Journal of Nuclear Materials 588 (2024) 154813

Contents lists available at ScienceDirect

Journal of Nuclear Materials

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

Automated analysis of grain morphology in TEM images using convolutional neural network with CHAC algorithm

Xinyuan Xu [a], Zefeng Yu [a], Wei-Ying Chen [b], Aiping Chen [c], Arthur Motta [a], Xing Wang [a,*]

[a] Ken and Mary Alice Lindquist Department of Nuclear Engineering, The Pennsylvania State University, University Park, PA 16802, USA
[b] Nuclear Science and Engineering Division, Argonne National Laboratory, Argonne, IL 60439, USA
[c] Center for Integrated Nanotechnologies (CINT), Los Alamos National Laboratory, Los Alamos, NM 87545, USA

ARTICLE INFO

Keywords:
Grain growth
UO_2
Transmission electron microscopy (TEM)
Convolutional neural network

ABSTRACT

The shape and size of grains significantly impact the properties of polycrystalline materials. In particular, high temperature and radiation exposure in nuclear reactors can lead to considerable grain growth of UO_2, thereby substantially modifying the fuel performance. Transmission electron microscopy (TEM) has been used to characterize the grain morphology of polycrystalline materials, but the manual analysis of TEM images is a time-consuming and labor-intensive process, which cannot meet the increasing demand for high-throughput data analytics.

This study presents an automated approach we developed for characterizing grain morphology recorded in bright field TEM images during ion irradiations performed *in situ*. Our approach combines a machine learning model for detecting grain boundaries and a computer vision algorithm named CHAC for selecting well-labeled grains for statistical analysis. Using TEM images acquired from *in-situ* ion irradiation experiments on nanocrystalline UO_2, we demonstrate that this automated approach can achieve comparable results to human experts while significantly reducing the analysis time. Moreover, the machine learning model functions as a "few-shot" model, requiring only a modest number of training images to perform effectively on a specific task. Consequently, researchers in need can efficiently train their own models following the procedures described in this study to automate grain morphology analysis of their own TEM images.

1. Introduction

It is widely acknowledged that properties of polycrystalline materials can be tailored by modifying the grain morphology [1]. Grain size engineering, which involves adjusting grain size without altering the chemical composition, has become an important approach for enhancing the mechanical, electrical, and optical properties of materials [2]. For example, the well-known Hall-Petch equation shows that yield strength increases with decreasing grain size due to the impeding effect of grain boundaries on dislocation movement [3]. Studies demonstrated that the charge carrier mobility would increase as the grain size of nanocrystalline indium tin oxide increases, since the number of trapping sites at grain boundaries is reduced during grain growth [4]. Research on nanocrystalline titanium oxide found that modifying grain sizes can change both the band gap and specific reaction area of this semiconductor as a photocatalyst [5]. All these studies highlight the importance for efficient characterization of grain morphology across various materials.

Grain size can significantly impact the performance of uranium dioxide (UO_2), which is the most-widely used nuclear fuel for the current fission reactor fleet [6,7] and a promising candidate fuel for some Generation IV reactors [8]. Due to the elevated operating temperature and intensive radiation, UO_2 fuels go through significant grain growth during reactor operation, which can lead to substantial fuel performance change, affecting thermal conductivity, fission gas release, swelling, and creep [9,10]. Previous studies have applied *in-situ* ion irradiation in transmission electron microscopy (TEM) to investigate grain growth in UO_2 and other materials, as this *in-situ* technique provides a unique opportunity to directly observe the kinetics of grain growth under well-controlled irradiation conditions [11–13]. Meanwhile, *in-situ* TEM experiments often result in a large number of images and even videos, making the data analysis time-consuming and potentially prone to

* Corresponding author.
E-mail address: xvw5285@psu.edu (X. Wang).

https://doi.org/10.1016/j.jnucmat.2023.154813
Received 2 July 2023; Received in revised form 1 November 2023; Accepted 3 November 2023
Available online 4 November 2023

human error if the images and videos are analyzed manually.

Applying machine learning in material science has significantly advanced the field as processing large datasets has become a common task for various modern material science research. From a material discovery point of view, machine learning has led to a new field of material informatics and resulted in the discovery of many new materials [14]. From a material data analysis point of view, machine learning in many cases is irreplaceable due to the large number of datasets produced in different imaging experiments using electron, neutron and X-ray [15,16]. Furthermore, the integration of computer vision, machine learning, and deep learning techniques for microstructural characterization and analysis holds considerable significance [17,18]. Convolutional Neural Networks (CNNs) for machine learning have recently gained much attention from the microscopy community because of their excellent performance in image analysis tasks, such as face recognition and image classification [19]. CNNs are a type of neural network that utilizes shared-weight convolutional kernels scanning across the input image to efficiently produce feature maps with translation-invariant responses. Researchers have successfully applied CNNs to analyze various microstructural features. For instance, Roberts et al. utilized a CNN variant called DefectSegNet to identify precipitates and voids [20], while Li et al. applied a classical CNN to identify <100> dislocation loops in scanning transmission electron microscopy (STEM) images of irradiated FeCrAl alloys [21]. More recently, an effective model was developed by incorporating a mature object-detection framework, Faster R-CNN, allowing for detecting different types of dislocation loops and black spots in STEM images [22]. Additionally, several different frameworks, including Mask R-CNN and Faster R-CNN, have been successfully applied to recognize voids and bubbles in TEM images [23–25].

A number of studies have focused on automating the identification of grain boundaries in micrographs, a task that essentially involves detecting object edges. The Canny algorithm is a classic edge detection method that recognizes pixels exhibiting sufficiently strong intensity gradients as edges. However, directly applying the Canny algorithm to micrographs of grain boundaries has yielded limited performance [26,27]. Campbell et al. proposed a set of processing techniques, including filtering, watershed transform, region merging, and phase separation, to automatically draw contours around grains in scanning electron microscopy (SEM) and optical microscopy (OM) images [28]. In addition to the conventional computer vision techniques, machine learning has also been explored to analyze grain morphology. Using the CNN-based holistically nested edge detection algorithm and post-processing, Bordas et al. quantified the grain size distribution in SEM images [29]. Their results demonstrated superior recognition of grain boundaries compared to the Canny algorithm. Similarly, Ma et al. trained a CNN-based framework named DeepLab to separate grains in microscopic images of lanthanum aluminum alloys, which substantially outperformed the conventional image segmentation methods, such as watershed, K-Means, and Graph-Cut [30]. Despite these advancements, few studies have been dedicated to analyzing grain morphology in TEM images. Compared to SEM and OM, TEM has more complex imaging mechanisms, resulting in more dynamic changes in image contrast depending on the crystallographic orientation, thickness, and composition of each grain. Therefore, it is more challenging to identify grain boundaries in TEM images than in SEM or OM. Nevertheless, because of its high spatial resolution (< 1nm), TEM has become a critical tool in studying nanomaterials. So far, most studies still rely on human experts to manually analyze TEM images. Automated TEM analysis of grain morphology can greatly expedite research by reducing manual labor, increasing consistency, and potentially revealing subtle patterns that might be missed by human eyes.

In this work, we developed a machine learning model that can automatically detect grain boundaries of nanocrystalline UO_2 in Bright Field (BF)-TEM images. The model utilized a unique CNN framework known as UNet [31]. Our results showed that the UNet model could achieve a good performance with a modest number of manually labeled TEM images as the training dataset, demonstrating the model's potential of being quickly trained for analyzing different nanocrystalline materials. It is important to note that achieving 100% accuracy with a machine learning model remains elusive; consequently, the UNet model mislabeled some grain boundaries. To obtain a reliable statistical analysis of grain size distribution, we designed a post-processing method that combined the Convex-Hull (CH) and Approximate Contour (AC) algorithms for selecting well-labeled grain contours by the machine learning model. We termed this method CHAC. By combining UNet with CHAC, we were able to automate the analysis of TEM images from *in-situ* irradiation of nanocrystalline UO_2. The automated approach revealed the grain growth kinetics that matched very well with results from manual analyses, and significantly accelerated the data analysis process. In the end, limitations and future improvements of this automated method for grain morphology analysis are discussed.

2. Methods

2.1. Fabrication of nanocrystalline UO_2 and in-situ irradiation

Nanocrystalline UO_2 thin films of electron transparent thickness were grown on an electron transparent silicon nitride membrane (from Norcada) using pulsed laser deposition (PLD) at the Center for Integrated Nanotechnologies, Los Alamos National Laboratory. The samples were prepared so that the grains were quite small to enhance the driving forces for grain growth and so that their size would be much smaller than the TEM sample thickness. The samples had an initial grain diameter of 10-15 nm. *In-situ* Kr ion irradiation was then performed to monitor the radiation-induced grain growth of UO_2. The thin films of nanocrystalline UO_2 were irradiated by 1 MeV Kr ions at a flux of 6.25×10^{15} $ions/m^2/s$ to a maximum fluence of 7.1×10^{19} $ions/m^2$ using the Intermediate Voltage Electron Microscopy (IVEM) facility at Argonne National Laboratory, where a linear ion accelerator with a 911 Danfysik ion source is coupled with a Hitachi 9000 TEM at 30° for simultaneous ion irradiation and electron imaging. The Stopping and Ranging of Ions in Matter (SRIM) was utilized to simulate the fraction of ions traversing the 50 nm thick film and implanting within it. The simulation, conducted in full cascade mode, involved 175,000 ions. At this energy level, the majority of Kr ions passed through the thin foil, with only a minimal fraction (0.48%) being implanted. To investigate the mechanisms for grain growth at different temperatures, the irradiation temperatures varied from 50 K to 1075 K [11]. BF-TEM images were acquired during the entire irradiation process. Fig. 1 shows typical BF-TEM images of nanocrystalline UO_2 during the irradiation at room temperature (300K). It is clear that the average grain size increased significantly due to the irradiation. More than 200 BF TEM images in total were collected. More details related to sample fabrication and *in-situ* irradiation have been provided before [19,32,33].

2.2. Introduction to CNN, UNet, and model training

Similar to conventional artificial neural networks, CNN consists of an input layer, multiple hidden layers, and an output layer. However, the unique convolution layers among the hidden layers render CNN particularly adept at handling sizable inputs, such as high-resolution images [34]. In conventional artificial neural networks, matrix multiplications are used that would result in an exponential increase in time, computational resources, and learning complexity with the size of the input dataset, making it challenging to develop an effective neural network for high-resolution images. In contrast, the convolution layers in CNN use small kernels (e.g., 3 × 3 matrices) for the convolution, leaving the data size nearly unchanged between different layers. For image analysis, the convolutional layers can extract various features from the input image and generate the feature maps, which are three-dimensional tensors, and the depth of the tensor is often referred

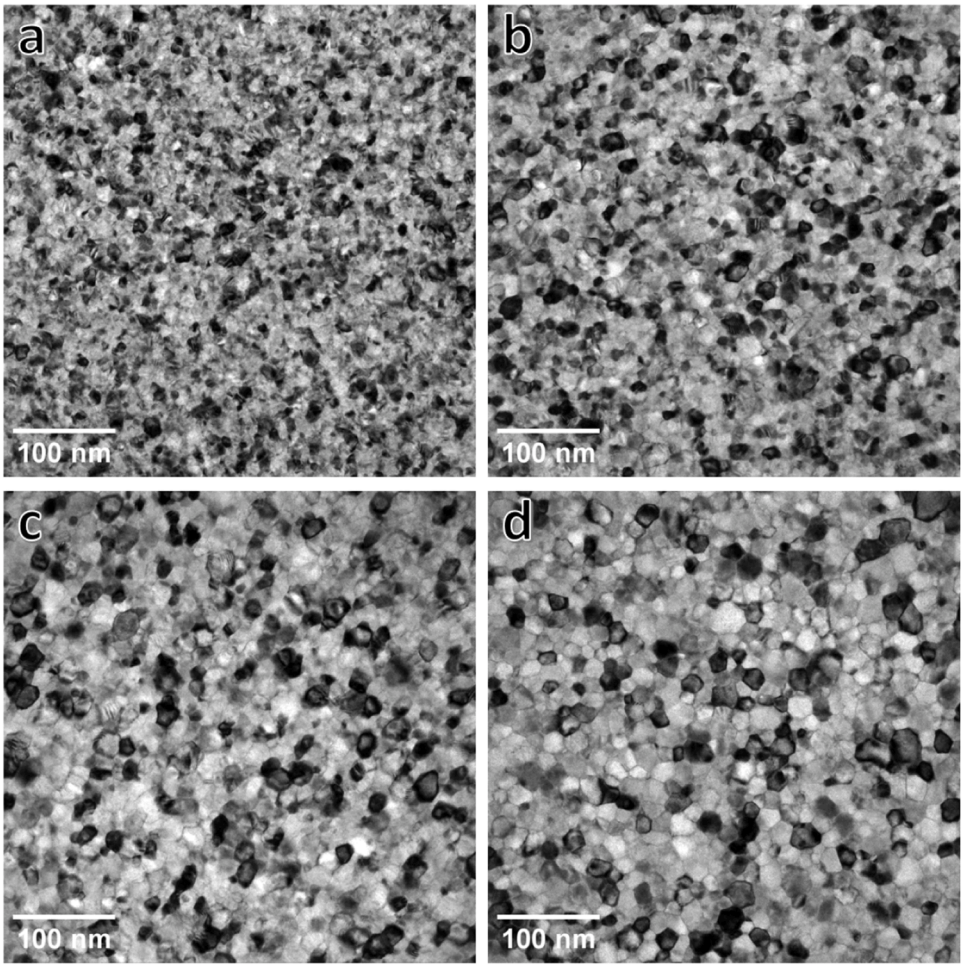

Fig. 1. Examples of BF-TEM images of nanocrystalline UO_2 irradiated to the different fluence of (a) 0 ions/m^2, (b)5.04 $\times 10^{18}$ ions/m^2, (c)1.51 $\times 10^{19}$ ions /m^2 and (d) 3.02 $\times 10^{19}$ ions/m^2 at room temperature.

to as channels of the feature map. The first convolutional layer may only identify basic features, such as edges and corners, but with successive convolutional layers, increasingly complex features like shapes can be extracted. Besides convolutional layers, the hidden layers in CNN also include non-linearity layers, pooling layers, and a fully-connected layer. The non-linearity layers, which often employ the Rectified Linear Unit (ReLU) [35] activation function, introduce nonlinearity to the neurons and enable the network to model non-linear relationships between

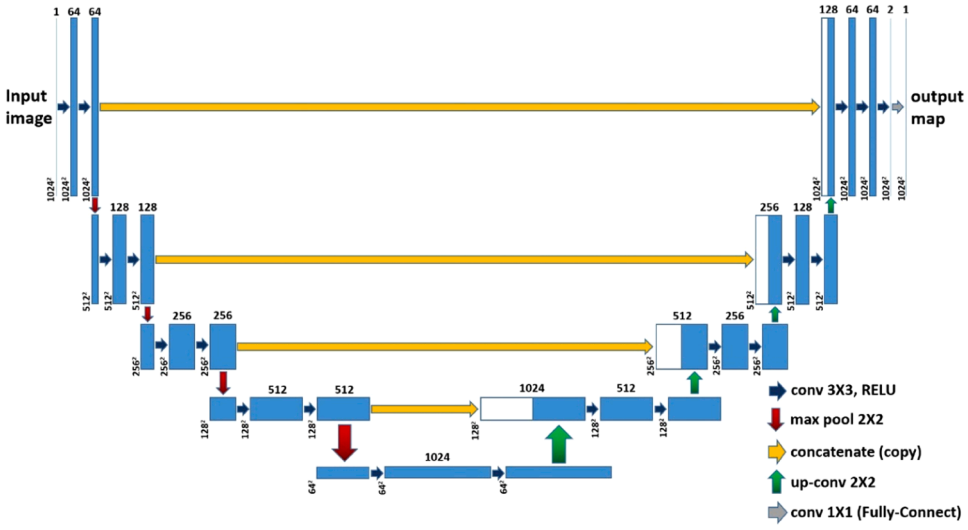

Fig. 2. Structure of UNet model in this work. Each blue or white-blue box represents a feature map. The number of channels in each feature map is shown on the top and the map dimensions are shown in the lower-left corner.

extracted features and the output. The pooling layer, typically located after the convolution layer, condenses the feature maps by summarizing the information within a smaller region (e.g., 2 × 2 window) to reduce the parameters and computational costs. Lastly, the fully-connected layer integrates all the local features extracted by the preceding layers into global features and mapping them to the appropriate output space.

UNet is a CNN with an innovative U-shaped architecture, making it highly effective for tasks such as image segmentation and particle analysis [31]. As illustrated in Fig. 2, the model includes a contracting path that captures the content of images, such as colors, objects, and scenes, and a symmetrical expansive path that transfers this contextual information to higher-resolution layers. As indicated by the yellow arrows in Fig. 2, the concatenate operation combines the feature maps from both the contracting path and the expansive path. This operation integrates high-level contextual information with low-level spatial details, enhancing the capability of UNet model to accurately recognize object boundaries. UNet has been widely used in segmentation of biomedical and microscopic images [31,36,37]. In particular, we have developed a customized UNet model to automatically recognize and measure UO_2 grains in dark field TEM images [11].

In this work, we used ten 1024 × 1024 pixels BF-TEM images as the training dataset and another two BF-TEM images as the validation dataset for tuning hyperparameters of the UNet model. The performance of the trained model was evaluated on a separate test dataset containing four representative BF-TEM images. The training images were randomly selected from augmented through rotation, mirroring and zoom-in by eight times. In all the BF-TEM images, grain boundaries were annotated by human experts. It is worth noting that human experts often labeled the grain boundaries as one pixel-wide lines, the width of which was smaller than the size of the convolution kernels (3 × 3-pixel matrices), causing difficulties for the kernels to extract the labeled grain boundaries. To address this issue, we first increased the width of labeled boundaries to five pixels for the model training and segmentation. After that, the grain boundary width was reduced to one pixel by keeping only the central pixels. The next step involves grain size measurement, we employed a well-established method called Equivalent Circle Diameter to measure grain diameters in this work [38,39]. Specifically, we used the diameter of an equivalent circle which has the same area as the grain to represent the grain diameter. The UNet model was trained on the Tensorflow platform using a single Tesla P-100 GPU. Binary cross-entropy loss function, which is commonly used by models for grayscale image segmentation, was chosen for training the model [40]. The training process was terminated when the decrease of validation loss plateaued, at which point the best model from the entire training phase was automatically saved. The training process lasted approximately five hours. The network structure was optimized by adjusting the number and size of convolutional layers along with filters, the number of maxpooling layers, and the size of fully connected layers. Fig. 2 illustrates the finely tuned UNet, which consists of one input layer, one output layer, four max-pooling layers, four concatenation layers, and 23 convolutional layers.

2.3. Convex grain assumption and CHAC method

In the classic theory of grain growth, grain boundary movement is driven by the difference in curvature between adjacent grains. The curvature difference creates a pressure that results in unbalanced atom fluxes jumping across the boundary, causing the boundary to migrate towards the center of the boundary curvature, minimizing the boundary area, and thus reducing the total free energy [41]. As a result of grain boundary migration, it is expected that all grains will adopt a convex shape after adequate growth, leading to an equilibrium state where the curvature difference between touching grains becomes negligible. This process indicates that only convex grains should be considered well-grown for counting and measurement.

Although the UNet model can reach a reasonably good performance for grain boundary identification, its accuracy is lower than 100%, as some grain boundaries are inevitably missed in the UNet-labelled results. As illustrated in Fig. 3, the missing grain boundaries can cause unclosed grain contours and even the merging of two adjacent grains, resulting in concave contours. To get more accurate grain size measurements in BF-TEM images, it is necessary to develop an automated method that eliminates concave contours while selecting an adequate number of convex contours. The CHAC method was developed to achieve this goal and is introduced in the following paragraphs.

A number of computer vision algorithms exist for constructing convex hulls of objects in images. For example, Sklansky's algorithm identifies the minimum convex polygon of a set of pixel points and returns its contours [42]. However, it is challenging to directly apply these convex hull algorithms to the grain boundary maps. Because of many concave contours as shown in Fig. 3, a large fraction of convex hulls constructed by Sklansky's algorithm would differ significantly from the actual shape of grains, leading to obvious errors in grain size measurement. Therefore, it is necessary to pick up the convex contours first. Nevertheless, this is not an easy task because many grain boundaries are labeled by human experts and the UNet model contains minor bends as illustrated in Fig. 4a–4d, which introduce a slight concavity to the convex contours. If these minor bends are not straightened, only a small fraction of convex contours with perfectly straight boundaries can be selected. For instance, among all the grains in Fig. 4a, only the small grain marked by the purple box was selected by the algorithm as shown in Fig. 4e, 4g.

To address the above issue, we adopted the Douglas-Peucker algorithm to straighten the grain boundaries labeled by the UNet model [43,44]. This algorithm generates an approximated polygon that preserves the rough shape of the original polygon but contains fewer vertices. The degree of approximation is controlled by a single parameter, ε, which defines the maximum allowable distance between the original vertices of the polygon edge and the simplified edge. For example, setting ε to 0.02 means that the maximum distance should be less than two percent of the circumference of the original polygons. In general, as ε increases, the similarity between the original polygon and the approximated polygon decreases but it is more likely to convert the polygon into a convex contour. Therefore, there is a trade-off between the accuracy of the polygon size and the number of convex contours that can be obtained. In this work, we applied $\varepsilon = 0.02$ to all images to obtain the approximated contours.

The CHAC method is illustrated in Fig. 5. The computation process starts with the Approximate Contour module, including contour detection, approximation of the contours into polygons, and selection of convex polygons. This module results in the straightening of the original contours and the generation of a sufficient number of perfect convex contours for statistical analysis. The Convex Hull module applies Sklansky's algorithm to construct polygons and draw them on a white canvas. With these closed and convex grain contours, a watershed algorithm is applied to separate the grains for extracting grain information, such as mean and standard deviation of grain diameters.

3. Results

3.1. Performance of UNet for identifying grain boundaries

The performance of UNet model for automated grain boundary detection is evaluated using four basic metrics, i.e., Accuracy, Precision, Recall, and F1 score. For calculating these metrics, we took the BF-TEM images labeled by human experts as the ground truth and defined the pixels on the grain boundaries as positive and all the other pixels as negative. As shown by the confusion matrix in Table 1 and Eq. (1), Accuracy measures the proportion of correctly predicted pixels among all pixels in the TEM image. Although Accuracy is commonly adopted for evaluating performance of machine learning models, it is not sufficient for our work due to the significant imbalance between the number of

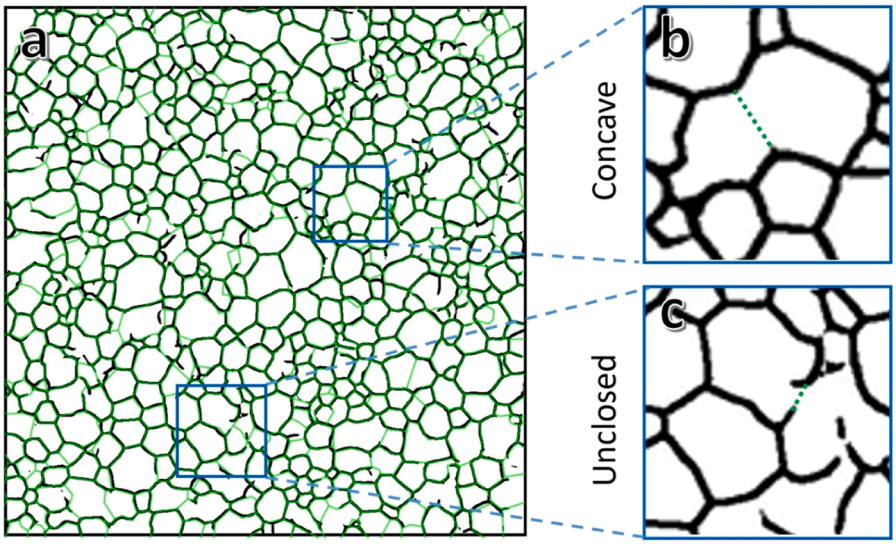

Fig. 3. An example of (a)a UNet-labeled grain map (black lines) overlaid by the "ground truth" grain map (green lines) with (b) two merged grains forming a concave shape and (c) an unclosed contour of grains. The green dashed lines indicate the possible locations of the missing grain boundaries.

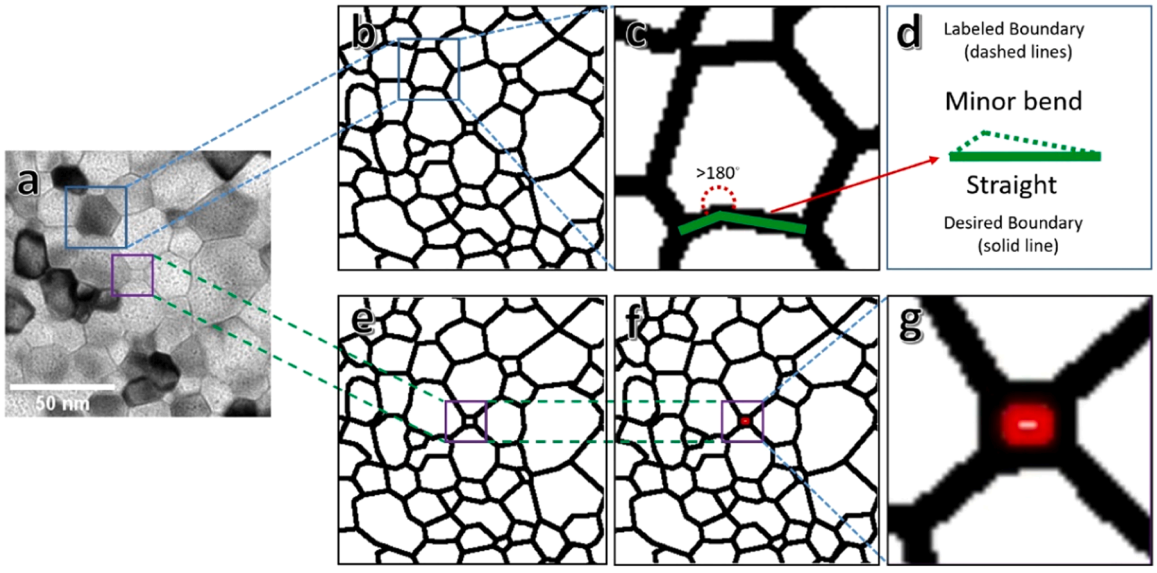

Fig. 4. Issues of directly applying the convex hull algorithm to UNet result (a) An example of BF-TEM image. A grain containing minor bends along the boundary is marked by the blue box; the grain with perfectly straight boundaries is marked by the purple box; (b) The corresponding grain boundary map of labeled by UNet; (c) The minor bend is annotated by the green solid lines, where the angle between two bent boundaries is larger than 180 degrees, making the contour concave; (d) Enlargement of the minor bend (dashed lines) and the desired straight grain boundary (solid line); (e) A perfect convex contour marked by a purple box; (f) Only perfect convex contours can be selected by the convexity algorithm; (g) Enlargement of the selected perfect convex contour.

positive (grain boundary) and negative (non-grain boundary) pixels. As shown in Fig. 6, the pixels on grain boundaries only account for a small fraction (~ 10% to 20%) in the TEM image. Therefore, even a very high Accuracy, e.g., >90%, does not guarantee that enough positive pixels are identified by the UNet model, since most of the correctly detected pixels could be negative. Therefore, the other three metrics are necessary for a comprehensive evaluation. As shown in Eqs. (2) and (3), Precision measures the fraction of positive pixels identified by UNet that are actually on grain boundaries; Recall measures the fraction of pixels on grain boundaries that are correctly captured by UNet. Precision and Recall are often conflicting performance quotas: the higher the Precision is, the lower the Recall tends to be [45]. Therefore, F1 score is employed to evaluate the overall performance. As shown in Eq. (4), either a low Precision (P) or Recall (R) could result in a low F1 score.

$$Accuracy = \frac{TP + TN}{TP + TN + FN + FP} \tag{1}$$

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

$$F1 = \frac{2PR}{P + R} \tag{4}$$

Table 2 summarizes the evaluation metrics of our UNet model using four test images. All four matrices exceed 0.70. Furthermore, the F1 score (0.74) is comparable to other CNN models developed for defecting different defects in BF-TEM images, such as black dots (0.68) [22],

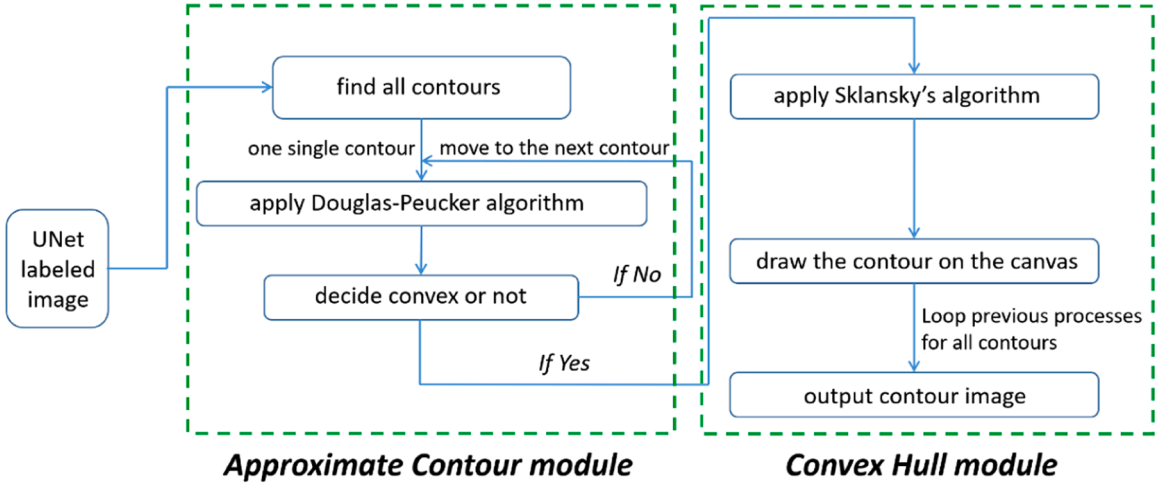

Fig. 5. The step-by-step schematic diagram of the CHAC method.

Table 1
Confusion matrix showing all possible prediction results of UNet model.

Ground Truth	Machine Identification Positive (Boundary)	Negative (Non-boundary)
Positive (Boundary)	True Positive (**TP**)	False Negative (**FN**)
Negative (Non-boundary)	False Positive (**FP**)	True Negative (**TN**)

dislocation loops (0.67-0.78) [22], and cavities (0.7-0.8) [23,24]. Note that the term 'black dots' in the reference includes stacking fault tetrahedra, dislocation loops, and defect clusters, which are not considered a separate class of defects. It is also worth mentioning that the F1 score we assessed is based on pixel-level identification, i.e., grain boundary segmentation, which is slightly different from the object-level identification in the above references. Nevertheless, this comparison still provided a valuable evaluation and showed that our UNet model demonstrated a reasonable performance level. To further support this claim, Fig. 6 compares a test image, the corresponding map of grain boundaries labeled by a human expert, and the UNet results. It is clear that the grain morphology from UNet closely resembles that from human labeling. Note that for general semantic segmentation tasks using CNN models, an F1 score above 0.8 or even 0.9 is considered excellent [46,47]. Compared to that standard, our F1 score is still low. Correspondingly, some grain boundaries are missing in the UNet results, as marked by the red boxes in Fig. 6b and 6c. Nevertheless, we can still accurately measure the grain size distribution since there are a few hundred grains in each TEM image, so the CHAC method can select adequate well-labeled grains. More details are provided in the next Section.

3.2. Performance of CHAC method for measuring grain sizes

As shown in Fig. 7, we combined the UNet model and the CHAC method for measuring grain size distribution. Fig. 7 (a) is a test BF-TEM image and Fig. 7 (b) shows the grain boundaries detected by the trained UNet model, with convex grains identified by the CHAC method marked in red. In Fig. 7 (c), the convex grains found via the UNet+CHAC approach are overlaid on the original TEM image, showing that more than 70% of the total grains in the original TEM image were successfully detected. The situation is similar in all the other three test images. In Fig. 7 (d), we applied the watershed algorithm so each grain can be separated and easily measured.

To quantitatively evaluate the performance of this UNet + CHAC approach, we compared the grain diameter distribution in four test BF-TEM images obtained from UNet results and the ground truth (i.e., the results from human experts). Note both the size distribution from UNet and the ground truth were processed via the CHAC method. As shown in Fig. 8, the histograms of grain diameter from UNet closely resemble that

Table 2
Summary of the model's identification performance evaluation metrics.

Evaluation Metrics	Accuracy	Precision	Recall	F1 Score
UNet	0.89	0.75	0.73	0.74

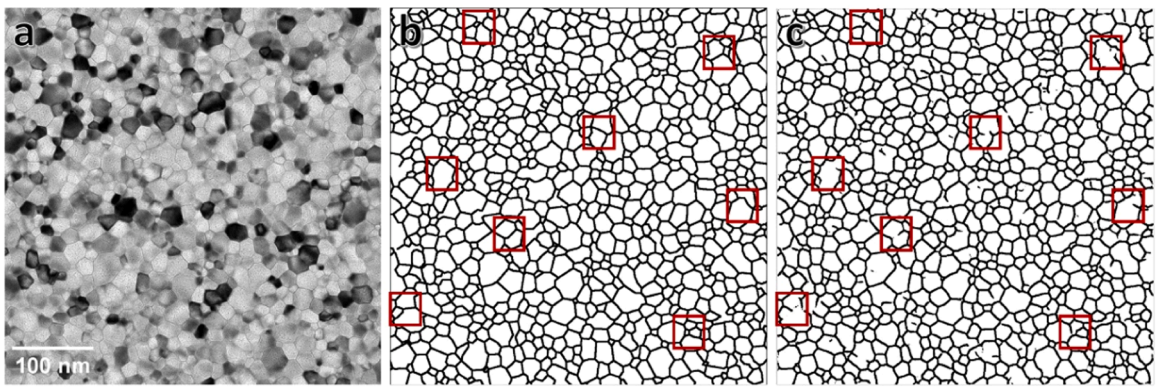

Fig. 6. Examples of identification results of one selected representative test BF image of UNet (a) Origin BF image; (b) Ground Truth; (c) UNet identification result.

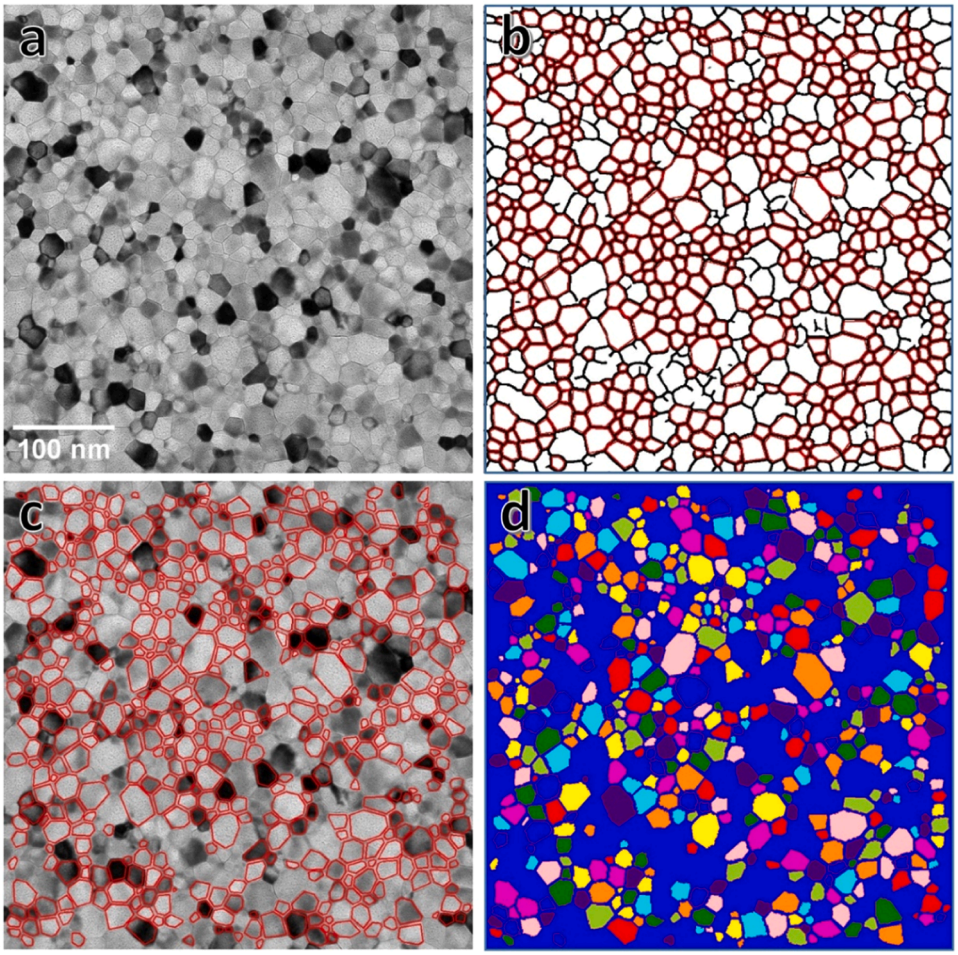

Fig. 7. Examples of convex grains selected by the CHAC method. (a) A test TEM image; (b) The corresponding map of grain boundaries labeled by UNet with the convex contours identified by CHAC colored in red;(c) The convex contours identified by CHAC overlapped on the test image; (d) Convex contours separated by the watershed algorithm. The contours are labeled in different colors for ease of observation.

of the ground truth for each test TEM image. It is worth mentioning that the discrepancy between the UNet results and the ground truth is more pronounced in Image 4, which is likely caused by the relatively larger average grain diameters and more blurred grain boundaries. However, even in this special case, the histograms from UNet and human experts are quite similar.

To further evaluate the effectiveness of the UNet+CHAC approach for automated grain size analysis, we applied this approach to a series of BF-TEM images of nanocrystalline UO_2 under *in-situ* irradiation at 875K to the ion fluence of 6.4×10^{19} $ions/m^2$. For a reliable evaluation, none of these BF-TEM images were used as training or validation data for the UNet model. The trend of grain growth obtained from the UNet+CHAC method was compared to the manual analysis in Fig. 9, which shows that the average grain sizes measured by both methods are very close. Therefore, the new automated analysis approach has achieved a comparable performance to that of human experts. Furthermore, while it took about eight hours for the human expert to acquire all the data points represented in Fig. 9, the implementation of the automatic method dramatically reduced the analysis time to just about ten minutes. This indicates the high efficiency of the automated approach, demonstrating that the UNet+CHAC approach can be applied to efficiently and accurately characterize grain size distribution in BF-TEM images.

3.3. *Effect of dataset size on UNet model performance*

When developing CNN models for semantic segmentation tasks, a frequently raised question is how much training data is necessary in order to achieve an acceptable performance. To answer this question, we trained our UNet model using varying sizes of training datasets and then calculated the model F1 score. The resulting curve is plotted in Fig. 10. Here we used the number of grains identified by human experts in the training images to represent the size of the training dataset, and the training datasets were randomly chosen among the ten 1024×1024 pixels BF-TEM images. The F1 score was calculated using the same four BF-TEM images. As shown in Fig. 10, the model struggled when trained by only a few hundred grains, but the F1 score rapidly increased to 0.7 with about 1000 labeled grains. Finally, the F1 score saturates around 0.75 with ~3000 grains. A comparable trend in the relationship between dataset size and performance was also observed by Jacob et al., who employed a region-based CNN model to detect dislocations in STEM images [47]. Since there are a few hundred grains on average in each of our BF-TEM images, it indicates that the UNet model only needs about a few well-labeled TEM images to achieve a reasonable performance. Therefore, the UNet model offers not only an effective approach to identifying grain boundaries, but also the advantage of being a "few-shot" model, which can be trained to accomplish a specific task using only a few examples [48,49].

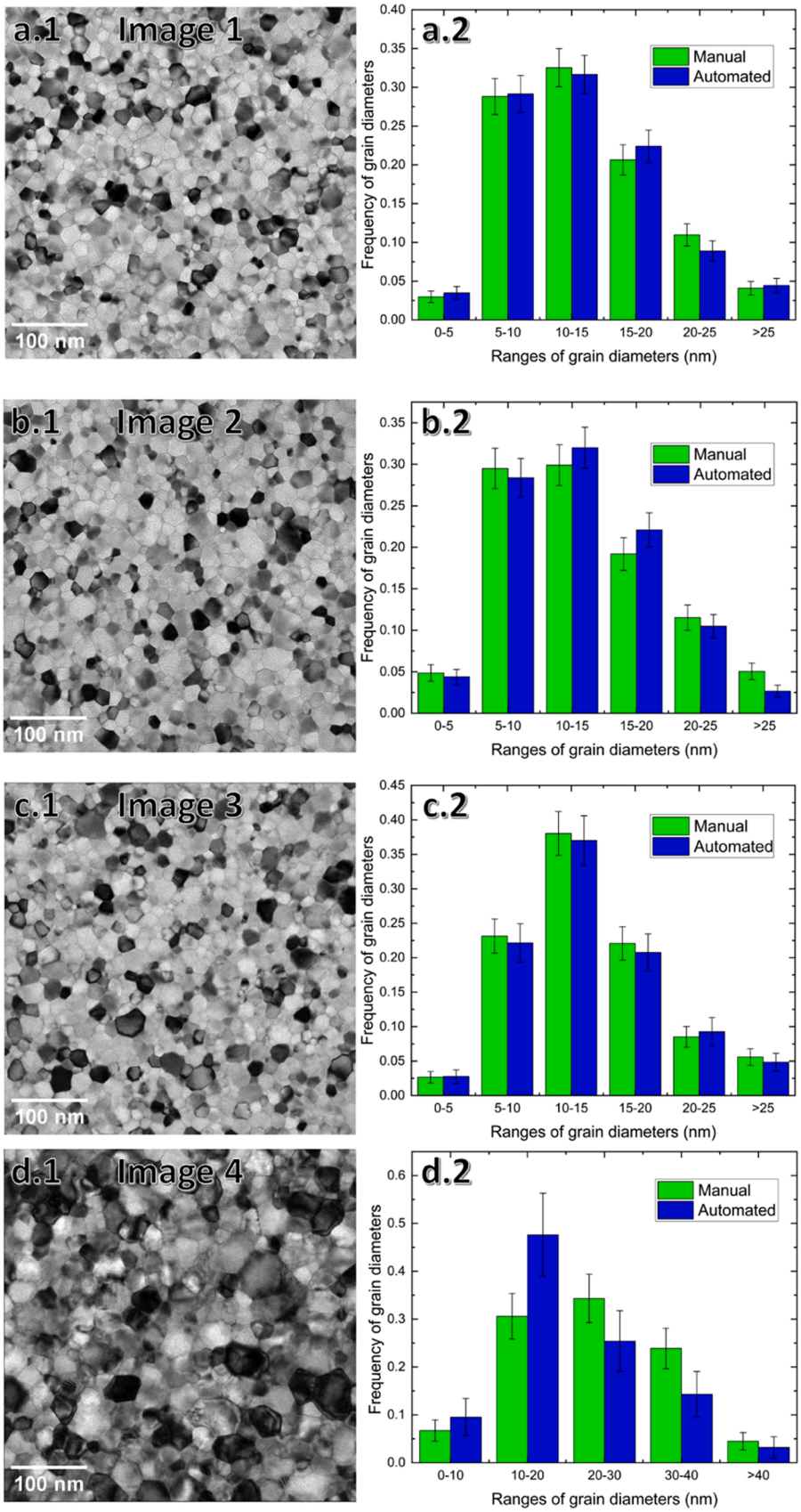

Fig. 8. (a.1) (b.1) (c.1) (d.1) are four test BF-TEM images and (a.2) (b.2) (c.2) (d.2) are the corresponding frequency distribution of grain diameters. Results of manual analysis and automated analysis are plotted in green and blue, respectively. The error bars were computed by assuming the number of grains within a grain diameter range follows the Poisson distribution.

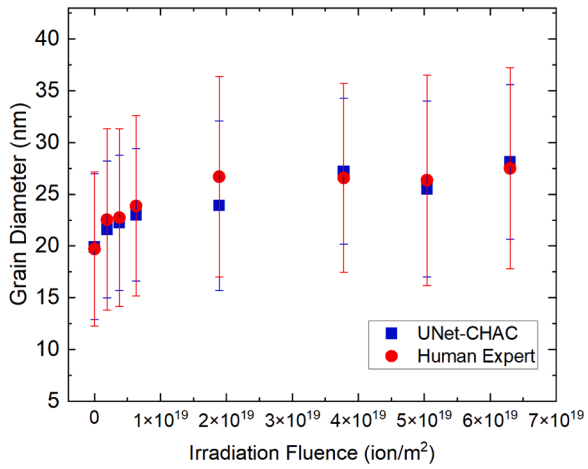

Fig. 9. Grain diameters measured by the UNet-CHAC method and human experts under different irradiation fluences at 875 K. The error bars correspond to the standard deviation of the measured grain diameters.

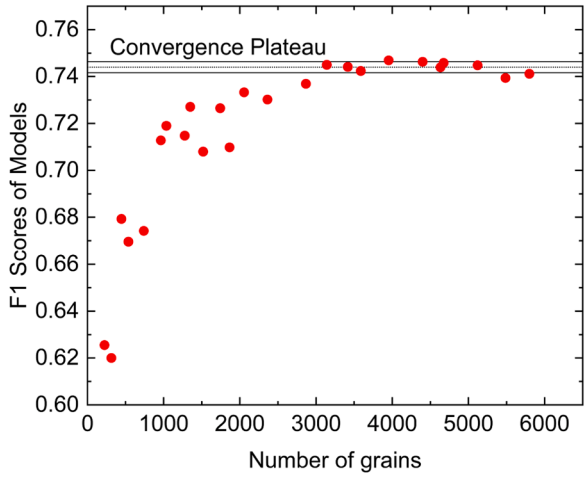

Fig. 10. F1 scores as a function of the number of well-annotated grain contours in the images of training dataset.

4. Discussion

The UNet+CHAC methodology conceived in this study brings two unique strengths to automated grain morphology analysis of BF-TEM images. First, it is inevitable for both machine learning models and human experts to mislabel some grain boundaries. In fact, even with the state-of-the-art performance (e.g., both precision and recall around 0.9), a considerable fraction of grain boundaries may still be mislabeled (around 10%). If not processed properly, these mislabeled grain boundaries can introduce significant errors in the grain size measurement. Based on the convex grain assumption, the CHAC method provides an effective approach for selecting well-labeled grain contours, thereby enabling accurate grain size measurement. Second, as shown in Section 3.3, the UNet model only needs a handful of BF-TEM images containing about 3000 grains in total to reach optimal performance. Therefore, UNet is a few-shot model, which is particularly useful for scenarios where collecting a large amount of labeled data is costly or time-consuming, such as identifying defects in TEM images. Currently, our UNet model was trained using only BF-TEM images of a nanocrystalline UO_2 film acquired from a specific TEM facility, so it may not achieve similar performance on BF-TEM images acquired from a different material or TEM facility. Nevertheless, by following the same training process and employing the UNet+CHAC approach, researchers can quickly train their own UNet with a few well-labeled images for automated grain size measurements.

To reduce the barriers for researchers to develop their own models, we will share all our data, including the original and labeled TEM images, the code of model training and the CHAC method, the trained UNet model, as well as a user manual. The details for accessing the data are provided in the Data Availability Section. Additionally, we call for the establishment of an online platform similar to the Materials Project [47,50], where researchers can contribute their images and annotations while the shared data can be reused by the wider scientific community. Such an infrastructure would greatly facilitate the development of more generalized and powerful CNN models. Moreover, recent progress in transfer learning has enabled machine learning models to effectively apply knowledge learned from a source task to another different but related new task, even if there is limited training data for the new task [51,52]. By leveraging the abundant well-labeled images on the platform, researchers could utilize transfer learning techniques to develop their own models for detecting various types of defects in TEM images.

Despite achieving commendable performance, the UNet+CHAC approach can be further improved. First, the current approach is ineffective in separating overlapped grains. Specifically, even after UNet correctly identifies the boundaries of two overlapped grains, the CHAC would usually separate them as three grains. Since overlapped grains are relatively rare in our TEM images, the impact on grain size measurement is negligible. We denote that resolving overlapped features is an intrinsic challenge for 2D imaging techniques. Application of 3D tomography techniques may help mitigate this limitation. Second, our TEM images were labeled by only two human experts, and we did not conduct the round-robin test to evaluate the errors of manual labels. As shown by recent studies, round-robin tests among multiple participants are effective in improving the accuracy of manual labels and reducing the bias of human experts, thereby enhancing the performance of the machine learning model [20,21]. We will include more human experts in future studies. Third, the F1 score of our UNet model (0.74) is relatively lower than state-of-the-art CNN models for general semantic segmentation tasks (typically above 0.8 or even 0.9). Moreover, F1 score would generally keep increasing as the training data size increases [40]. However, as shown in Fig. 10, the F1 score appeared to plateau when more than 3000 grains were used for training. Two factors may contribute to this phenomenon: (1) Model Limitations: The finite number of artificial neurons in the UNet model may constrain its ability to approximate the optimal function. Recent advances in CNN structures, such as U-Net++ [53], have shown improved performance than conventional UNet. (2) Data Noise: Training data, inherently carrying bias and errors from manual labeling, inevitably includes a fraction of noise. As size of training data expands, the value of additional data for performance improvement wanes due to the noise [54,55]. Therefore, potential solutions to this limitation could be the deployment of more advanced CNN models and the utilization of annotated training data with higher quality. It is worth noting that gathering high-quality annotated TEM images could be a labor-intensive and resource-demanding task. To tackle the data scarcity, we propose two solutions: a data-sharing platform, as discussed earlier, and the use of synthetic images. Notably, deep learning-based generative models like generative adversarial networks (GANs) have shown considerable potential in augmenting the volume and diversity of training images [56]. Researchers have successfully applied GANs to produce SEM images of polycrystalline iron and enhance the performance of their UNet model for identifying grain morphology in SEM images [57]. Similar GANs will be explored to generate synthetic TEM images for improving the model performance in our future studies.

5. Conclusions

In this work, we developed an automated approach for quantifying the grain size evolution of nanocrystalline UO_2 under *in-situ* ion

irradiation by combining the UNet model for machine learning and a computer vision algorithm named CHAC. The conclusions are:

1 The UNet+CHAC approach has proved to be effective and efficient in characterizing grain morphology in BF-TEM images, enabling high-throughput analysis of the data produced by *in-situ* irradiation experiments.
2 The UNet model only needs a few well-annotated BF-TEM images to achieve a performance comparable to human experts. Meanwhile, the CHAC method enables accurate grain size measurement even with potential grain boundary mislabeling by the UNet model.

Our studies indicate that the number of well-annotated grains plays a key role in controlling the performance of UNet model, so the primary challenge in enhancing the model performance lies in acquiring high-quality annotated TEM images. Establishing data sharing platforms and using synthetic data are promising approaches to addressing the data scarcity challenge.

Data availability statement

The authors confirm that all the data and codes supporting the findings of this study are available on the GitHub website https://github.com/SinnyJay1996/UNet-CHAC.

CRediT authorship contribution statement

Xinyuan Xu: Conceptualization, Methodology, Formal analysis, Software, Writing – original draft, Writing – review & editing. **Zefeng Yu:** Resources, Data curation, Writing – review & editing. **Wei-Ying Chen:** Resources, Writing – review & editing. **Aiping Chen:** Resources, Writing – review & editing. **Arthur Motta:** Conceptualization, Writing – review & editing. **Xing Wang:** Conceptualization, Supervision, Funding acquisition, Project administration, Writing – original draft, Writing – review & editing.

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.

Data availability

I have shared the link to my data/code in the manuscript.

Acknowledgment

X. Xu and X. Wang were supported by the faculty development grant from the Pennsylvania State University. This work was funded by the U. S. Department of Energy's Nuclear Engineering University Program project number 17-12797. Mr. Peter Baldo and Mr. Richard Sisson were thanked for operating the ion accelerator at Argonne National Laboratory. The work at Los Alamos National Laboratory was supported by the NNSA's Laboratory Directed Research and Development Program and was performed, in part, at the Center for Integrated Nanotechnologies, an Office of Science User Facility operated for the U.S. Department of Energy Office of Science. Z. Yu and A. Motta are CINT users and acknowledge the support from CINT users' program.

References

[1] H. Clemens, S. Mayer, C. Scheu, Microstructure and properties of engineering materials. Neutrons and Synchrotron Radiation in Engineering Materials Science, Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany, 2017, pp. 1–20, https://doi.org/10.1002/9783527684489.ch1.
[2] M. Ohring, How engineering materials are strengthened and toughened. Engineering Materials Science, Elsevier, 1995, pp. 431–500, https://doi.org/10.1016/B978-012524995-9/50033-7.
[3] E.O. Hall, The deformation and ageing of mild steel: III discussion of results, Proc. Phys. Soc. Sect. B 64 (1951) 747–753, https://doi.org/10.1088/0370-1301/64/9/303.
[4] J.H. Lee, Y.H. Kim, S.J. Ahn, T.H. Ha, H.S. Kim, Grain-size effect on the electrical properties of nanocrystalline indium tin oxide thin films, Mater. Sci. Eng. B 199 (2015) 37–41, https://doi.org/10.1016/j.mseb.2015.04.011.
[5] H. Lin, C. Huang, W. Li, C. Ni, S. Shah, Y. Tseng, Size dependency of nanocrystalline TiO_2 on its optical property and photocatalytic reactivity exemplified by 2-chlorophenol, Appl. Catal. B 68 (2006) 1–11, https://doi.org/10.1016/j.apcatb.2006.07.018.
[6] W. Goll, H.P. Fuchs, R. Manzel, F.U. Schlemmer, Irradiation behavior of UO_2 /PuO_2 fuel in light water reactors, Nucl. Technol. 102 (1993) 29–46, https://doi.org/10.13182/NT93-A34800.
[7] S.M. Mirvakili, M. Alizadeh Kavafshary, A. Joze Vaziri, Comparison of neutronic behavior of UO_2, (Th- 233 U)O_2 and (Th- 235 U)O_2 fuels in a typical heavy water reactor, Nucl. Eng. Technol. 47 (2015) 315–322, https://doi.org/10.1016/j.net.2014.12.014.
[8] G. Locatelli, M. Mancini, N. Todeschini, Generation IV nuclear reactors: current status and future prospects, Energy Policy 61 (2013) 1503–1520, https://doi.org/10.1016/j.enpol.2013.06.101.
[9] J.A. Turnbull, The effect of grain size on the swelling and gas release properties of UO_2 during irradiation, J. Nucl. Mater. 50 (1974) 62–68, https://doi.org/10.1016/0022-3115(74)90061-0.
[10] M. Liu, Q. Kang, H. Xu, Grain-scale study of the grain boundary effect on UO_2 fuel oxidation and fission gas release under reactor conditions, Chem. Eng. Sci. 229 (2021), 116026, https://doi.org/10.1016/j.ces.2020.116026.
[11] Z. Yu, X. Xu, W.Y. Chen, Y. Sharma, X. Wang, A. Chen, C.J. Ulmer, A.T. Motta, *In-situ* irradiation-induced studies of grain growth kinetics of nanocrystalline UO_2, Acta Mater. 231 (2022), 117856, https://doi.org/10.1016/j.actamat.2022.117856.
[12] J. Liu, A.H. Mir, G. He, M. Danaie, J. Hinks, S. Donnelly, H. Nordin, S. Lozano-Perez, C.R.M. Grovenor, *In-situ* TEM study of irradiation-induced damage mechanisms in monoclinic-ZrO_2, Acta Mater. 199 (2020) 429–442, https://doi.org/10.1016/j.actamat.2020.08.064.
[13] F. Cappia, M. Cullison, T. Chen, B. Kombaiah, K. Bawane, F. Teng, J. Madden, E. Perez, T. Yao, P. Lei, J. Lian, Y. Miao, K. Mo, Grain subdivision and structural modifications by high-energy heavy ions in UO_2 with different initial grain size, Nucl. Instrum. Methods Phys. Res. B 515 (2022) 48–60, https://doi.org/10.1016/j.nimb.2021.06.020.
[14] A. Agrawal, A. Choudhary, Perspective: materials informatics and big data: realization of the "fourth paradigm" of science in materials science, APL Mater. 4 (2016), 053208, https://doi.org/10.1063/1.4946894.
[15] S.V. Kalinin, C. Ophus, P.M. Voyles, R. Erni, D. Kepaptsoglou, V. Grillo, A. R. Lupini, M.P. Oxley, E. Schwenker, M.K.Y. Chan, J. Etheridge, X. Li, G.G.D. Han, M. Ziatdinov, N. Shibata, S.J. Pennycook, Machine learning in scanning transmission electron microscopy, Nat. Rev. Methods Primers 2 (2022) 11, https://doi.org/10.1038/s43586-022-00095-w.
[16] Z. Chen, N. Andrejevic, N.C. Drucker, T. Nguyen, R.P. Xian, T. Smidt, Y. Wang, R. Ernstorfer, D.A. Tennant, M. Chan, M. Li, Machine learning on neutron and X-ray scattering and spectroscopies, Chem. Phys. Rev. 2 (2021), 031301, https://doi.org/10.1063/5.0049111.
[17] E.A. Holm, R. Cohn, N. Gao, A.R. Kitahara, T.P. Matson, B. Lei, S.R. Yarasi, Overview: computer vision and machine learning for microstructural characterization and analysis, Metall. Mater. Trans. A 51 (2020) 5985–5999, https://doi.org/10.1007/s11661-020-06008-4.
[18] R. Jacobs, Deep learning object detection in materials science: current state and future directions, Comput. Mater. Sci. 211 (2022), 111527, https://doi.org/10.1016/j.commatsci.2022.111527.
[19] X. Xu, Z. Yu, A. Motta, X. Wang, Automated analysis of grain growth under *in-situ* irradiation using convolutional neural network, Microsc. Microanal. 28 (2022) 2036–2037, https://doi.org/10.1017/S1431927622007899.
[20] G. Roberts, S.Y. Haile, R. Sainju, D.J. Edwards, B. Hutchinson, Y. Zhu, Deep learning for semantic segmentation of defects in advanced STEM images of steels, Sci. Rep. 9 (2019) 12744, https://doi.org/10.1038/s41598-019-49105-0.
[21] W. Li, K.G. Field, D. Morgan, Automated defect analysis in electron microscopic images, npj Comput. Mater. 4 (2018) 36, https://doi.org/10.1038/s41524-018-0093-8.
[22] M. Shen, G. Li, D. Wu, Y. Liu, J.R.C. Greaves, W. Hao, N.J. Krakauer, L. Krudy, J. Perez, V. Sreenivasan, B. Sanchez, O. Torres-Velázquez, W. Li, K.G. Field, D. Morgan, Multi defect detection and analysis of electron microscopy images with deep learning, Comput. Mater. Sci. 199 (2021), 110576, https://doi.org/10.1016/j.commatsci.2021.110576.
[23] R. Jacobs, P. Patki, M.J. Lynch, S. Chen, D. Morgan, K.G. Field, Materials swelling revealed through automated semantic segmentation of cavities in electron microscopy images, Sci. Rep. 13 (2023) 5178, https://doi.org/10.1038/s41598-023-32454-2.
[24] C.M. Anderson, J. Klein, H. Rajakumar, C.D. Judge, L.K. Béland, Automated detection of helium bubbles in irradiated X-750, Ultramicroscopy 217 (2020), 113068, https://doi.org/10.1016/j.ultramic.2020.113068.
[25] X. Wang, K. Jin, C.Y. Wong, D. Chen, H. Bei, Y. Wang, M. Ziatdinov, W.J. Weber, Y. Zhang, J. Poplawsky, K.L. More, Understanding effects of chemical complexity on helium bubble formation in Ni-based concentrated solid solution alloys based

on elemental segregation measurements, J. Nucl. Mater. 569 (2022), 153902, https://doi.org/10.1016/j.jnucmat.2022.153902.
[26] Y. Meng, Z. Zhang, H. Yin, T. Ma, Automatic detection of particle size distribution by image analysis based on local adaptive canny edge detection and modified circular Hough transform, Micron 106 (2018) 34–41, https://doi.org/10.1016/j.micron.2017.12.002.
[27] S. Banerjee, P.C. Chakraborti, S.K. Saha, An automated methodology for grain segmentation and grain size measurement from optical micrographs, Measurement 140 (2019) 142–150, https://doi.org/10.1016/j.measurement.2019.03.046.
[28] A. Campbell, P. Murray, E. Yakushina, S. Marshall, W. Ion, New methods for automatic quantification of microstructural features using digital image processing, Mater. Des. 141 (2018) 395–406, https://doi.org/10.1016/j.matdes.2017.12.049.
[29] A. Bordas, J. Zhang, J.C. Nino, Application of deep learning workflow for autonomous grain size analysis, Molecules 27 (2022) 4826, https://doi.org/10.3390/molecules27154826.
[30] B. Ma, X. Ban, H. Huang, Y. Chen, W. Liu, Y. Zhi, Deep learning-based image segmentation for Al-La alloy microscopic images, Symmetry 10 (2018) 107, https://doi.org/10.3390/sym10040107 (Basel).
[31] O. Ronneberger, P. Fischer, T. Brox, U-Net: convolutional networks for biomedical image segmentation, in: 2015: pp. 234–241. 10.1007/978-3-319-24574-4_28.
[32] E. Enriquez, G. Wang, Y. Sharma, I. Sarpkaya, Q. Wang, D. Chen, N. Winner, X. Guo, J. Dunwoody, J. White, A. Nelson, H. Xu, P. Dowden, E. Batista, H. Htoon, P. Yang, Q. Jia, A. Chen, Structural and optical properties of phase-pure UO_2, α-U_3O_8, and α-UO_3 epitaxial thin films grown by pulsed laser deposition, ACS Appl. Mater. Interfaces 12 (2020) 35232–35241, https://doi.org/10.1021/acsami.0c08635.
[33] Y. Sharma, B. Paudel, A. Huon, M.M. Schneider, P. Roy, Z. Corey, R. Schönemann, A.C. Jones, M. Jaime, D.A. Yarotski, T. Charlton, M.R. Fitzsimmons, Q. Jia, M. T. Pettes, P. Yang, A. Chen, Induced ferromagnetism in epitaxial uranium dioxide thin films, Adv. Sci. 9 (2022), 2203473, https://doi.org/10.1002/advs.202203473.
[34] L. Alzubaidi, J. Zhang, A.J. Humaidi, A. Al-Dujaili, Y. Duan, O. Al-Shamma, J. Santamaría, M.A. Fadhel, M. Al-Amidie, L. Farhan, Review of deep learning: concepts, CNN architectures, challenges, applications, future directions, J. Big Data 8 (2021) 53, https://doi.org/10.1186/s40537-021-00444-8.
[35] A.F. Agarap, Deep learning using rectified linear units (relu), ArXiv Preprint ArXiv:1803.08375. (2018).
[36] H. Cao, Y. Wang, J. Chen, D. Jiang, X. Zhang, Q. Tian, M. Wang, Swin-unet: Unet-like pure transformer for medical image segmentation, in: Proceedings of the Computer Vision–ECCV 2022 Workshops, Tel Aviv, Israel, October 23–27, 2022, 2023, pp. 205–218. Part III.
[37] M. Sangalli, S. Blusseau, S. Velasco-Forero, J. Angulo, Scale Equivariant U-Net, (2022).
[38] Y.A. Coutinho, S.C.K. Rooney, E.J. Payton, Analysis of EBSD grain size measurements using microstructure simulations and a customizable pattern matching library for grain perimeter estimation, Metall. Mater. Trans. A 48 (2017) 2375–2395, https://doi.org/10.1007/s11661-017-4031-z.
[39] K.P. Mingard, B. Roebuck, E.G. Bennett, M.G. Gee, H. Nordenstrom, G. Sweetman, P. Chan, Comparison of EBSD and conventional methods of grain size measurement of hardmetals, Int. J. Refract. Met. Hard Mater. 27 (2009) 213–223, https://doi.org/10.1016/j.ijrmhm.2008.06.009.
[40] I. Goodfellow, Y. Bengio, A. Courville, Deep Learning, MIT Press, 2016.
[41] W.W. Mullins, Two-dimensional motion of idealized grain boundaries, J. Appl. Phys. 27 (1956) 900–904, https://doi.org/10.1063/1.1722511.
[42] J. Sklansky, Finding the convex hull of a simple polygon, Pattern Recognit. Lett. 1 (1982) 79–83, https://doi.org/10.1016/0167-8655(82)90016-2.
[43] U. Ramer, An iterative procedure for the polygonal approximation of plane curves, Comput. Graph. Image Process. 1 (1972) 244–256, https://doi.org/10.1016/S0146-664X(72)80017-0.
[44] D.H. Douglas, T.K. Peucker, Algorithms for the reduction of the number of points required to represent a digitized line or its caricature, Cartogr. Int. J. Geogr. Inf. Geovis. 10 (1973) 112–122, https://doi.org/10.3138/FM57-6770-U75U-7727.
[45] M. Buckland, F. Gey, The relationship between recall and precision, J. Am. Soc. Inf. Sci. 45 (1994) 12–19, https://doi.org/10.1002/(SICI)1097-4571(199401)45:1<12::AID-ASI2>3.0.CO;2-L.
[46] S. Xie, Z. Tu, Holistically-nested edge detection, in: Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2015, pp. 1395–1403, https://doi.org/10.1109/ICCV.2015.164.
[47] R. Jacobs, M. Shen, Y. Liu, W. Hao, X. Li, R. He, J.R.C. Greaves, D. Wang, Z. Xie, Z. Huang, C. Wang, K.G. Field, D. Morgan, Performance and limitations of deep learning semantic segmentation of multiple defects in transmission electron micrographs, Cell Rep. Phys. Sci. 3 (2022), 100876, https://doi.org/10.1016/j.xcrp.2022.100876.
[48] P. Bateni, R. Goyal, V. Masrani, F. Wood, L. Sigal, Improved Few-Shot Visual Classification, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2020, pp. 14481–14490, https://doi.org/10.1109/CVPR42600.2020.01450.
[49] S. Akers, E. Kautz, A. Trevino-Gavito, M. Olszta, B.E. Matthews, L. Wang, Y. Du, S. R. Spurgeon, Rapid and flexible segmentation of electron microscopy data using few-shot machine learning, npj Comput. Mater. 7 (2021) 187, https://doi.org/10.1038/s41524-021-00652-z.
[50] A. Jain, S.P. Ong, G. Hautier, W. Chen, W.D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder, K.A. Persson, Commentary: the materials project: a materials genome approach to accelerating materials innovation, APL Mater. 1 (2013), 011002, https://doi.org/10.1063/1.4812323.
[51] K. Weiss, T.M. Khoshgoftaar, D. Wang, A survey of transfer learning, J. Big Data 3 (2016) 9, https://doi.org/10.1186/s40537-016-0043-6.
[52] H.C. Shin, H.R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, R. M. Summers, Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning, IEEE Trans. Med. Imaging 35 (2016) 1285–1298, https://doi.org/10.1109/TMI.2016.2528162.
[53] Z. Zhou, M.M.R. Siddiquee, N. Tajbakhsh, J. Liang, UNet++: a nested U-Net architecture for medical image segmentation, (2018).
[54] C. Northcutt, L. Jiang, I. Chuang, Confident learning: estimating uncertainty in dataset labels, J. Artif. Intell. Res. 70 (2021) 1373–1411, https://doi.org/10.1613/jair.1.12125.
[55] S. Gupta, A. Gupta, Dealing with noise problem in machine learning data-sets: a systematic review, Procedia Comput. Sci. 161 (2019) 466–474, https://doi.org/10.1016/j.procs.2019.11.146.
[56] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial networks, Commun. ACM 63 (2020) 139–144, https://doi.org/10.1145/3422622.
[57] 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) 125, https://doi.org/10.1038/s41524-020-00392-6.