we found that clustering shapes post foreground/background segmentation was not able to distinctly separate microstructural features in an unsupervised manner, i.e. without tedious and manual intervention.

Rather than adding a shape clustering routine to an already segmented image, we implemented cluster based methods on either the raw image, or neighborhoods of the raw image in the bottom row of Fig. 3. A common unsupervised K-Nearest Neighbors (KNN) clustering method is shown in the bottom row right, where we again see clustering results based on pixel intensity or background/foreground separation. Bottom row middle shows the first non-intensity based approach. An average structural similarity index measure (SSIM) is computed pairwise for 100×100 pixel non-overlapping neighborhoods as a measure of similarity between regions. The average SSIM for each neighborhood is a bimodal distribution that can be grouped into two classes as shown in bottom row center of Fig. 3. However, the cutoff in SSIM must be manually determined. Lastly, the few-shot segmentation technique described in this manuscript is shown in bottom row left, where we see perfect segmentation between the two regions of distinct micrographs. We have included few-shot as a clustering technique, since a neighborhood is compared to a prototype analogous to the way clustering techniques compare to a centroid.

The results presented here have been assessed qualitatively, given that no standard comparison data sets exist to address the fidelity of few-shot learning in standard quantitative practice. This gap also points directly to the high cost of manually labeling data and perhaps opens an opportunity for few-shot assisted annotation. We have attempted a comparison to hand-labeled data, as shown in Supplementary Note 4, which shows overall good classification with some variation at feature boundaries, but further work is needed to more rigorously assess performance. Additionally, we have shown only one type of few-shot architecture here, though a number of other architectures and configurations may well outperform the simplicity of the Prototypical Network, including Siamese Networks[67], Relation Networks[68], Conditional Networks[56], etc. However, we have found that a Relational-Conditional Network may be more sensitive to noise and less adept to out of distribution samples. (See Supplementary Note 2 for a Relational-Conditional Network result).

In summary, here we developed a flexible few-shot learning approach to STEM image segmentation that can significantly accelerate mapping and identification of phases, defects, and other microstructural features of interest in comparison to more traditional image processing methods. We studied three different materials systems (STO/Ge, LSFO, and MoO_3), with varying atomic-scale features and hence diversity in image data for model development. Segmented images using the few-shot learning approach show good qualitative agreement with original micrographs.

When compared to other techniques, we find that noise sensitivity and/or labeling capability remain challenges for adaptive segmentation and clustering algorithms. We note that in the present study we have not examined the robustness of the few-shot approach to image artifacts or high noise levels because of a lack of established ground truth for quantitatively benchmarking the present data sets. These topics merit additional future study. In addition, the effect of orientational disorder was not systematically considered, but recent work[69,70] has shown that rotationally invariant variational autoencoders may help address such disorder. However, analysis of a larger MoO_3 image set shown in Supplementary Note 3, demonstrates the qualitative robustness of the approach to both feature intensity variation and rotation. The few-shot techniques explored in this manuscript provide powerful resources to combat these issues and remain flexible enough to accommodate a suite of materials. While few-shot machine learning has been increasingly successful in rapidly generalizing to new classification tasks containing only a few samples with supervised information, it is a known problem that the empirical risk minimizer can be slightly unreliable[71], leading to uncertainty in the reliability of the model for any given support set. We can mitigate some of this uncertainty with careful selection of the support set in order to avoid driving the model toward a non-optimal solution with mistakes in the support set, for instance. Another way we can alleviate some of the uncertainty in solutions is to actually train the few-shot model with large volumes of labeled data, when available, so that new classification tasks already have the benefit of optimization, even in a completely new and different material systems. A post-processing routine of spatial structural statistics may also provide a means for smoothing spurious segmentation predictions from an otherwise uniform material region; however, it is possible that a true signal could be smoothed out as an artifact with such approaches.

Aside from uncertainties in performance, the benefits in model generalizability also limit the amount of physically meaningful information we can extract from the model itself. One route for future exploration is the use of multiple data streams, such as spectral (electron energy loss spectroscopy (EELS) and energy-dispersive X-ray spectroscopy (EDS)), as well as diffraction (4D-STEM) in a multi-modal few-shot model. Insights in the multi-modal context would help to extract more physically meaningful information and likely enhance the segmentation performance. In theory, this same model should generalize to a wide variety of STEM images and our preliminary results on other systems are indicative of this. Yet still unknown are the effects of the size and variety of the examples in a support set for a given class. Simulation studies are being designed to help answer these questions and more, including the possibility of using other models for embedding and the benefit of model training. Recent work in few-shot for texture segmentation[47] also indicates that pixel level annotation may be possible, pushing the accuracy of this approach beyond chip-level resolution. Ideally, few-shot segmentation can also help curate and annotate these large training sets, something that is otherwise extremely costly to do manually and at scale. In summary, this approach offers a potentially powerful means to standardize and automate the analysis of materials microstructures for a single image, paving the way for high-throughput characterization architectures.

METHODS

Experimental materials and methods

The three experimental systems were prepared as follows. $SrTiO_3$ films were deposited onto Ge substrates using molecular beam epitaxy (MBE), as described elsewhere[44]. $La_{0.8}Sr_{0.2}FeO_3$ films were deposited onto $SrTiO_3$ (001) substrates using MBE, according to a procedure described elsewhere[45]. Cross-sectional STEM samples of the thin films were prepared using a FEI Helios NanoLab DualBeam Focused Ion Beam (FIB) microscope and a standard lift out procedure. Bulk MoO_3 particles were drop cast onto a lacey carbon grid from suspension in ethanol. High-angle annular dark field (STEM-HAADF) images of the STO/Ge were collected on a probe-corrected JEOL ARM-200CF microscope operating at 200 kV, with a convergence semi-angle of 20.6 mrad and a collection angle of 90–370 mrad. STEM-HAADF images of the LSFO and MoO_3 were collected on a probe-corrected JEOL GrandARM-300F microscope operating at 300 kV, with a convergence semi-angle of 29.7 mrad and a collection angle of 75–515 mrad. The original image data analyzed in this work varied between 3042×3044 pixels (for STO/Ge), 2048×2048 (for LSFO), and 512×512 for MoO_3. Because of its beam sensitivity, the STO/Ge images shown were collected using a frame-averaging approach; a series of 10 frames were acquired with a 1024×1024 px sampling and 2 µs px^{-1}, then

non-rigid aligned and upsampled 2 × using the SmartAlign plugin[72]. Tens of images were collected from each material system and a range of selected defect features were used in this study.

Computational methods

The specific implementation of the preprocessing techniques and parameters for the few-shot model are described in Tables 1 and 2, respectively. All methods were implemented using the Python programming language v.3.6 available at http://www.python.org. Each image was processed using a 16 GB RAM 2.7 GHz Intel Core i7 MacBook Pro.

DATA AVAILABILITY

The raw images, support sets, and query sets for each of our analyzed images, as well as additional classification of MoO_3 nanoparticles, are available on FigShare[73].

CODE AVAILABILITY

The exact codebase for this work is unavailable to the public due to proprietary reasons. However, the Prototypical Network code[57] is available on Github (https://github.com/jakesnell/prototypical-networks).

REFERENCES

1. MacLaren, I. & Ramasse, Q. M. Aberration-corrected scanning transmission electron microscopy for atomic-resolution studies of functional oxides. *Int. Mater. Rev.* **59**, 115–131 (2014).
2. Pennycook, S., Varela, M., Hetherington, C. & Kirkland, A. Materials advances through aberration-corrected electron microscopy. *MRS Bull.* **31**, 36–43 (2006).
3. Varela, M. et al. Materials characterization in the aberration-corrected scanning transmission electron microscope. *Annu. Rev. Mater. Res.* **35**, 539–569 (2005).
4. Pennycook, S. J. The impact of STEM aberration correction on materials science. *Ultramicroscopy* **180**, 22–33 (2017).
5. Oxley, M. P., Lupini, A. R. & Pennycook, S. J. Ultra-high resolution electron microscopy. *Rep. Prog. Phys.* **80**, 026101 (2017).
6. Aguiar, J. A., Gong, M. L., Unocic, R. R., Tasdizen, T. & Miller, B. D. Decoding crystallography from high-resolution electron imaging and diffraction datasets with deep learning. *Sci. Adv.* **5**, eaaw1949 (2019).
7. Voyles, P. M. Informatics and data science in materials microscopy. *Curr. Opin. Solid State Mater. Sci.* **21**, 141–158 (2017).
8. Spurgeon, S. R. et al. Towards data-driven next-generation transmission electron microscopy. *Nat. Mater.* **20**, 274–279 (2016).
9. Baker, M. 1,500 scientists lift the lid on reproducibility. *Nature* **533**, 452–454 (2016).
10. Gunkel, F., Christensen, D. V., Chen, Y. Z. & Pryds, N. Oxygen vacancies: The (in) visible friend of oxide electronics. *Appl. Phys. Lett.* **116**, 120505 (2020).
11. Huang, Z. et al. Interface engineering and emergent phenomena in oxide heterostructures. *Adv. Mater.* **30**, 1802439 (2018).
12. Tuller, H. L. & Bishop, S. R. Point defects in oxides: tailoring materials through defect engineering. *Annu. Rev. Mater. Res.* **41**, 369–398 (2011).
13. Belianinov, A. et al. Big data and deep data in scanning and electron microscopies: deriving functionality from multidimensional data sets. *Adv. Struct. Chem. Imaging* **1**, 6 (2015).
14. Vlcek, L., Maksov, A., Pan, M., Vasudevan, R. K. & Kalinin, S. V. Knowledge extraction from atomically resolved images. *ACS Nano* **11**, 10313–10320 (2017).
15. Horwath, J. P., Zakharov, D. N., Mégret, R. & Stach, E. A. Understanding important features of deep learning models for segmentation of high-resolution transmission electron microscopy images. *NPJ Comput. Mater.* **6**, 108 (2020).
16. Ovchinnikov, O. S. et al. Detection of defects in atomic-resolution images of materials using cycle analysis. *Adv. Struct. Chem. Imaging* **6**, 3 (2020).
17. Maksov, A. et al. Deep learning analysis of defect and phase evolution during electron beam-induced transformations in WS2. *NPJ Comput. Mater.* **5**, 12 (2019).
18. Vasudevan, R. K., Ziatdinov, M., Jesse, S. & Kalinin, S. V. Phases and interfaces from real space atomically resolved data: physics-based deep data image analysis. *Nano Lett.* **16**, 5574–5581 (2016).
19. Kautz, E. et al. An image-driven machine learning approach to kinetic modeling of a discontinuous precipitation reaction. *Mater. Charact.* **166**, 110379 (2020).
20. Otsu, N. A threshold selection method from gray-level histograms. *IEEE Trans. Syst. Man Cybern. Syst.* **9**, 62–66 (1979).
21. Digabel, H. & Lantuéjoul, C. Iterative algorithms. In *Proc. 2nd European Symp. Quantitative Analysis of Microstructures in Material Science, Biology and Medicine*, Vol. **19**, 8 (Riederer Verlag, 1978).
22. MacQueen, J. et al. Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, Vol. **1**, 281–297 (Oakland, CA, USA, 1967).
23. Ede, J. M. Deep learning in electron microscopy. *Mach. Learn.: Sci. Technol.* **2**, 011004 (2021).
24. Kalidindi, S. R. & De Graef, M. Materials data science: current status and future outlook. *Annu. Rev. Mater. Res.* **45**, 171–193 (2015).
25. DeCost, B. L., Francis, T. & Holm, E. A. High throughput quantitative metallography for complex microstructures using deep learning: a case study in ultrahigh carbon steel. *Microsc. Microanal.* **25**, 21–29 (2019).
26. Chowdhury, A., Kautz, E., Yener, B. & Lewis, D. Image driven machine learning methods for microstructure recognition. *Comput. Mater. Sci.* **123**, 176 - 187 (2016).
27. Azimi, S. M., Britz, D., Engstler, M., Fritz, M. & Mücklich, F. Advanced steel microstructural classification by deep learning methods. *Sci. Rep.* **8**, 2128 (2018).
28. DeCost, B. L. & Holm, E. A. A computer vision approach for automated analysis and classification of microstructural image data. *Comput. Mater. Sci.* **110**, 126–133 (2015).
29. Baskaran, A., Kane, G., Biggs, K., Hull, R. & Lewis, D. Adaptive characterization of microstructure dataset using a two stage machine learning approach. *Comput. Mater. Sci.* **177**, 109593 (2020).
30. Ziatdinov, M. et al. Deep learning of atomically resolved scanning transmission electron microscopy images: chemical identification and tracking local transformations. *ACS Nano* **11**, 12742–12752 (2017).
31. Roberts, G. et al. Deep learning for semantic segmentation of defects in advanced stem images of steels. *Sci. Rep.* **9**, 12744 (2019).
32. Chen, D., Guo, D., Liu, S. & Liu, F. Microstructure instance segmentation from aluminum alloy metallographic image using different loss functions. *Symmetry (Basel)* **12**, 639 (2020).
33. Felzenszwalb, P. F. & Huttenlocher, D. P. Pictorial structures for object recognition. *Int. J. Comput. Vis.* **61**, 55–79 (2005).
34. Krizhevsky, A., Sutskever, I. & Hinton, G. E. In *Advances in Neural Information Processing Systems* Pereira, (eds Pereira, F., Burges, C. J. C., Bottou, L. & Weinberger, K. Q.) *Advances in Neural Information Processing Systems*, vol. 25, 1097–1105 (Curran Associates, Inc., 2012).
35. Deng, J. et al. Imagenet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition* (2009).
36. Kaufmann, K., Lane, H., Liu, X. & Vecchio, K. S. Efficient few-shot machine learning for classification of ebsd patterns. *Sci. Rep.* **11**, 1–12 (2021).
37. Pelt, D. M. & Sethian, J. A. A mixed-scale dense convolutional neural network for image analysis. *Proc. Natl Acad. Sci. USA* **115**, 254–259 (2018).
38. Larmuseau, M. et al. Race against the Machine: can deep learning recognize microstructures as well as the trained human eye? *Scr. Mater.* **193**, 33–37 (2021).
39. Díez-Pastor, J.-F., Latorre-Carmona, P., Arnaiz-González, Á., Ruiz-Pérez, J. & Zurro, D. "You are not my type": an evaluation of classification methods for automatic phytolith identification. *Microsc. Microanal.* **26**, 1158–1167 (2020).
40. Bloom, P. *How Children Learn the Meanings of Words* (MIT press, 2002).
41. Altae-Tran, H., Ramsundar, B., Pappu, A. S. & Pande, V. Low data drug discovery with one-shot learning. *ACS Cent. Sci.* **3**, 283–293 (2017).
42. Rutter, E. M., Lagergren, J. H. & Flores, K. B. A convolutional neural network method for boundary optimization enables few-shot learning for biomedical image segmentation. In *Domain Adaptation and Representation Transfer and Medical Image Learning with Less Labels and Imperfect Data. DART 2019, MIL3ID 2019. Lecture Notes in Computer Science, Vol 11795* (2019).
43. Kautz, E. J., Hagen, A. R., Johns, J. M. & Burkes, D. E. A machine learning approach to thermal conductivity modeling: a case study on irradiated uranium-molybdenum nuclear fuels. *Comput. Mater. Sci.* **161**, 107–118 (2019).
44. Du, Y. et al. Layer-resolved band bending at the n-SrTiO3(001)/p-Ge(001) interface. *Phys. Rev. Mater.* **2**, 094602 (2018).
45. Wang, L. et al. Hole-induced electronic and optical transitions in La1-xSrxFeO3 epitaxial thin films. *Phys. Rev. Mater.* **3**, 025401 (2019).
46. Ma, W. et al. Image-driven discriminative and generative machine learning algorithms for establishing microstructure–processing relationships. *J. Appl. Phys.* **128**, 134901 (2020).
47. Ustyuzhaninov, I., Michaelis, C., Brendel, W. & Bethge, M. One-shot texture segmentation. Preprint at https://arxiv.org/abs/1807.02654 (2018).
48. Pizer, S. M. et al. Adaptive histogram equalization and its variations. *Computer Vis. Graph. Image Process.* **39**, 355–368 (1987).
49. Zuiderveld, K. Contrast limited adaptive histogram equalization. *Graphics Gems* (Academic Press, 1994) pp 474–485.
50. Marsh, B. P., Chada, N., Gari, R. R. S., Sigdel, K. P. & King, G. M. The hessian blob algorithm: Precise particle detection in atomic force microscopy imagery. *Sci. Rep.* **8**, 1–12 (2018).

51. Van der Walt, S. et al. scikit-image: image processing in python. *PeerJ* **2**, e453 (2014).
52. Vinyals, O., Blundell, C., Lillicrap, T., Kavukcuoglu, K. & Wierstra, D. Matching networks for one shot learning. *Advances in Neural Information Processing Systems* (Curran Associates, Inc., 2016).
53. Rahman, S., Khan, S. & Porikli, F. A unified approach for conventional zero-shot, generalized zero-shot, and few-shot learning. *IEEE Trans. Image Process.* **27**, 5652–5667 (2018).
54. Edwards, H. & Storkey, A. Towards a neural statistician. *In 5th International Conference on Learning Representations (ICLR 2017)* (2017) 1–13.
55. Ren, M. et al. Meta-learning for semi-supervised few-shot classification. *In 5th International Conference on Learning Representations (ICLR 2018)* (2018). 1–15.
56. Hilliard, N. et al. Few-shot learning with metric-agnostic conditional embeddings. Preprint at https://arxiv.org/abs/1802.04376 (2018).
57. Snell, J., Swersky, K. & Zemel, R. Prototypical networks for few-shot learning. In *Advances in Neural Information Processing Systems (Curran Associates, Inc.)*, 4077–4087 (2017).
58. He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, IEEE, 770–778 (2016).
59. Ronneberger, O., Fischer, P. & Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, 234–241 (Springer, 2015).
60. Paszke, A. et al. Pytorch: An imperative style, high-performance deep learning library. (eds Wallach, H. et al.) *Advances in Neural Information Processing Systems 32*, 8024–8035 (Curran Associates, Inc., 2019).
61. Deng, J. et al. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, 248–255 (IEEE, 2009).
62. Horwath, J. P., Zakharov, D. N., Megret, R. & Stach, E. A. Understanding important features of deep learning models for segmentation of high-resolution transmission electron microscopy images. *NPJ Comput. Mater.* **6**, 1–9 (2020).
63. Achanta, R. et al. Slic superpixels compared to state-of-the-art superpixel methods. *IEEE Trans. Pattern Anal. Mach. Intell.* **34**, 2274–2282 (2012).
64. Hughes, A., Liu, Z., Raftari, M. & Reeves, M. E. A workflow for characterizing nanoparticle monolayers for biosensors: machine learning on real and artificial SEM images. Preprint at https://peerj.com/preprints/671/ (2014).
65. Gonzalez, R.C. & Woods, R.E. (2002) Digital Image Processing. 2nd Edition, Prentice Hall, Upper Saddle River.
66. Buades, A., Coll, B. & Morel, J. M. On image denoising methods. Preprint available at https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.100.81&rep=rep1&type=pdf (2004).
67. Koch, G., Zemel, R. & Salakhutdinov, R. Siamese neural networks for one-shot image recognition. In *ICML Deep Learning Workshop*, vol. 2 (Lille, 2015).
68. Sung, F. et al. Learning to compare: Relation network for few-shot learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, 1199–1208 (2018).
69. Oxley, M. P. et al. Probing atomic-scale symmetry breaking by rotationally invariant machine learning of multidimensional electron scattering. *NPJ Comput. Mater.* **7**, 1–6 (2021).
70. Kalinin, S. V., Dyck, O., Jesse, S. & Ziatdinov, M. Exploring order parameters and dynamic processes in disordered systems via variational autoencoders. *Sci. Adv.* **7**, eabd5084 (2021).
71. Wang, Y., Yao, Q., Kwok, J. T. & Ni, L. M. Generalizing from a few examples: a survey on few-shot learning. *ACM Comput. Surv.* **53**, 1–34 (2020).
72. Jones, L. et al. Smart Align-a new tool for robust non-rigid registration of scanning microscope data. *Adv. Struct. Chem. Imaging* **1**, 8 (2015).
73. Akers, S. et al. Supplemental data for rapid and flexible segmentation of electron microscopy data using few-shot machine learning. Data repository available at https://doi.org/10.6084/m9.figshare.14850102.v2 (2021).

AUTHOR CONTRIBUTIONS
S.A. and S.R.S. planned the study. S.A. and A.T.-G. developed the machine learning approach. S.R.S, M.O., and B.M. collected the microscope data. L.W. and Y.D. synthesized the thin film crystal samples. All authors contributed to the data interpretation and writing of the manuscript.

COMPETING INTERESTS
The authors declare no competing interests.

ADDITIONAL INFORMATION
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41524-021-00652-z.

Correspondence and requests for materials should be addressed to Steven R. Spurgeon.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.