

Deep Learning Object Detection in Materials Science: Current State and Future Directions

Ryan Jacobs^{1,*}

¹ Department of Materials Science and Engineering, University of Wisconsin-Madison, Madison, WI, USA.

*Corresponding author email: rjacobs3@wisc.edu

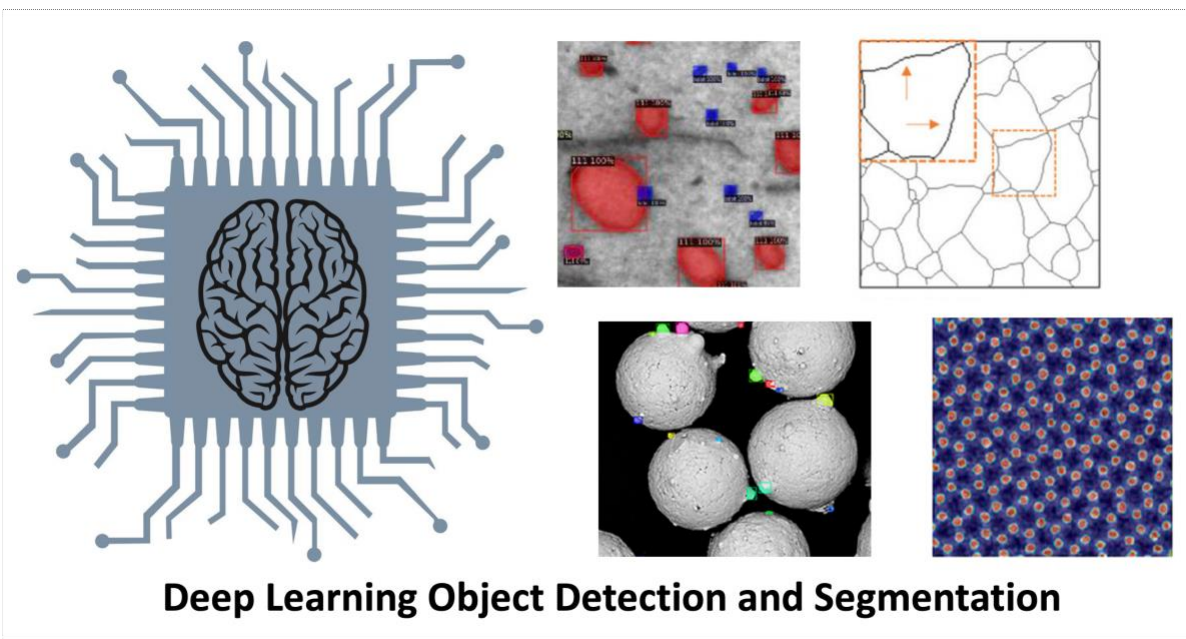
Keywords: object detection, semantic segmentation, deep learning, computer vision, machine learning, electron microscopy

Abstract:

Deep learning-based object detection models have recently found widespread use in materials science, with rapid progress made in just the past two years. Scanning and tunneling electron microscopy methods are among the most important and widely used characterization techniques for understanding fundamental materials structure-property-performance linkages from the micron to atomic scale. Dramatic increases in dataset size and complexity from modern electron microscopy instruments have necessitated the development and use of automated methods of extracting pertinent features of images. Here, the use of object detection in materials science, with a focus on the analysis of features in electron microscopy images, is reviewed. Key findings and limitations of recent seminal studies using object detection to characterize and quantify defects in irradiated metal alloys, segment and analyze micro and nanoparticles, find individual atoms at the nanoscale, and detect and track objects from *in situ* video are reviewed. Opportunities and challenges presently facing the materials community are highlighted, where discussion of best practices for model assessment and applicability are presented, along with the potential of improved model training with synthetic data. This review concludes with offering more speculative, forward-looking thoughts on the potential of the broader materials community to construct a living ecosystem integrating community-consensus curated data and validated

models as tools to best inform application of object detection and segmentation models to specific materials domains.

Graphical Abstract:



1 Introduction

Materials characterization using electron microscopy (EM) methods such as scanning electron microscopy (SEM), transmission electron microscopy (TEM), scanning TEM (STEM) and high-resolution TEM (HRTEM) are some of the most widely used and important characterization techniques in all of materials science. EM methods are routinely used to analyze the structure of materials at lengths scales ranging from millimeters using SEM to the atomic scale, such as single atom imaging in aberration corrected STEM, where the resolution and field-of-view depend on the technique used, the conditions of the electron beam such as the probe size, and the desired feature of interest size. Recent advances in imaging and analytical detectors continues to push towards improved resolutions while maintaining or even increasing the field-of-view in EM based experiments. The continued growth of techniques, probe-forming technologies, and detector capabilities has led to EM methods being an essential method for developing understanding of

material structure-property-performance linkages, therefore informing rational materials design and engineering strategies for a breadth of modern technologies for the past several decades.

Typically, analysis of EM images is performed manually by researchers, using image analysis programs such as ImageJ[1] and cisTEM.[2] The use of such image analysis programs enables a researcher to manually annotate and quantify objects in images, such as the size and shape of nanoparticles or dislocation loops. However, the use of such manual methods is error prone (humans make mistakes), inconsistent (different people, even domain experts in the same field, will interpret results differently), time consuming (humans are slow compared to computers at certain tasks), and not scalable to large dataset sizes. In the long term the most problematic of these four issues of manual image analysis is probably that of scalability to large dataset sizes. The importance of scaling emerges from the fact that modern EM instruments have witnessed an exponential increase in data acquisition rates with the development of new detector technologies, resulting in not only higher resolution images and larger overall data sizes, but also more physically complex image data. Modern instruments are capable of acquiring multiple terabytes of data in a single session, in the form of real-time video and both spatially- and momenta-resolved (4D) data (see **Figure 1**).[3–6] Therefore, there is a need for automated analysis tools to efficiently analyze this ever-increasing amount of data, and a detailed analysis and understanding of the performance capabilities and applicability of such methods.

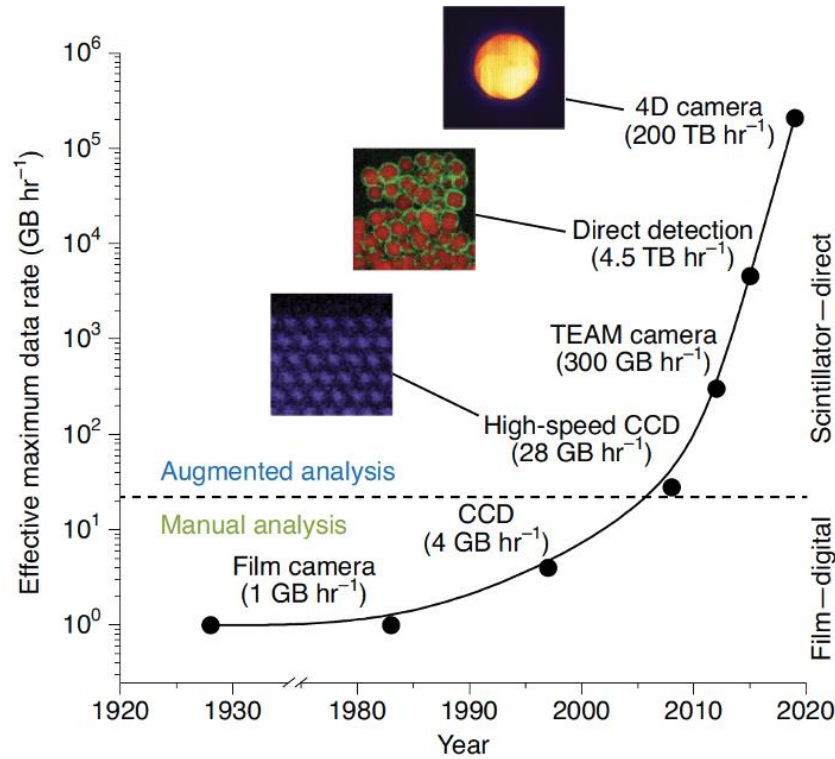


Figure 1. Growth of EM data acquisition over time, driven mainly by advancements in detector technology and further enabled by advancements in modern computing and data storage. Modern EM instruments are capable of cataloguing tens to hundreds of terabytes of data in a single session, motivating the need for automatic image analysis methods. Figure adapted with permission from Ref [3].

In the past ten years, the maturation of deep learning[7–9] methods has led to a series of stunning advances across a number of scientific and medical fields, ranging from the development of real-time language translation, autonomous vehicles,[10] superhuman facial recognition,[11] and mastery of games such as Go, Poker and Chess.[12] Deep learning methods are typically defined as neural networks with two or more hidden layers, resulting in at least four total layers including input and output layers. Deep neural networks may consist of fully connected dense layers, called multilayer perceptrons, but for deep learning tasks of image classification and object detection, convolutional neural network (CNN) models are typically used. As a basic description, provided a series of labeled examples, such as images that contain pictures of cats, CNN models perform a series of what are called convolution and pooling operations (to be described below, one operation is one layer in the network) which enable the CNN to learn which features of the image make it recognizable to the model as a cat. These

features may be aspects of the image like brightness, contrast, or patterns of edges. As shown in **Figure 2A**, each convolution operation consists of a sliding window (e.g., a 5x5 pixel box) which rasters over the image, convolving nearby pixels together to effectively create a lower dimensional embedding of the original image. The pooling operation down-samples the feature maps produced from convolution by typically taking the average pixel intensity value (average pooling) or maximum pixel intensity (max pooling) of the convolved window. Repeated sequences of convolution and pooling enable the CNN to learn abstract features of the image which become encoded as tunable weights for each convolution and pooling operation. The resulting lower dimensional embedding of the original image is finally flattened to a one-dimensional numerical vector, which is subsequently fed into a fully connected network to produce a final classification label or regression value (see **Figure 2A**). [13] Modern CNN models may contain dozens or even more than one hundred layers and millions of adjustable parameters. The power of deep learning lies in the ability of CNN-based models to learn abstract features of objects of interest in an image which may be broadly applicable to a more general problem via transfer learning. As an example, a CNN model used to classify different images of common objects such as cats, people, and stop signs can be transferred and fine-tuned to also detect dislocation loops, nanoparticles, or individual atoms in EM images.

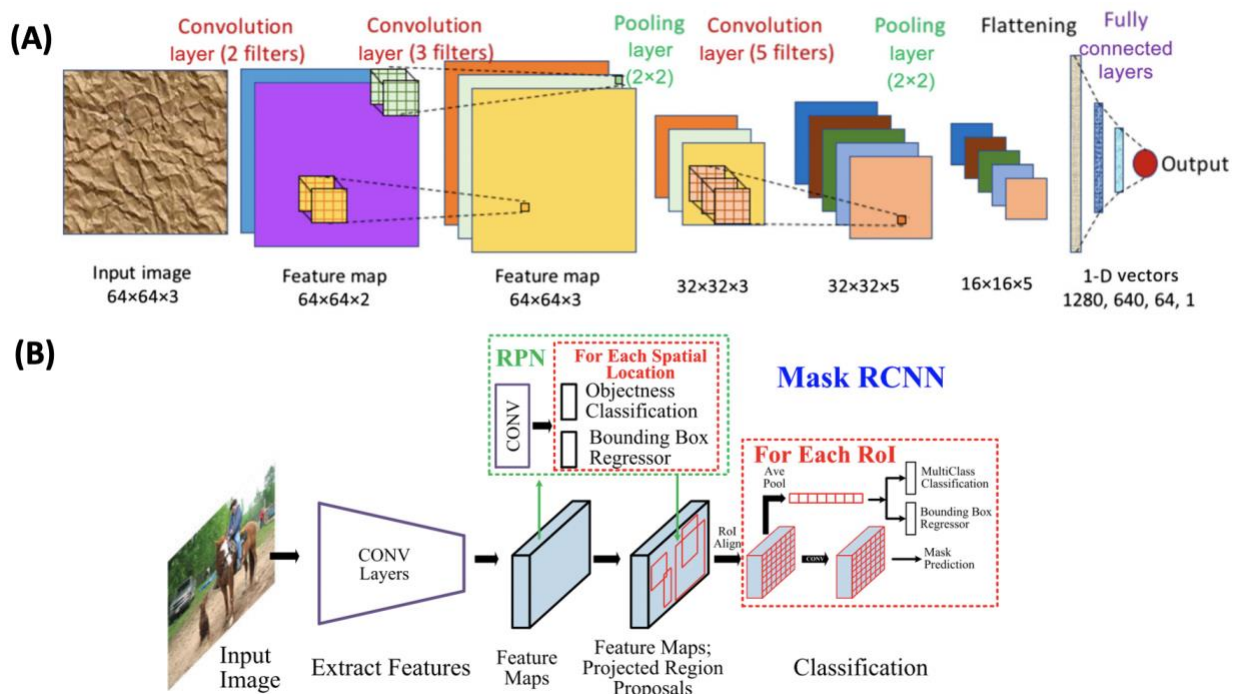


Figure 2. (A) Example CNN model consisting of three convolution layers, two pooling layers, and a flattening operation for feeding into a final three-layer fully connected network. The original input image of 64x64 pixels and three channels is transformed into 16x16 lower dimensional embeddings with five channels. (B) Architecture overview of the Mask R-CNN model. Figures adapted with permission from Ref [13] and Ref [14].

Deep learning has enabled revolutionary changes to the field of computer vision. Traditional computer vision image analysis steps such as blurring, masking, thresholding, etc. for classifying and detecting objects of interest are rapidly being supplanted by deep learning-based models. Deep CNNs such as ResNet50, ResNet101[15] and VGG16[16] are typically used to extract detailed underlying feature sets from tens of thousands of images in canonical databases such as ImageNet[17] and Common Objects in Context (CoCo).[18] As shown in **Figure 2B**, these networks are then employed as “backbones” in more complex **object detection frameworks**, such as the regional convolutional neural network (R-CNN) models like Faster R-CNN[19] and Mask R-CNN,[20] which use the deep CNN backbones mentioned above, combined with additional neural networks used to suggest regions of interest in the image and classify and segment individual objects within each region of interest. Additional information on the development and technical details for a complete list of network types used for object detection can be found in excellent recent reviews such as Ref [21] and Ref [14].

Figure 3 presents an overview of different object detection tasks. **Figure 3A** is an example of an image classification task, where the contents of the image are classified (here, as material powder particles), though no analysis of the exact location or other characteristics of the objects is performed. Image classification may be performed using a deep CNN architecture such as ResNet50 or VGG16. **Figure 3B** performs semantic segmentation of the image contents with pixel masking, where the background is shown as black, and the powder particles shown as yellow. One of the most popular models to perform semantic segmentation is U-Net.[22] While the use of U-Net can classify and segment objects at the pixel-level, the use of a single mask for an entire image makes per-object analysis (e.g., quantifying size and shape of each particle) challenging. **Figure 3C** shows an example of object detection, where a bounding box is drawn around each detected instance of a particle, but without pixel-level information. This type of object detection is obtainable with, for example, the Faster R-CNN model. Finally, **Figure 3D** is a demonstration of

instance segmentation, which provides a marriage of the pixel-level segmentation and object detection, providing pixel-level masks for each detected particle. This level of object detection is obtained by employing models such as Mask R-CNN, Cascade R-CNN, and more recently developed models such as vision transformers (ViT).[23] It is worth noting there are tradeoffs one must consider when choosing the model and type of object detection task to pursue, where the preferred method will generally depend on the application at hand. For example, while the instance segmentation approach shown in **Figure 3D** provides more information in the form of detailed pixel-level and per-object detection information compared to image classification in **Figure 3A**, the models needed to perform instance segmentation (e.g., Mask R-CNN) take much longer to train than those for image classification (e.g. ResNet50), and the preparation of the training dataset for instance segmentation is much more time consuming, where every object of interest must be annotated at the pixel level, as opposed to supplying a single class label for each image.

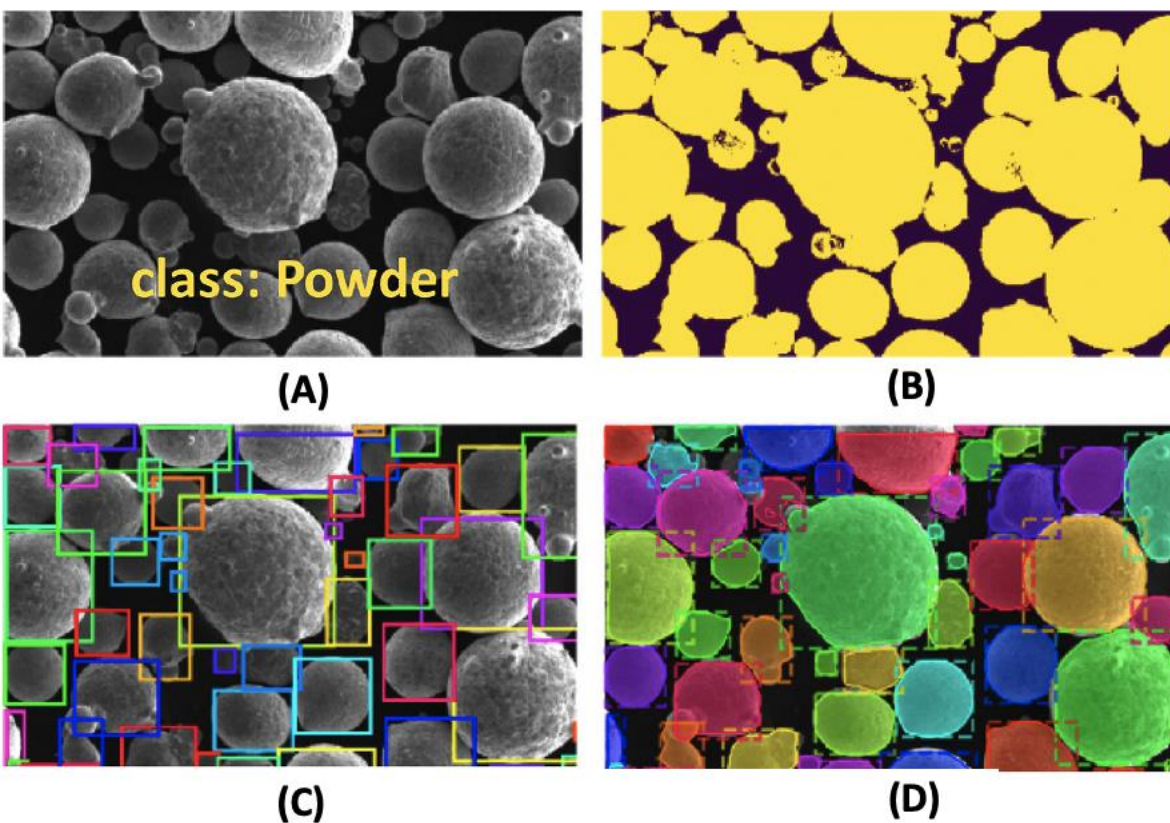


Figure 3. Overview of different types of object detection tasks. (A) EM image of nanoparticles in a powder, with classification of “powder”. (B) Pixel-level segmentation of the image from (A)

showing semantic segmentation of background (black) vs. particle (yellow), obtained from, for example, a U-Net model. (C) Object detection of individual particles in the powder, where each identified particle is encapsulated with a bounding box obtained from, for example, a Faster R-CNN model. (D) Instance segmentation combining object detection with pixel-level detail to obtain bounding boxes and segmentation masks for each particle, obtained using, for example, a Mask R-CNN model. Figure adapted with permission from Ref [24].

Deep learning has witnessed increased adoption in materials science in the past five years,[13,25–30] and the use of deep learning-based object detection in materials science began in earnest in 2018 and has advanced rapidly in just the past two years.[24,31–33] Currently, there is significant simultaneous progress in the computer science field of developing new object detection algorithms,[21,23,34,35] and the subsequent application and assessment of object detection models in materials science.[31,36–40] The purpose of this review is to provide a brief summary of the present state-of-the-art use of object detection and quantification in materials science, with a particular emphasis on studies conducted in the past two years using object detection models to analyze EM images. The methods and key findings of a set of pioneering studies employing object detection to EM images are discussed (Section 2), including applications of quantifying defects in images of irradiated metal alloys (Section 2.1), characterizing micro and nanoparticles (Section 2.2), finding individual atoms in STEM images (Section 2.3), and tracking detected objects from *in situ* EM video (Section 2.4). Following this, in Section 3 some opportunities and challenges presently facing the materials community are highlighted, with discussion of some suggested best practices for evaluating the performance of object detection models (Section 3.1), the use of synthetic data generation to improve model training (Section 3.2), the possible influence of subjectivity in ground truth labels and the prospect of achieving community consensus of data labeling (Section 3.3), and prospects for community-based model development and deployment (Section 3.4).

2 Applications of Object Detection in Materials Science

2.1 Quantifying Defects in Electron Microscopy Images of Irradiated Metal Alloys

One key application area of automatically detecting and analyzing defects in EM images is that of metal alloys which have undergone some form of irradiation. Metal alloys used in nuclear reactor components undergo irradiation, and the type, shape, size and number distribution of defects formed result in hardening, embrittlement and swelling of the material.[41] Such changes in the mechanical properties of the alloys directly impact its performance in the reactor, and understanding such property changes **is** critically important for the safe and reliable operation of reactor facilities.[42] Up until recently, radiation-induced defects in metal alloy EM images were manually quantified by domain-expert researchers.[43–45] Since 2018, with the pioneering work of Li et al.,[31] a number of informative studies have successfully characterized and analyzed multiple radiation-induced defect types in EM images of metal alloys used in nuclear applications. In this section, the key methods used, main results and some limitations of each of these studies are summarized.

Beginning in 2018, the work of Li et al. sought to quantify defects in irradiated ferritic FeCrAl alloys.[31] While the detection of both $\langle 111 \rangle$ - and $\langle 100 \rangle$ -type dislocation loops, along with interstitial cluster (so-called “black dot”) defects were of interest for detection, Li et al. focused on detecting only the $\langle 111 \rangle$ -type dislocation loops as a proof of concept, with the hypothesis that if the model can correctly perform detection of the $\langle 111 \rangle$ loops, extension of the method to detect other defect types would be straightforward. A series of methods were used to detect loops in this work. First, proposed bounding boxes were obtained by training a non-deep learning based cascade object detector (using an AdaBoost model) to locate probable regions of an image containing a defect. Next, a 15-layer CNN classifier was used to refine the bounding box predictions by classifying whether proposed bounding boxes likely contained a defect or not. Finally, the watershed flood algorithm (a traditional computer vision method) was used to segment the exact defect locations and morphologies. **Figure 4A** provides an example EM image of the defect loops and the model predictions from this study. Generally, this work showed that automated defect detection methods hold promise for comparable performance to human researchers, with orders of magnitude faster inference. Two key limitations to the work from Li et al. were the ability to only identify a single type of defect, and the lack of pixel-level segmentation information from the deep learning model. Only predicting a single defect type

limits the information the model can provide for subsequent materials property modeling of radiation-induced hardening, which would benefit from detailed size and number distributions of all defect types. Second, the use of traditional computer vision methods (e.g., watershed flood algorithm) required extensive manual tuning to obtain the desired performance, thus limiting the domain of applicability, where subtle differences in image conditions, defect geometry, and material type may result in significant deterioration in model performance.[40]

The study of Shen et al. extended the work of Li et al. by using the Faster R-CNN object detection model on the same FeCrAl alloy data from Li et al., advancing the state of the art by characterizing multiple defect types with a fully deep learning approach (see **Figure 4B**). This work used model evaluation statistics such as the defect find F1 score and defect identification F1 score to quantify model performance. In general, the F1 score is the harmonic mean of the precision and recall. Here, the defect find F1 score considers whether the model found a defect at the correct location in an image, regardless if the defect type is correct. The defect identification F1 score is a more demanding test, as it considers whether the model found a defect at the correct location *and* classified the defect type correctly. The work of Shen et al. showed a defect find F1 score and a defect identification F1 score of about 0.8 and 0.7, respectively, with defect size errors of about 10% and defect areal density errors in the 25-50% range. This work demonstrated the power of leveraging modern object detection algorithms and showcased very good overall predictive ability on par with domain-expert human labelers. However, a shortcoming with the use of Faster R-CNN is the lack of pixel-level information, so traditional computer vision methods were again used to obtain the pixel-level segmentation information to extract details of predicted defect sizes.[37] Furthermore, only a single set of training and test image data was used in the analysis of model performance, raising questions regarding the variation in model performance with factors such as dataset size, image type (e.g. imaging condition, amount of irradiation), image quality and train vs. test image set.

Defect detection of dislocation loops and black dots in FeCrAl alloys was further refined with the recent work of Jacobs et al.,[40] who employed refined and expanded versions of the dataset from Shen et al. and used a Mask R-CNN model to realize a fully end-to-end deep learning approach to quantify the size and shape distributions and densities of these defects (see **Figure**

4C). Jacobs et al. found generally comparable performance to Shen et al. on defect find and identification tasks with F1 scores of about 0.8. Up until this work, only the quality of the detections were quantified, e.g., through precision, recall and F1 scores. As a key result from this work, the instance segmentation capabilities of Mask R-CNN were used to form detailed per-defect size and shape distributions and density calculations, which, when combined with a simplified dispersion hardening model, resulted in hardening predictions with errors of about 10-20 MPa, which is about 10% of total material hardening and within the margin of error of experiments. This result is a concrete example of automated object detection in EM images directly being used to inform materials property predictions. Assessing the final quantification of detected objects is a unique aspect of using deep learning approaches in materials science compared to the field of computer science, which is generally more **concerned** with assessing detection quality. This work also provided a number of detailed tests to aid in informing the expected performance and possible limitations of this and similar models, which will be discussed in more detail in Section 3.1. While this work helped codify useful tests and expected performance of automated defect detection models, limitations remain, such as analysis to interpret and understand decisions made by the Mask R-CNN model through, for example, feature and activation maps. In addition, given the rapid advancement of object detection algorithms, more state-of-the-art methods than Mask R-CNN, such as vision transformers (ViTs),[23,35] may offer improved performance, offering a potential means toward identifying small features in TEM images (e.g., those that **are** less than 10 pixels in size), which are presently difficult to detect reliably using R-CNN models.

There are additional seminal studies employing object detection models for characterizing defects in irradiated alloys. The work of Anderson et al. used the Faster R-CNN model to detect microstructural voids (also sometimes referred to as cavities) formed by helium bubbles in irradiated X-750 Ni-based superalloys, as shown in **Figure 4D**. Similar to the detection of dislocation loops discussed above, the defect find F1 score for void detection is about 0.8, and the defect size distributions are well-reproduced by the model. In addition, like the works of Li et al. and Shen et al., this study also used additional post-processing methods separate from the deep learning model to extract the void size information using the predicted bounding boxes.[38]

This study was the first of its kind to apply automated object detection techniques to predict void formation, but additional work should be done to assess model applicability to a more varied dataset, and explore segmentation methods to achieve a fully deep-learning based approach. Further, analysis of material swelling, the key material property to track during void formation, was not performed in this work. Swelling is a function of both void size and number distributions, and being able to quickly and reliably predict the swelling is key to informing service and deployment of these materials in nuclear applications.

As a final example, the work of Roberts et al. employed a model called DefectSegNet, a custom U-Net model architecture, and was the first study to demonstrate pixel-level segmentation of multiple defect types in EM images, as shown in **Figure 4E**. In this work, the custom U-Net model was trained to detect line dislocations, precipitate particles, and voids in HT-9 martensitic steels. The model produced extremely high pixel classification accuracies of >90%, with pixel accuracy approaching 99% for void identification, with performance rivaling or exceeding human domain experts. However, as discussed in Section 1, U-Net based methods, while providing pixel-level information, have the drawback of not providing individual object detection, making subsequent analysis of object properties like size and shape difficult. In addition, given the single-mask classification nature of U-Net models, a separate model must be trained for each defect type (dislocation line, void, precipitate), in contrast to Faster and Mask R-CNN, which can classify multiple object types within a single trained model. The nature of U-Net produces the limitation that classifying multiple defect types in an image, for example, a mixture of dislocation lines and precipitates, would likely require the application of two separate models and some form of post-processing for dealing with pixels that are classified as more than one object type.

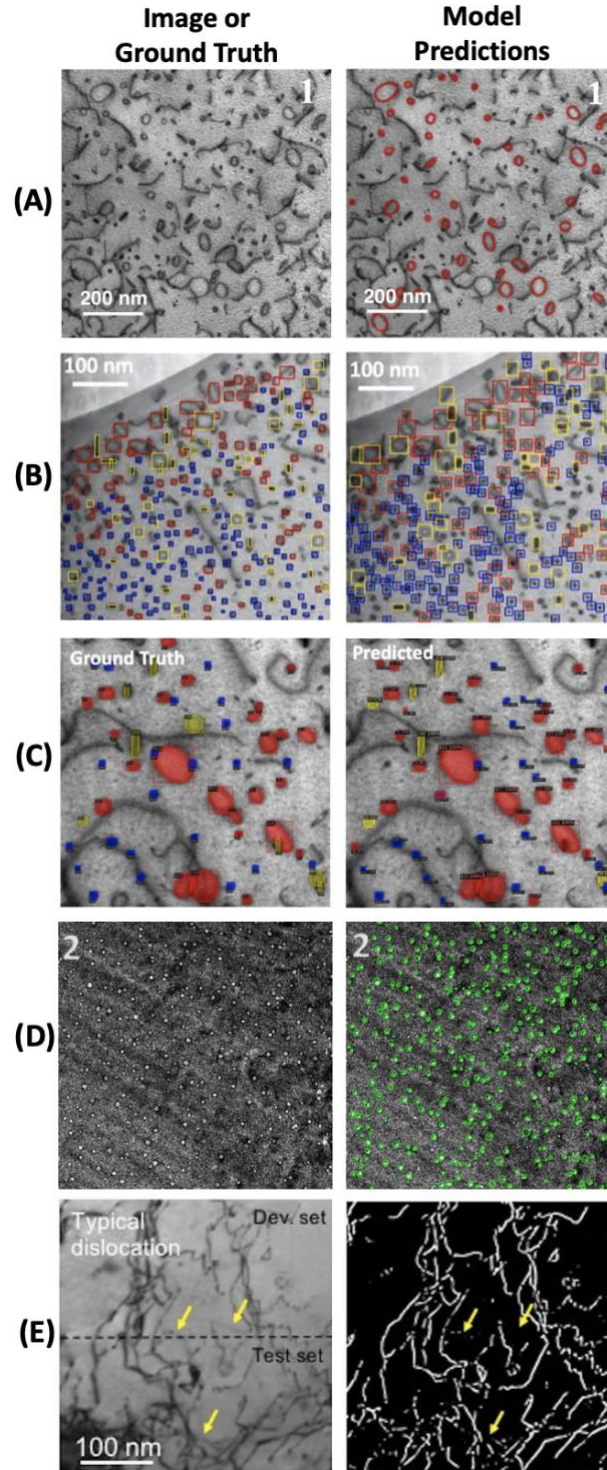


Figure 4. Survey of object detection studies for EM images in materials science. The left column of images denotes original micrographs or micrographs with ground truth labels. The right column denotes corresponding model predictions. (A) Dislocation loops in FeCrAl alloy from Li et al. using a non-deep learning based cascade object detector and classifier CNN.[31] (B) $\langle 111 \rangle$, $\langle 100 \rangle$ and black dot defects in FeCrAl alloy from Shen et al. using Faster R-CNN.[37] (C) $\langle 111 \rangle$,

$\langle 100 \rangle$ and black dot defects in FeCrAl alloy from Jacobs et al. using Mask R-CNN.[40] (D) Voids in X-750 Ni superalloy from Anderson et al. using Faster R-CNN.[38] (E) Line dislocations in HT-9 martensitic steel from Roberts et al. using a custom U-Net model.[39] In panels (B) and (C), the red, yellow and blue annotations denote $\langle 111 \rangle$, $\langle 100 \rangle$ and black dot defects, respectively. All images were adapted from their corresponding references with permission.

2.2 Characterizing Particles in Electron Microscopy Images

Characterizing the sizes and shapes of microparticles and nanoparticles is important for applications ranging from designing feedstocks for additive manufacturing to understanding rates of catalytic reactions and electronic properties of quantum dots.[32,46] Object detection models are well-suited for performing instance segmentation of particles in EM images. A particularly noteworthy study from Cohn et al. used the Mask R-CNN model to perform instance segmentation on SEM images of gas-atomized nickel superalloy powders.[46] They calculated both the precision and recall for pixel segmentation accuracy and detection of particles, and found by 5-fold random cross validation average F1 scores of 0.976 and 0.862 for pixel segmentation and particle detection, respectively. This finding not only demonstrates the robust particle detection capabilities of the Mask R-CNN model as shown in **Figure 5A**, but also highlights the potential differences between examining what constitutes well-performing model accuracies at the pixel vs. **object** level. Cohn et al. also performed a series of more demanding tests on their particle detection model. The first test involved the detection of particle satellites, which are smaller particles agglomerated to the surface of large particles (see **Figure 5A**). Here, they found **a** cross validation satellite detection F1 score that was 0.610, notably lower than the 0.862 value for finding full particles. This lower value is expected given the higher difficulty of detecting the much smaller satellite particles, where the lower F1 score was attributed mainly to the highly subjective nature of what features constitute a satellite. A second test was performed on test images of powder samples of different materials than those used in the training set, such as Ti- and Al-Si-Mg-based powders, which not only differ in composition (thus altering their contrast in EM images), but also have different size and shape distributions than the Ni-based powders in the training data. Generally, the Mask R-CNN trained only on the Ni-based powders performed qualitatively very well on the out-of-domain test images by showing reasonable segmentation

masks on most particles. As expected, from a visual standpoint more errors are made on these out-of-domain images, particularly for powders with high concentrations of satellites. Overall, this work provides an important demonstration not only of the utility of Mask R-CNN for particle detection in EM images but also provides insight regarding the expected performance on test data outside of the initial training domain, i.e., that the model was able to reliably predict different particle shapes and simulated particles not present in the training data.

Regarding the detection of nanoparticles in TEM images, Oktay and Gurses used a mixture of deep learning and traditional computer vision methods to detect and segment individual Fe_3O_4 and SiO_2 -coated Fe_3O_4 nanoparticles, as shown in **Figure 5B**.^[47] They used a multiple output CNN (MO-CNN) model to detect the location of individual particles and segment the particle boundaries. The segmentation output was then fed into a Hough transform to calculate the sizes of each particle. In general, the combination of MO-CNN and Hough transform yielded high pixel-level accuracies, and a comparison test to segmentation performance from a U-Net model showed the MO-CNN plus Hough transform outperformed U-Net from the standpoint of accurately predicting the segmented nanoparticle sizes. A second example from Groschner et al. segmented the boundaries of Au and CdSe nanoparticles in HRTEM images using a U-Net model, as shown in **Figure 5C**.^[48] In this work, the U-Net model trained on HRTEM images of both CdSe and Au nanoparticle images produce a pixel-wise F1 score of 0.8. Interestingly, the model performed better at segmenting Au nanoparticles than CdSe nanoparticles, where pixel-wise F1 scores of 0.89 and 0.59 were obtained for Au and CdSe test images, respectively. The increased difficulty of segmenting CdSe over Au nanoparticles makes sense, as the lower atomic numbers of Cd and Se compared to Au make the CdSe particles more difficult to visually resolve in the HRTEM images, resulting in a lower signal to noise ratio for the CdSe images. As a helpful demonstration of the general power of deep learning-based methods, Groschner et al. also compared the performance of their U-Net segmentation model with a number of traditional computer vision segmentation methods and found that U-Net significantly outperformed all non-deep learning methods to such an extent that the use of such traditional methods should no longer be recommended, at least for this particular segmentation task. As a final example, Nartova et al. employed the Cascade R-CNN model to detect nanoparticles deposited on supports

used for heterogeneous catalysis.[49] Their model resulted in a particle detection F1 score of approximately 0.71. They found that the F1 score was improved when just considering the particles localized to the support surface, as opposed to also including an analysis of particles that were partially visible or whose intensity was partially occluded by other particles or the catalyst support. When these two types of particles were treated as separate classes, the model F1 score saw improvement to about 0.81 for detections of particles on the support surface.

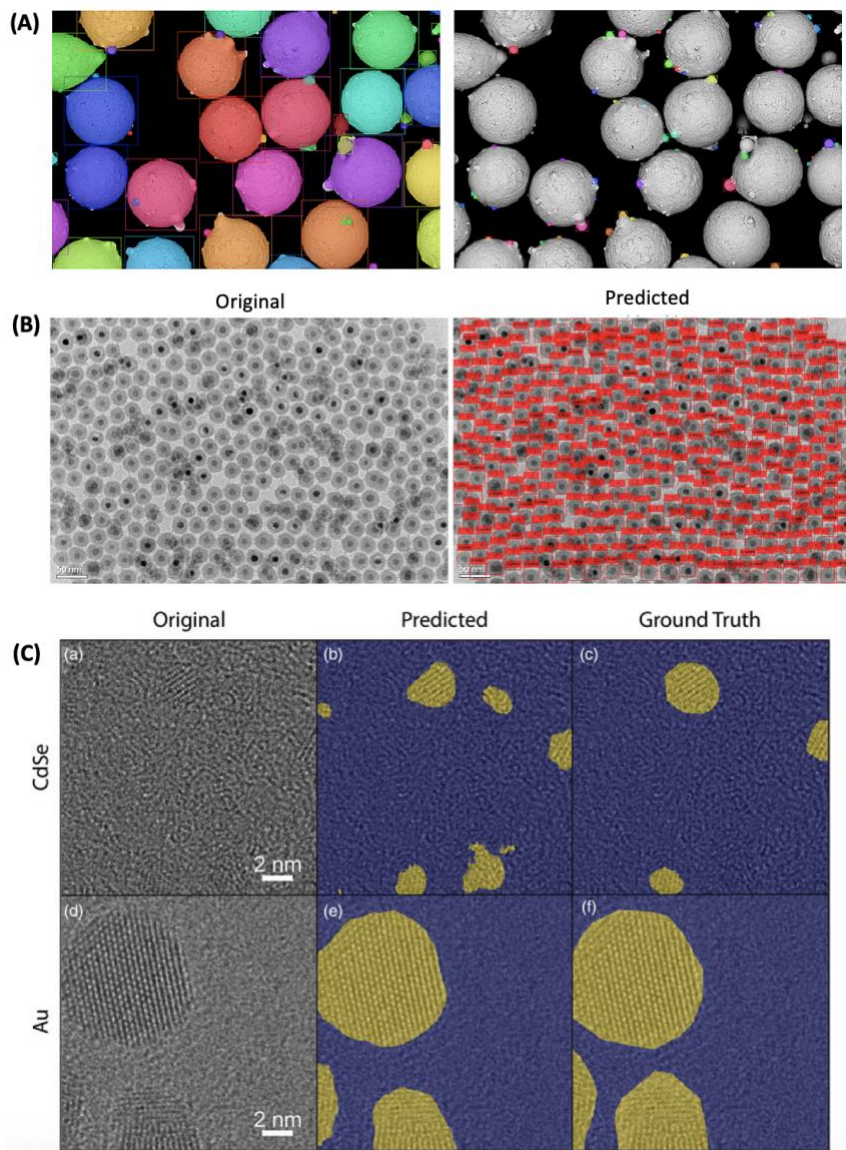


Figure 5. Some examples of object detection of micro and nanoparticles in SEM and TEM images. (A) Mask R-CNN for particle detection of gas-atomized Ni superalloy powders, where detected particle bounding boxes and segmentation masks are shown for all particles and satellites (left image) and highlighting the ability of the model to predict the satellites (right image).[46] (B) TEM micrograph of SiO_2 -coated Fe_3O_4 nanoparticles (left image) and the detection output marking

each found nanoparticle using an MO-CNN model (right image).[47] (C) HRTEM micrographs (left images) of CdSe (top) and Au (bottom) nanoparticles, with U-Net predicted segmentations (middle images) and ground truth segmentations (right images).[48] All images were adapted from their corresponding references with permission.

2.3 Finding Individual Atoms in STEM Images

Materials properties and functions are directly tied to the underlying micro, nano, and atomic structure of the material, where the finest degree of structural information one can obtain is knowing the position and elemental identity of every atom in a material. Given both the increased fidelity and quantity of available data from modern EM instruments as described in Section 1, it would be extremely helpful to develop object detection methods which provided automatic output of every atom position in STEM images. Toward this goal, seminal work by Ziatdinov et al. employed fully convolutional networks (FCNs, so named because they do not contain any fully connected dense layers) to detect individual atoms in STEM images of the two-dimensional materials graphene and $\text{Mo}_{1-x}\text{W}_x\text{Se}_2$. [50] The FCN used in this work employed an encoder-decoder architecture, very similar to U-Net, and was trained on simulated STEM data of pristine and defected graphene structures, and augmented variants thereof obtained using standard methods (e.g., flipping, rotation, noise addition, etc.). An example of the FCN model applied to an experimental STEM test image of graphene is shown in **Figure 6A**. The output produced by the FCN is a probability density map of a given pixel being an atom, where here red indicates high probability of an atom, progressing to blue and black for low probability. This probability density output from the final layer of the FCN model is then used as input to a Laplacian of Gaussian (LoG)-based blob detection algorithm, which yields the final segmented atom positions, as seen in **Figure 6A** (right image). The model was also successfully applied to finding atom positions in a $\text{Mo}_{1-x}\text{W}_x\text{Se}_2$ STEM image as well as detecting vacancy and dopant positions in graphene and $\text{Mo}_{1-x}\text{W}_x\text{Se}_2$, demonstrating it can accurately discern different element types (or a missing atom) based on intensity variations in the STEM images.

A drawback of the work by Ziatdinov et al. discussed above is the focus on hexagonal symmetry of 2D materials in the training data, limiting the potential domain of model applicability

to finding atoms in other material structure types or sample morphologies. In particular, as the model discussed above primarily learned atom type and position based on variations in image intensity, samples with varying thickness or more complex morphologies (e.g., nanoparticles), will produce additional challenges with atom type identification as the brightness of the atomic column depends on both sample thickness and element type. As a step toward building a more generalizable atom finding model, Ge and Xin have used a similar methodology as Ziatdinov et al., where here a U-Net model for atom detection is used followed by circular Hough transform to produce the final atom segmentations.[51] The key advancement put forth by Ge and Xin lies in the more diverse set of training data employed, effectively expanding the domain of model applicability. Here, Ge and Xin developed an extensive simulated STEM image database of 10,000 images containing a variety of noise levels, sample thicknesses, imaging conditions, crystal structures, imaging zone axis, and image field of view. The use of simulated STEM images is necessary for construction of a sufficiently large training database for training atom finding models like those of Ge and Xin, and the important role of synthetic data generation for training improved object detection models is discussed in more detail in Section 3.2. By qualitative visual inspection, the U-Net model and Hough transform method employed by Ge and Xin appears to accurately locate atom positions in materials with varying symmetries (e.g., cubic vs. hexagonal), and morphologies (e.g., bulk vs. samples with surfaces or extended defects).

As an extension of the above discussed studies, Wei et al. performed detailed benchmarking of the models originally fit by Ge and Xin and Ziatdinov et al.[52] Benchmark studies were performed by quantifying atom finding metrics such as precision, recall, and distance from correct atom position for [100]-oriented perovskite SrTiO_3 and monolayer WS_2 . Generally, it was found that the model Ziatdinov et al. slightly out-performed the model from Ge and Xin for these particular test cases, though both models failed on more challenging images, and it was found they failed for different reasons. It was observed that the model of Ziatdinov et al. has low recall (i.e., misses many atoms) if there is significant contrast variation in the image, and the model of Ge and Xin performs poorly if the image has large pixel sizes. Both models have trouble identifying atoms for cases in which there is significant overlap of intensity peaks. Interestingly, it was found that for many images, both models perform extremely well with recall

and precision scores nearing perfect performance, and for select images the models (particularly Ge and Xin) fail in such a manner that very few, if any, atoms are detected, indicating some underlying fragility in the applicability domain of both models. This work provides the first step toward the important goal of establishing a community repository of models and data, from which detailed benchmarking and iterative improvements to both the data and trained models can be made to facilitate the production of the best atom-finding model. Additional discussion of this important opportunity for the materials community is provided in Section 3.4.

The use of deep learning models to detect the position of each atom in STEM images enables the extraction of other useful properties, such as sample thickness and nanoparticle morphology. For example, Zhang et al. used a VGG16-based CNN to determine the sample thickness of SrTiO_3 samples based on the Sr column intensities, with the capability to provide thickness predictions within about one unit cell across the provided image.[53] Ragone et al. used CNN models to calculate the atomic column heights of gold nanoparticles imaged at the atomic scale using HRTEM, enabling detailed inference of the overall nanoparticle morphology, as shown in **Figure 6B**. [54] Similar to other atom detection studies, Ragone et al. leveraged simulated STEM image data, and the trained CNN model was able to translate from simulated images to reasonable predictions of atomic column heights for experimentally imaged nanoparticles. Lee et al. built upon the work of Ziatdinov et al. and used an FCN residual U-Net model to detect and classify individual point defects in the 2D material $\text{WSe}_{2-2x}\text{Te}_{2x}$. [55] As shown in **Figure 6C**, their network is capable of identifying different types of individual points defects in experimental STEM images, and can even discern between single and double Se vacancies based on the contrast differences of the Se site column. Moreover, by employing class averaged images with the FCN defect predictions, information of local lattice expansion and contraction (as a result of the point defects) and the resulting strain fields was obtained. This work demonstrates the utility of deep learning methods to help produce new insight on atomic-scale materials structure and its coupling to materials properties (here, variations in elastic properties from point defects).

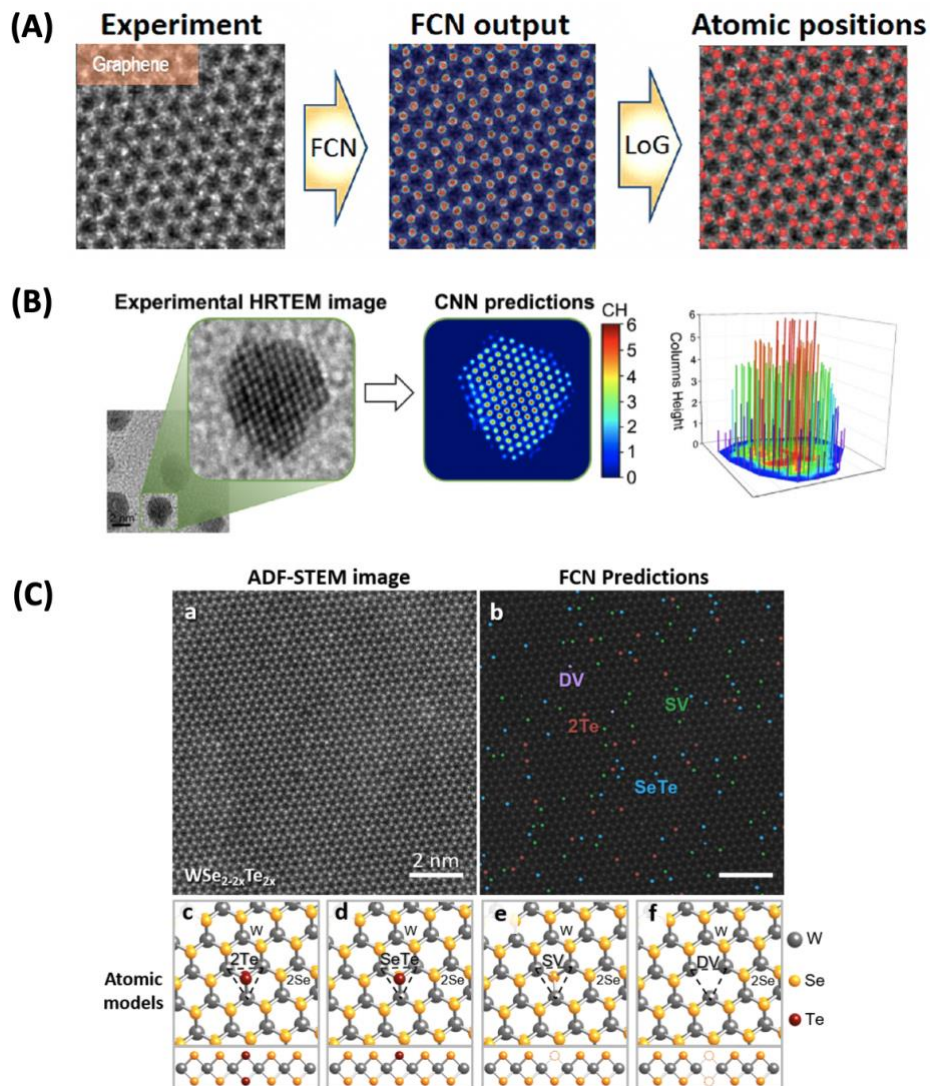


Figure 6. Application examples of using deep learning methods to detect individual atoms in STEM images. (A) Experimental STEM image of graphene (left image), with atomic positions predicted by the fully convolutional network (FCN) (middle image), and final, refined atomic positions marked using the Laplacian of Gaussian (LoG) blob detection method on the FCN output (right image).[50] (B) Experimental HRTEM image of an individual gold nanoparticle that is about 2.8 nm in size (left image), with corresponding CNN predictions of the atomic column heights for each column of gold atoms comprising the nanoparticle (center image), and the inferred 3D morphology of the nanoparticle based on the CNN-extracted column heights (right image).[54] (C) Annular dark field (ADF) STEM image of the 2D material WSe_{2-x}Te_{2x} (upper left), and different detected defects using an FCN residual U-Net model, where single Se vacancy (SV), double Se vacancy (DV), Se on a Te site (antisite, SeTe), and 2 Te defects on the Se site (2Te) are shown as green, pink, blue and red dots, respectively (right upper image). The corresponding atomic structures of the detected defect types are shown along the bottom.[55] All images were adapted from their corresponding references with permission.

2.4 Tracking Detected Objects in Electron Microscopy Video

All of the above discussed applications of object detection in EM images have centered around automatic analysis of individual or static images. As discussed in Section 1, the capability of drastically increased EM image data generation now includes the generation of full videos obtained from *in situ* EM experiments. Such *in situ* experiments can provide information of dynamic materials behavior which can be used to better understand the evolution of material structure and properties.[32,56] For example, the coupling of high temperature defect migration to grain boundaries with the resulting mechanical properties, or the nucleation and growth of voids during material irradiation, thus determining materials swelling which can then inform operational conditions such as temperature or radiation dose. As videos are a time-resolved sequence of images, objects in each frame of a video can be detected using the same deep learning-based object detection tools described throughout this section. Two challenges which emerge from having to perform object detection on a large number of images comprising a video are (1) long object detection model inference times, making on-the-fly detection and tracking difficult, and (2) linking detected objects between frames via a tracking algorithm. There are also challenges with acquiring stable, low noise TEM video suitable for reliable object detection. The You Only Look Once (YOLO)[57] model has emerged as a leading model for performing object detection and tracking in videos. Contrary to two-stage object detection models such as Faster R-CNN and Mask R-CNN, YOLO is a single stage detection framework which enables both faster training and inference times compared to the R-CNN models. YOLO provides only bounding box information (similar to Faster R-CNN), so additional image analysis on detected objects is required if more detailed geometrical information is desired. Application of tracking detected objects in EM images is presently in the nascent stages, though some notable recent studies have emerged.

Shen et al. was the first study of its kind to employ the YOLO object detection model to demonstrate real-time identification and tracking of defect loops in FeCrAl alloys for sets of TEM images extracted from video.[36] The YOLO model was shown to be extremely effective at detecting dislocation loops, with very high F1 scores in the range of 0.83-0.93 depending on the test image examined. As the images used in this work were the result of *in situ* irradiation, the

total irradiation dose (as measured in displacements per atom, or dpa) increases with time. As irradiation damage increases, new dislocation loops are nucleated in the microstructure, and previously present loops can grow, migrate, and coalesce. Shen et al. were able to detect and track the dislocation loops throughout the *in situ* TEM video and calculate key materials properties like defect density and defect size evolution as a function of dpa (i.e., as a function of time under irradiation), with YOLO-detected results as a function of irradiation dose compared with the ground truth calculation displayed in **Figure 7**. The results of defect density and defect size as a function of dpa shown in **Figure 7** demonstrate the YOLO model can accurately reproduce both qualitative and quantitative trends of defect size and density changes with irradiation. Similar work as Shen et al. has been reported by Sainju et al., who were interested in tracking the radiation defect dynamics of defect clusters (e.g., dislocation loops and stacking-fault tetrahedra) in pure nickel metal which has undergone krypton ion irradiation. They use a multiple object-tracking computer vision model to detect and track the defect clusters over time at high temperature, revealing the expected lifetime of such clusters, which was found to decay monotonically (i.e., the defects are disappearing) in less than 4 seconds' time, providing new insights of irradiation-induced defect dynamics.[58]

The level of analysis made possible by models like YOLO not only provides understanding of application-specific materials properties such as hardening as a function of irradiation dose, but may unlock a deeper level of materials data analysis by virtue of efficiently analyzing large amounts of *in situ* TEM data, which was not previously possible by traditional analysis methods. For example, recent work by Nathaniel et al. performed manual analysis of defect density as a function of distance from grain boundaries in copper bicrystal samples, and characterized grain boundary defect absorption characteristics based on grain boundary type.[59] Similar studies to the work of Nathaniel et al. would likely be conducive to deep learning-based object detection and tracking analysis, such as application of YOLO to track defects in the vicinity of grain boundaries over time to assess the grain boundary sink strength. Finally, fast inference of the YOLO model offers the intriguing possibility of deploying a trained detection model and accompanying software on new TEM instruments for real-time detection and tracking of defects during irradiation experiments.

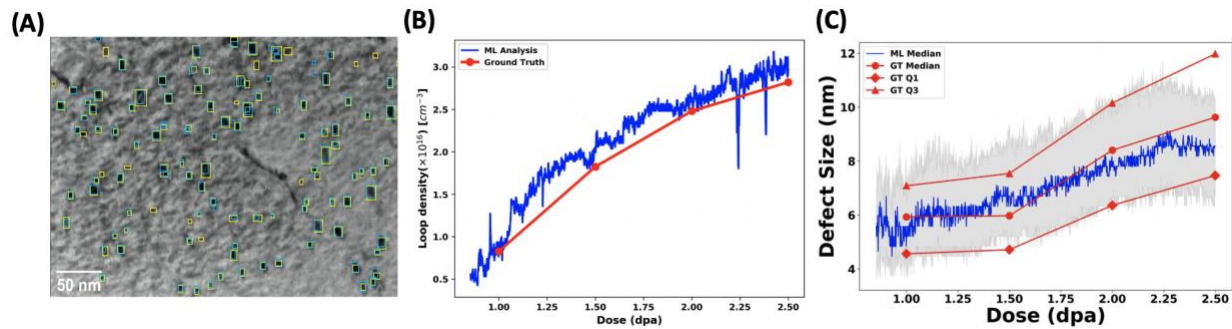


Figure 7. Materials-specific properties extracted from detecting and tracking defects using a YOLO model on *in situ* TEM video from the work of Shen et al.[36] (A) A single frame of the video showing ground truth (cyan) and YOLO predicted (yellow) bounding boxes. (B) The calculated defect density as a function of dose from the YOLO results (blue) compared to ground truth calculations (red). (C) The calculated average defect sizes from the YOLO results and subsequent geometry fitting (blue) compared to ground truth calculations (red). For the ground truth labels in (C), the triangle, circle and diamond data correspond to the third quartile, median, and first quartile, respectively, and the grey shaded region corresponds to the third (upper boundary) and first (lower boundary) quartile of defect size from the YOLO results. All images were adapted from Ref [36].

3 Opportunities and Challenges for Object Detection in Materials Science

3.1 Best Practices of Model Evaluation and Applicability

The studies described throughout Section 2 highlight application of a variety of object detection and segmentation techniques that were found to be successful in characterizing and quantifying a range of objects of interest in EM images. These and other similar studies tend to have two factors in common: (1) model performance is typically evaluated for a limited set of testing data, potentially giving a false impression of model performance quality on new unseen data and (2) model quality is assessed using primarily statistics-based metrics (e.g., precision, recall, F1 score), as opposed to materials-centric metrics which seek to connect the model results to the prediction of a material property or performance metric (e.g., material hardening or

swelling). Now that application of **several** object detection and segmentation models is becoming increasingly straightforward, the arc of research using these tools to characterize EM images will likely shift from the demonstration these models broadly work for characterizing particular EM image features to more detailed, meaningful materials domain-based assessment of model performance.

It is understandable that in initial studies, model performance was evaluated for a limited set of testing data. This is likely the result of both computational cost of training and evaluating object detection models, and also the fact that labeled data for EM image analysis tends to be limited in quantity and quality, where well-curated databases can take years to develop. Given the increased computational power of new graphical processing units (GPUs), together with more widespread availability of computing resources such as Google Colab and Amazon Web Services, more thorough evaluation of model fit statistics is quickly becoming a reality. For example, the work from Jacobs et al. discussed in Section 2.1 above performed both random and targeted group cross validation using Mask R-CNN models. Similar to studies employing traditional machine learning methods for materials property prediction,[60–62] they found that the nature of the training and test data can have a dramatic impact on the model performance. As a concrete example, through random cross validation splits of train and test images, they found the model predicted errors in average defect size can range from about 2 to 11%, while density errors ranged from about 8 to 22%. Further, as shown in **Figure 8**, they found the model performance by defect type was quite sensitive to dataset size and cross validation type, where random leave out cross validation tended to result in improved performance compared to targeted group cross validation, where images were separated based on factors such as alloy type and extent of irradiation.

While results like those shown in **Figure 8** are not surprising, there are two takeaway messages that are worth keeping in mind. First, baseline model performance on any metric should be evaluated at least by some kind of random cross validation technique.[40,46] This way, key metrics like object size distribution can be quoted with an average, standard deviation, and range of values. Such baseline performance measures give potential users of the model a sense of the scale of errors to expect on test images that are, to a first approximation, drawn from the

same domain as the training data. More targeted, grouped cross validation is useful for understanding weak areas in the model domain of applicability. In general, the test used to evaluate the model should mirror how the model is expected to be used in production, where random (grouped) cross validation performance is qualitatively reflective of test images drawn from inside (outside) the training image domain.

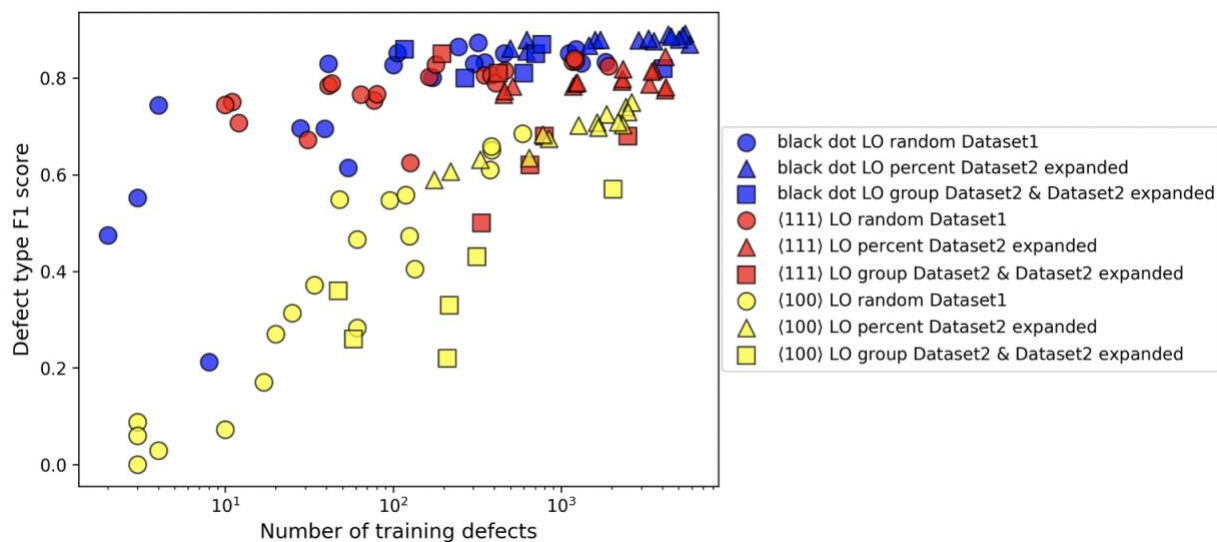


Figure 8. Mask R-CNN F1 score of identifying different defect types in EM images of FeCrAl alloys as a function of dataset size and type of cross validation split. Blue, red and yellow shapes denote black dot, <111> and <100> dislocation loop defect types, respectively. The circle and triangle shapes are different types of random cross validation assessment, while the squares denote targeted grouped cross validation. Figure adapted with permission from Ref [40].

Model quality is often assessed using primarily statistics-based metrics such as precision and recall. This is a sensible starting point to gauge general model performance, but such statistics-based metrics alone are not sufficient to understand how model performance correlates to materials application-specific quantities. For example, for detecting and segmenting nanoparticles, one statistics-based metric reported may be pixel accuracy. However, it is difficult to connect how, for example, a 90% pixel classification accuracy for detecting nanoparticle vs. background in an image can be used to form desired connections with the underlying materials properties. An example of such a property from nanoparticle segmentation may be quantifying the distribution of nanoparticle sizes and shapes and relating these size and shape distributions to the resulting catalytic activity for a particular **chemical** reaction. For materials science

applications of object detection, gauging the usefulness of a model will be contingent on not only analyzing basic detection statistics, but forming connections on how model performance on these basic detection statistics corresponds to prediction accuracy of the resulting material properties and performance for a particular application.

3.2 Generating synthetic data for improved model development

The performance of object detection models can generally only be as good as the input data used to train the model, and acquiring, curating and annotating a large amount of high-quality training image data is a time- and resource-intensive task. While most object detection model training uses traditional methods of image augmentation such as random flipping, rotations, cropping, resizing, and contrast or brightness adjustments of images, the inclusion of synthetic data is a highly promising method toward expanding the amount and domain of image training data. Currently, the generation of synthetic data can broadly be placed into one of three groups: (1) Physics-based simulation of image features, (2) deep learning-based generative models such as generative adversarial networks (GANs),[63] and (3) direct rendering. The use of synthetic data to improve model training has already found many notable successes in the broader scientific and medical communities. For example, GANs have been successfully used to simulate computed tomography and X-ray images of various organs, improving classification models of cancer detection.[64] In addition, the autonomous vehicle company Waymo uses a direct rendering computer program called Simulation City to simulate a diverse array of events and situations likely to be encountered by an autonomous vehicle, and also employs specialized GAN models to create synthetic camera and LiDAR sensor data. In this case, the scalability of synthetic data generation is a key advantage, and as of 2020, Waymo's database contains 15 billion miles of simulated driving data, compared to 20 million miles of real driving.[65,66] Therefore, in a broad sense synthetic data will likely play a key role in the coming years for developing the most accurate and reliable object detection models for a range of applications.

For generating synthetic EM images for materials science applications, one promising avenue is the use of physics-based modeling. This modeling may take the form of multislice

simulations, where the interaction of an electron beam in an electron microscope with the atomic structure of a material is modeled to produce a simulated STEM image.[67] The multislice method has been successful in simulating a number of materials at the atomic scale, for example pristine perovskite SrTiO_3 and Si with a dislocation core,[53,68] high resolution TEM images of individual carbon nanotubes,[69] and III-V semiconductor quantum well heterostructures,[70] as shown in **Figure 9A**. Multislice simulated images such as these depict a complete atomic-scale STEM image, and can be used as input for segmentation and object detection tasks such as those tasked with finding every atom in an image, as discussed in Section 2.3. Multislice simulations have also been used to create synthetic images of helium voids (sometimes called cavities), with a representative example of a synthetic underfocused void shown in **Figure 9B** from work of Yao et al.[71] Helium voids were the subject of object detection by Anderson et al.[38] as discussed in Section 2.1, and inclusion of synthetic cavities can be used to augment existing experimental databases to expand the model training domain to include different size distributions, focusing and imaging conditions, and noise levels to improve model training. Initial work in this space from Field et al. has combined simulated voids onto experimental images containing real voids in an effort to improve object detection model training.[72] A present challenge is how to best integrate individual simulated synthetic voids with real or simulated background EM images (see **Figure 9C**) with reliable, automatic annotation for fast construction of larger synthetic training datasets, and present work is ongoing toward addressing this challenge and testing the impact of synthetic data on model performance.

In addition to physics-based simulation, another promising avenue toward effective synthetic data generation is through the use of GANs. Briefly, GANs consist of two neural networks, a generator and a discriminator. The generator network seeks to create new synthetic images which are of sufficient quality as to fool the discriminator, which seeks to classify true vs. synthetic images. In materials science, different types of GANs have been successfully used to create synthetic microstructural images for numerous materials. As a first example, Lee et al. employed three types of GAN models, namely a deep convolutional GAN (DCGAN),[73] a cycle-consistent GAN (cycle-GAN)[74] and a conditional GAN (pix2pix)[75] to generate synthetic optical microscopy and EM images of steel surfaces and different lithium ion battery electrode

morphologies.[76] While their DCGAN-generated images were qualitatively sensible in their visual features, the cycle-GAN generated images ended up being superior. An advantage to the trained cycle-GAN model is the ability to make generated EM images using optical microscopy images as training, and vice versa, which could aid in expanding the domain of training data used in subsequent object detection or segmentation tasks. As the cycle-GAN employs two sets of image types in the training data which contain similar visual features (e.g., images of zebras and horses have similar features but are different animals), one may use cycle-GAN and similar models to generate synthetic EM images of different imaging modes, e.g., secondary vs. backscatter SEM images, or underfocused vs. overfocused helium cavities in TEM images. As a second example, Ma et al. employed both simulated and synthetic data made using the pix2pix conditional GAN to augment their real data of polycrystalline iron microstructures.[77] In this case, the simulated data were too pristine and did not accurately reflect the imperfect features of real micrographs, and the use of the pix2pix GAN trained on real and simulated data resulted in synthetic microstructure images with more realistic features. Using the real and synthetic data, U-net segmentation models were trained and evaluated on real images. Initial tests showed best performance (measured as mean average precision (mAP)) with U-Net trained only on real data (mAP=0.585), poor performance when trained only on simulated data (mAP=0.112), and good but slightly degraded performance (compared to using only real data) when real and simulated data were used together with training (mAP=0.504). These findings make sense, considering the overly idealized nature of the simulated data. Given this, Ma et al. then used different fractions of real and simulated data to train a pix2pix GAN to turn the simulated images into more realistic synthetic images, and trained a new U-Net model this time combining real and synthetic data. They found that using just 35% of the real image dataset, combined with the synthetic data, was sufficient to yield the same level of performance (mAP=0.586) as the previous U-Net model trained using all of the real data, and which resulted in improved grain boundary segmentation as shown in **Figure 9D**. These findings demonstrate the power of synthetic data methods for data augmentation, lessening the need for large amounts of curated experimental data for segmentation model training. As a third and final example, Hsu et al. used a Wasserstein GAN (WGAN)[78] to generate synthetic 3D microstructures of porous solid oxide fuel cell anodes, with

an example shown in **Figure 9E**.^[79] These electrodes are chemically and structurally complex, consisting of interconnected grain structures of Ni, yttria-stabilized zirconia, and porous spaces. They found that WGAN-based synthetic images more closely resembled experimental images than corresponding simulated images based on a number of microstructure properties, including volume fraction, particle size, tortuosity factor, etc. This finding suggests, at least for this application, that WGAN-generated synthetic images have closer physical resemblance to real images than 3D simulated microstructures. These and similar studies highlight the power of using GANs in conjunction with, or totally in place of, real data for subsequent materials analysis. However, it is worth noting that one drawback to using GANs as to opposed to physics-based simulations is the need for some initial seed of training data, which must be sufficiently large to train a reliable, stable GAN to generate synthetic data.

A third method researchers are presently exploring for synthetic data generation is that of direct rendering of images. As a first example, Cid-Mejias et al. used the animation software Blender^[80] to create artificial SEM images containing different sizes and geometries of nanoparticles, e.g., cuboids, spherical particles, platelets, etc.^[81] This was accomplished by creating 3D representations of different types of nanoparticle shapes, and embedding them into a Blender scene comprising the objects, lights and camera. The camera was positioned to mimic the angle of an SEM detector, and the lighting was also made to mimic typical SEM bright field image contrast. Since the type and size of each nanoparticle placed in the synthetic image is known, such a synthetic dataset has the needed annotation information for use in object detection. Another study that employed Blender for direct rendering of synthetic images was Mill et al.^[82] In this work, synthetic microscopy images of SiO₂ and TiO₂ nanoparticles were created. Compared to the study from Cid-Mejias et al., the work by Mill et al. was able to automate the Blender scene generation process, streamlining and accelerating the task of synthetic data generation. U-Net models were separately trained on real and rendered synthetic nanoparticle images. Encouragingly, both U-Net models showed high quality results on experimental test data, and while the U-Net model trained on experimental data slightly outperformed the model trained on synthetic data, the model trained on synthetic data still displayed high performance metrics with pixel F1 scores of 0.923 and 0.930 for segmenting TiO₂ and SiO₂ nanoparticles,

respectively. As a final example, Trampert et al. rendered synthetic dense grain microstructures using Voronoi tessellations combined with traditional computer vision techniques to add low and high frequency noise components to yield more realistic-looking textured microstructures, two examples of which are shown in **Figure 9F**.^[83] Using about 5000 synthetic images, they trained a ResNet50-based U-Net model to perform segmentation of the grain boundaries. Using 12 experimental images as test data, the model trained solely using synthetic data resulted in high pixel-wise segmentation accuracy (about 99%), and visually provided visually reasonable segmentations of grain boundaries for the real images.

In addition to standard methods like transfer learning, data augmentation, and the more novel uses of synthetic data generation as discussed above, there are additional emerging methods of overcoming the scarcity of labeled data. Two such methods gaining traction are the use of single- or few-shot learning methods^[84] and the introduction of physics-based priors into object detection models. Few-shot learning may be a promising avenue for training new models using very few instances of new labeled data. For example, recent work from Akers et al.^[85] used few-shot learning approaches to successfully classify and segment different material phases comprising STEM-imaged heterostructures and nanoparticles. In this approach, a single STEM image containing regions of different material types (e.g., thin film SrTiO_3 deposited on Ge) is broken up into several sub-images (sometimes called chips or superpixels, and these chips may be on the order of 100×100 pixels in size), which are classified by a domain expert and the visual features are encoded into representative class prototypes using, e.g., a ResNet101 network. The classification and segmentation are performed at the superpixel level, and the work of Akers et al. demonstrated that reliable accuracy can be obtained by training a model on approximately 10 chips for each class (i.e., material) type in the STEM image. In addition to few-shot learning approaches, methods which seek to encode physical rules governing the structure of materials into their models can reduce the possible feature space the model must learn to accurately detect objects in an image, thus reducing the amount of required training data. As an example of this, the work of Fu et al. encoded object shape and symmetry rules as physical priors into the Faster R-CNN detection framework.^[86] They demonstrated the use of this physics-aware object detection framework on the problem of segmenting oriented fibers in fiber-reinforced

composites and dendrites in microstructural images of Ni-based superalloys. They found that the inclusion of the physics-based priors in the Faster R-CNN model resulted in improved performance with respect to detected pixel precision, recall, and F1 scores compared to a conventional Faster R-CNN model, suggesting that less labeled data may be needed when the physics-based priors were included in the model.

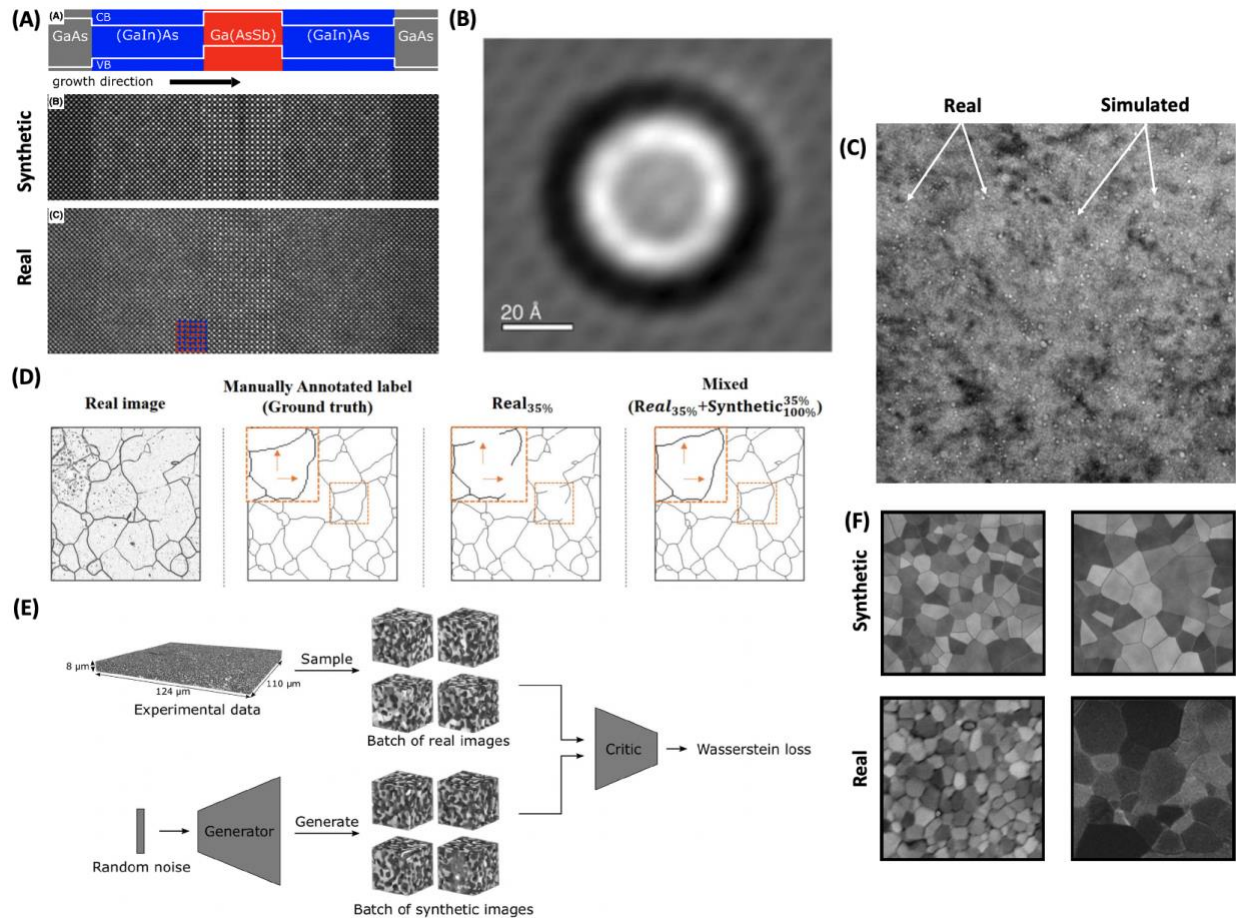


Figure 9. Survey of examples of synthetic image data generation for materials applications. (A) III-V semiconductor quantum well heterostructure showing sequence of materials and their respective band alignment (top), synthetic STEM image from multislice simulation (middle) and real STEM image (bottom).[70] (B) An individual underfocused helium void from multislice simulations.[71] (C) An experimental underfocused STEM image with simulated helium voids interspersed. A couple examples of real and simulated voids are marked with white arrows.[72] (D) A polycrystalline iron micrograph comparing the raw (left image) and annotated real image (middle left image) with a U-Net segmented image trained only using 35% of the real database (middle right image) and combination of real and synthetic data (right image).[77] (E) Schematic workflow of a WGAN to create synthetic 3D solid oxide fuel cell electrode microstructures,

showing examples of real and synthetic images.[79] (F) Examples of synthetic and real grain structures created using simulated Voronoi tessellations with added noise to provide more realistic textural features.[83] All images were adapted from their respective references with permission.

3.3 Influence of ground truth subjectivity and achieving community consensus labels

The complex nature of EM images often makes ground truth labeling and annotation of images not only a highly time intensive task but one that is challenging due to its subjectivity. Domain experts can disagree about, for example, whether a dislocation loop in an FeCrAl alloy STEM micrograph is correctly oriented and closed such that it should be annotated as a $\langle 111 \rangle$ loop or instead should be designated as a line dislocation, or whether a black dot defect is indeed a black dot or should be classified as a small $\langle 111 \rangle$ loop. Another example of subjective labeling was discussed in the work of Cohn et al., where labelers had to decide whether surface features of powder particles should be classified as satellites, or whether they were simply abnormal shaped protrusions from the host particle surface.[46] As a final example, Nartova et al. separately labeled and classified nanoparticles based on whether it appeared the particle was residing on the catalyst support surface and had well-defined edges, or whether the nanoparticle was occluded by the support or other nearby nanoparticles.[49]

One step toward understanding the impact of ground truth labeling on model performance is to perform round-robin tests where different researchers, preferably experienced domain experts, each separately label and annotate a set of images, and the distributions of key dataset features like number and size of each type of object are compared. Then, basic statistics such as the average, range, and standard deviation of the numbers and sizes of labels across all labelers can be established and compared against the model predictions and any previously developed ground truth labels. Informative examples of such round robin tests were performed in the work of Li et al.[31] and Roberts et al.,[39] with key results of their round robin tests shown in **Figure 10**. Plots like those shown in **Figure 10** are helpful to illustrate not just comparison of model performance against a ground truth, but, since the ground truth itself is subjective, what level of uncertainty may be present in the ground truth labels. Establishing

this baseline uncertainty is important, as it can help one establish whether model errors compared to a particular set of ground truth labels are significant, and whether the performance of a model may be deemed acceptable, for instance, if it is expected to perform more or less like a typical domain expert labeler.

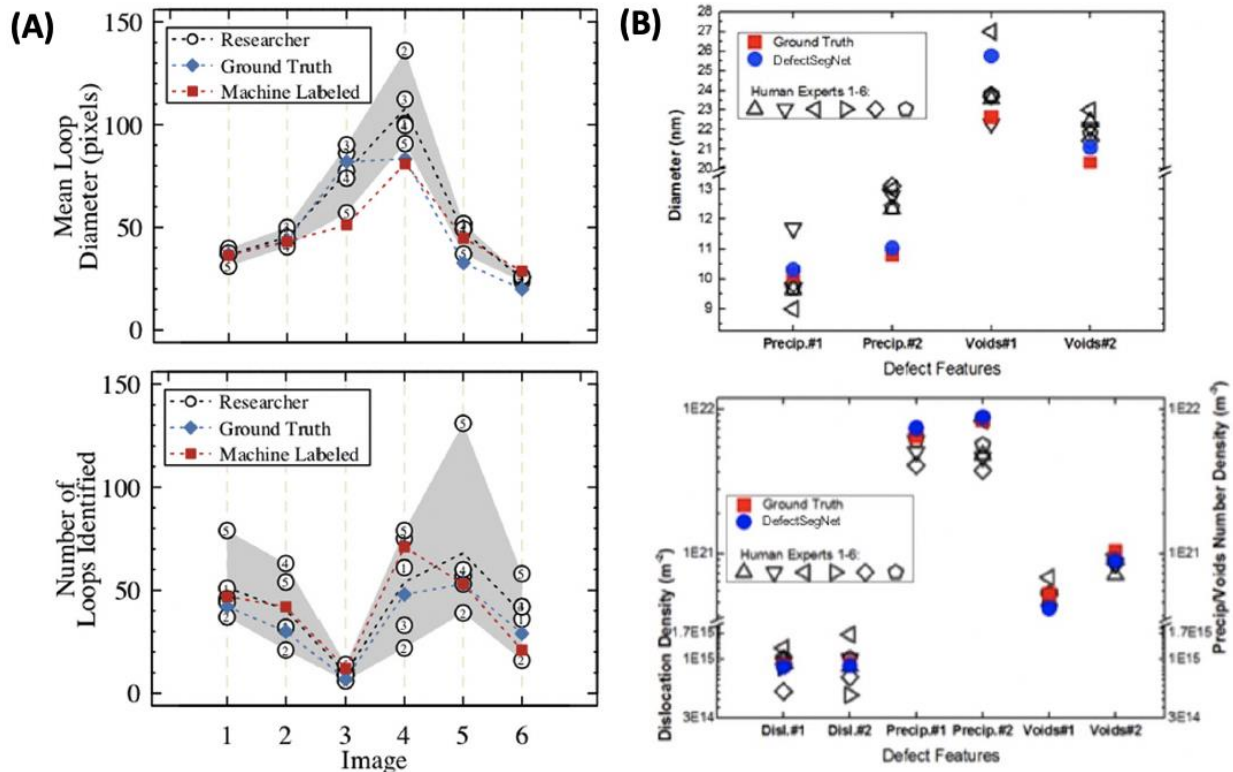


Figure 10. Examples of image labeling round robin tests establishing some basic statistics of ground truth subjectivity, and comparison with model predictions. (A) From the work of Li et al. identifying dislocation loops in irradiated FeCrAl alloys, plots of mean loop diameter (top) and number of identified loops (bottom) identified by five different labelers compared to the previously established ground truth (blue diamonds) and model predictions (red squares).[31] (B) From the work of Roberts et al. identifying dislocation lines, precipitates and voids, plots of the defect diameter (top) and defect density (bottom) identified by six different labelers compared to the established ground truth (red squares) and DefectSegNet model predictions (blue circles).[39]

At present, researchers taking part in individual studies may refine labels on their own data over time, perform round robin tests like those shown in **Figure 10**, and examine incorrect model predictions to better understand underlying features which may have caused the model

to misclassify or totally exclude certain features of an image. These tasks are all worthwhile to better understand model performance, but as object detection tools and methods continue to develop and are more widely adopted throughout the greater materials science community, it will be highly desirable to develop and maintain canonical benchmark datasets which can be iteratively refined to achieve community-consensus labeling. Datasets with community-consensus labels would be highly valuable in the pursuit of developing the most highly accurate and transferable object detection models. Opportunities for the community to establish tools for the improved development and dissemination of both data and models will be discussed more in Section 3.4.

Regarding establishing community-consensus labels and annotations for EM image datasets, there is an instructive study by Hattrick-Simpers et al. involving classifying X-ray diffraction (XRD) data from high throughput experiments which contains general concepts that could be applied to EM image object detection.[87] XRD is used to identify the crystal structure of materials, which can change as a function of temperature and composition. The motivation for the work of Hattrick-Simpers et al. centered around the fact that modern high-throughput experiments generate large amounts of XRD data which, analogous to the large amounts of EM image data discussed here, is too much for manual analysis by domain experts. It is of interest to train machine learning models to predict material structure from the XRD results to inform future experiments, however even domain expert analysis of the XRD data can be difficult and subjective, much like the annotation of objects in EM images. This case study by Hattrick-Simpers et al. compared the ability of domain experts and materials data science experts for the task of material phase identification of Nb-doped VO_2 as a function of temperature and Nb doping, where each peak in the XRD spectra were to be labeled as belonging to a tetragonal, monoclinic, or mixed phase. Consensus labels were determined by calculating the mode of the classification across all labelers, and the uncertainty of each label was quantified by computing the Shannon entropy value for the distribution of labels provided across all XRD data. Using these methods, it was found that the human labelers broadly agreed with each other regarding the correct peak labels, while there was some substantial disagreement between human labelers and the machine learning models. Moreover, the degree of disagreement was quantified, yielding useful insights

regarding which portions of the data resulted in the highest uncertainty. These and similar methods could be applied to EM image datasets used for training object detection models, where large-scale community voting results could be quantified to achieve consensus for the most accurate object labels. Furthermore, the uncertainty of community-wide object labels can be quantified, and labeled objects which show high uncertainty (from higher rate of disagreement between researchers) and are thus more subjective can be automatically flagged for further inspection and label refinement. It is worth noting here that such consensus assessment for dataset labeling could also be applied to evaluating a large set of trained object detection models from the community. It is foreseeable that researchers will train different types of object detection models (e.g. Mask R-CNN vs. vision transformer, or YOLO vs. Faster R-CNN), and seek to compare performance between the models on many datasets. Here, sets of predicted labels from many models could be compared with the community consensus labels, and uncertainty in the labels may also be quantified to provide a more detailed assessment both between model types and against the community-established ground truth.

3.4 Development and deployment of data and models for the community

The final step of machine learning model development is deployment of the model to a broader community. Many of the application-specific studies discussed throughout this review have sought to deploy their models to the public, and their efforts to do so are briefly summarized here. As an application-specific example, Nartova et al. developed the web service ParticlesNN, which enables segmentation of nanoparticles in an image uploaded by the user.[49] Cohn et al. have a software package called AMPIS (Additive Manufacturing Powder Instance Segmentation) for particle detection and segmentation in powder images using their trained Mask R-CNN model.[46] Ziatdinov et al. developed the python package AtomAI, a toolkit to leverage their deep CNNs for individual atom detection in STEM images.[88,89] Doty et al. packaged the few-shot learning methods described in the work of Akers et al.[85] into a web-based graphical user interface called pyCHIP, which lowers the barrier for classifying and segmenting material microstructures in STEM images using few-shot learning methods.[90] Finally, Jacobs et al.

packaged their trained Mask R-CNN model and provided a Google Colab notebook which facilitates straightforward inference and analysis of predicted dislocation loops and black spot defects on new test images without the need for any python coding.[40] There are also examples of additional open source toolkits well-equipped to perform a variety of object detection and segmentation tasks on EM images, such as PyJAMAS,[91] ZeroCostDL4Mic,[92] and ImageDataExtractor,[93] among others. It is worth noting that as the barrier required to perform meaningful object detection tasks has substantially reduced in the past few years, several companies and research institutions have developed software packages to aid in performing both traditional computer vision analysis and deep learning analysis of images, including semantic segmentation of objects in images. These tools include Reactiv IP's Smart Image Processing package,[94] Object Research Systems' Dragonfly package,[95] and EPFL's DeepImageJ package,[96] to name a few. As a final example, Theia Scientific is developing web-based object detection software used to run on electron microscopes to provide real-time object detection, quantification analysis and tracking in electron microscopy images.[97]

The above discussion of community-consensus labeled data and present methods of model deployment brings forth the broader question: How can materials science researchers interact with image data and models in a way which most productively advances the state of the art for the entire community? Development, refinement, and deployment of the best performing state-of-the-art object detection models will depend critically on the organization and accessibility of data and models. As an answer to the above question and a means toward accomplishing this vision, there is an ongoing initiative as part of the National Science Foundation's Cyberinfrastructure for Sustained Scientific Innovation (CSSI) to develop the *Foundry*, which will provide flexible, integrated, cloud-based management of machine learning projects in materials science, from organizing, publishing, and structuring data to hosting models to disseminating results that are machine and human accessible and reproducible in ways that support a networked materials innovation ecosystem.[98,99] While the Foundry will operate on data and machine learning models in a general way across disciplines and data and model modalities (e.g., tabular vs. image data, random forest vs. Mask R-CNN), it also holds the potential to revolutionize the use of object detection models for EM image and video analysis. The successful invocation of

Foundry would make updating existing datasets and re-training a previously hosted model straightforward, and enable cross-model and cross-dataset comparisons of predictions and uncertainties essentially turn-key. As a first step toward this goal, the work of Wei et al.[52] discussed in Section 2.3 involved the re-training and evaluation of atom finding CNNs from Ziatdinov et al. and Ge and Xin on datasets used in their work. With these datasets and models in hand, more detailed assessments of new model performance, benchmarked against previous models, and expansion of model domain by inclusion of new data will be made possible.

Given the rapid pace of development of new object detection models in the computer science field and their adoption and use in materials science applications, enabling the cross-model evaluations in a community-based infrastructure tool like the Foundry would be particularly worthwhile. The pace of model advancement has been fast, with a new model type releasing almost each year on average. More specifically, the Faster R-CNN,[19] Mask R-CNN,[20] and Cascade R-CNN[100] models were published in 2017, 2017, and 2018, respectively, while the newest models based on vision transformers (ViTs), the DETection TRANSformer (DETR) and Deformable DETR models were both released in 2020.[35,101] Generally, researchers in materials science have taken 3-4 years since the release of a new model to publish work using the model, where the Faster R-CNN model was used by Anderson et al. in 2020 and Shen et al. in 2021 (3 and 4 years after model release, respectively),[37,38] the Mask R-CNN model was used by Cohn et al. in 2021 and Jacobs et al. in 2021 (4 years after release in both cases), and the Cascade R-CNN model was used by Nartova et al. in 2022 (4 years after release).[40,46,49] A very recent study by Zhang et al. published in 2022 employed the Deformable DETR ViT model to detect particles in cryo-EM images, and appears to be the first such publication of using a ViT model to detect objects in EM images.[102] Moving forward, accelerating adoption of state-of-the-art object detection models from the computer science field to application on materials science problems and benchmarking performance against different R-CNN, ViT, and new state-of-the-art model types will be essential, and ideally performed on a series of canonical, community-consensus labeled datasets.

4 Summary

In this review, the current state of the use of deep learning-based object detection and segmentation methods for analyzing EM images is summarized. The motivation for the use of deep learning methods to analyze EM images has been necessitated by recent advancements in EM detector technology, enabling modern instruments to collect terabytes of data in a single session, including acquisition of real-time video from *in situ* experiments. Four key application spaces of object detection models in materials EM images were discussed: quantification of defects such as dislocation loops and voids in EM images relevant for nuclear materials applications, characterizing micro and nanoparticles for applications in additive manufacturing and catalysis, finding individual atoms in STEM images, thus informing atomic scale structure and properties, and tracking detected objects from *in situ* video, enabling new understanding of dynamic materials properties such as defect migration as a function of time and irradiation dose. Following these case study examples of the use of object detection in materials science, a series of topics related to ongoing opportunities and challenges for the continued development of these methods in materials research are examined. Of the many fruitful areas of continued research, here the focus was placed on topics such as enumerating best practices of model evaluation and applicability, where targeted cross validation schemes to assess model domain will play an important role in evaluating model application to new test data, and model evaluation beyond typical statistics-based metrics like pixel accuracy to more materials application focused metrics like predicted hardening or swelling to evaluate model performance can provide more informed assessment of model performance. The role of synthetic data to improve model domain and overall performance was discussed, including data generated from physical simulation, from generative models like GANs, and through direct rendering using animation tools like Blender. Looking to the future for what an ideal materials community interacting with codified EM image databases and trained object detection models might look like, key issues like the subjectivity of ground truth labels and their refinement through community-consensus labeling and uncertainty quantification is described, and a vision for a data- and model-sharing infrastructure which allows for the iterative refinement, assessment and dissemination of state-of-the-art models and methods becomes commonplace, enabling accelerated understanding of materials phenomena.

CRediT authorship contribution statement

Ryan Jacobs: Conceptualization, Writing- review & editing.

Acknowledgement

Support for R.J. was provided by Idaho National Laboratory as part of the Department of Energy (DOE) Office of Nuclear Energy, Nuclear Materials Discovery and Qualification Initiative (NMDQi). Additional support for R.J. was provided by the National Science Foundation (NSF) under NSF Award No. 1931298.

The author thanks Dane Morgan and Kevin Field for helpful discussions in preparing this review.

Data Availability

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

Annotated References

1. ****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). doi:10.1038/s41524-020-00392-6.** This study employed both simulated and synthetic data made using the pix2pix conditional GAN to augment their real data of polycrystalline iron microstructures. Using the real and synthetic data, U-net segmentation models were trained and evaluated on real images. They found that using just 35% of the real image dataset, combined with the synthetic data, was sufficient to yield the same level of performance as the previous U-Net model trained using all of the real data, highlighting the power of including synthetic data in model training.
2. ****R. Cohn, I. Anderson, T. Prost, J. Tiarks, E. White, E. Holm, Instance Segmentation for Direct Measurements of Satellites in Metal Powders and Automated Microstructural Characterization from Image Data, Jom. 73 (2021) 2159–2172. doi:10.1007/s11837-021-04713-y.** This study used the Mask R-CNN model to perform instance segmentation on SEM images of gas-atomized nickel superalloy powders. Five-fold random cross validation average F1 scores of 0.976 and 0.862 for pixel segmentation and particle detection, respectively. This finding not only demonstrates the robust particle detection capabilities of the Mask R-CNN model but also highlights the potential differences between examining what constitutes well-performing model accuracies at the pixel vs. detection level. Overall, this work provides an important demonstration not only of the utility of Mask R-CNN for particle detection in EM images but also provides insights regarding the expected performance on test data outside of the initial training domain.

3. **** R. Jacobs, M. Shen, Y. Liu, W. Hao, X. Li, R. He, J.R. 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 Reports Physical Science. 3 (2022) 100876. doi:10.1016/j.xcrp.2022.100876.** This study used refined and expanded versions of the dataset from previous works on object detection in FeCrAl alloys and used a Mask R-CNN model to realize a fully end-to-end deep learning approach to quantify the size and shape distributions and densities of dislocation loop and black spot defects. Detailed per-defect size and shape distributions and density calculations, which, when combined with a simplified dispersion hardening model, resulted in hardening predictions with errors of about 10-20 MPa, which is about 10% of total material hardening and within the margin of error of experiments. This result is a concrete example of automated object detection in EM images directly being used to inform materials property predictions.
4. **** M. Shen, G. Li, D. Wu, Y. Yaguchi, J.C. Haley, K.G. Field, D. Morgan, O. Ridge, O. Ridge, A deep learning based automatic defect analysis framework for In-situ TEM ion irradiations, Comput. Mater. Sci. 197 (2021) 110560. doi:10.1016/j.commatsci.2021.110560.** This was the first study of its kind to employ the YOLO object detection model to demonstrate real-time identification and tracking of defect loops in FeCrAl alloys for sets of TEM images extracted from video. The YOLO model was shown to be extremely effective at detecting dislocation loops, with very high F1 scores in the range of 0.83-0.93 depending on the test image examined. The authors were able to detect and track the dislocation loops throughout the *in situ* TEM video and calculate key materials properties like defect density and defect size evolution as a function of irradiation dose.
5. **** T. Hsu, W.K. Epting, H. Kim, H.W. Abernathy, G.A. Hackett, A.D. Rollett, P.A. Salvador, E.A. Holm, Microstructure Generation via Generative Adversarial Network for Heterogeneous, Topologically Complex 3D Materials, Jom. 73 (2021) 90–102. doi:10.1007/s11837-020-04484-y.** This study used a WGAN to generate synthetic 3D microstructures of porous solid oxide fuel cell anodes. They found that WGAN-based synthetic images more closely resembled experimental images than corresponding simulated images based on a number of microstructure properties, including volume fraction, particle size, tortuosity factor, etc. This finding suggests that WGAN-generated synthetic images have closer physical resemblance to real images than 3D simulated microstructures. These and similar studies highlight the power of using GANs in conjunction with, or totally in place of, real data for subsequent materials analysis.
6. ****J. Wei, B. Blaiszik, D. Morgan, P. Voyles, Benchmark tests of atom-locating CNN models with a consistent dataset, Microsc. Microanal. 27 (2021) 2518–2520. doi:10.1017/s1431927621008989.** This study performed detailed benchmarking of the atom finding models originally fit by Ge and Xin and Ziatdinov et al. This work provides the first step toward the important goal of establishing a community repository of models and data, from which detailed benchmarking and iterative improvements to both the data and trained models can be made to facilitate the production of the best atom-finding model.

References

- [1] C.A. Schneider, W.S. Rasband, K.W. Eliceiri, NIH Image to ImageJ: 25 years of image analysis, *Nat. Methods*. 9 (2012) 671–675. doi:10.1038/nmeth.2089.
- [2] T. Grant, A. Rohou, N. Grigorieff, cisTEM, user-friendly software for single-particle image processing, *Elife*. (2018). doi:https://doi.org/10.7554/eLife.35383.001.
- [3] S.R. Spurgeon, C. Ophus, L. Jones, A. Petford-Long, S. V. Kalinin, M.J. Olszta, R.E. Dunin-Borkowski, N. Salmon, K. Hattar, W.C.D. Yang, R. Sharma, Y. Du, A. Chiaramonti, H. Zheng, E.C. Buck, L. Kovarik, R.L. Penn, D. Li, X. Zhang, M. Murayama, M.L. Taheri, Towards data-driven next-generation transmission electron microscopy, *Nat. Mater*. 20 (2021) 274–279. doi:10.1038/s41563-020-00833-z.
- [4] Y. Jiang, Z. Chen, Y. Han, P. Deb, H. Gao, S. Xie, P. Purohit, M.W. Tate, J. Park, S.M. Gruner, V. Elser, D.A. Muller, Electron ptychography of 2D materials to deep sub-ångström resolution, *Nature*. 559 (2018) 343–349. doi:10.1038/s41586-018-0298-5.
- [5] D. Chatterjee, J. Wei, A. kvit, B. Bammes, B. Levin, R. Bilhorn, P. Voyles, An Ultrafast Direct Electron Camera for 4D STEM, *Microsc. Microanal*. 27 (2021) 1004–1006. doi:10.1017/s1431927621003809.
- [6] C. Ophus, Four-Dimensional Scanning Transmission Electron Microscopy (4D-STEM): From Scanning Nanodiffraction to Ptychography and Beyond, *Microsc. Microanal*. (2019) 563–582. doi:10.1017/S1431927619000497.
- [7] I. Goodfellow, Y. Bengio, A. Courville, *Deep Learning*, MIT Press, 2016.
- [8] A. Krizhevsky, I. Sutskever, G.E. Hinton, ImageNet Classification with Deep Convolutional Neural Networks, *Adv. Neural Inf. Process. Syst.* (2012) 1–9. doi:http://dx.doi.org/10.1016/j.protcy.2014.09.007.
- [9] Y. Lecun, Y. Bengio, G. Hinton, Deep learning, *Nature*. 521 (2015) 436–444. doi:10.1038/nature14539.
- [10] S. Grigorescu, B. Trasnea, T. Cocias, G. Macesanu, A survey of deep learning techniques for autonomous driving, *J. F. Robot*. 37 (2020) 362–386. doi:10.1002/rob.21918.
- [11] S. Li, W. Deng, Deep Facial Expression Recognition: A Survey, *IEEE Trans. Affect. Comput*. 3045 (2020) 1–20. doi:10.1109/TAFFC.2020.2981446.

- [12] J. Schrittwieser, I. Antonoglou, T. Hubert, K. Simonyan, L. Sifre, S. Schmitt, A. Guez, E. Lockhart, D. Hassabis, T. Graepel, T. Lillicrap, D. Silver, Mastering Atari, Go, chess and shogi by planning with a learned model, *Nature*. 588 (2020) 604–609.
doi:10.1038/s41586-020-03051-4.
- [13] A. Agrawal, A. Choudhary, Deep materials informatics: Applications of deep learning in materials science, *MRS Commun.* 9 (2019) 779–792. doi:10.1557/mrc.2019.73.
- [14] L. Liu, W. Ouyang, X. Wang, P. Fieguth, J. Chen, X. Liu, M. Pietikäinen, Deep Learning for Generic Object Detection: A Survey, *Int. J. Comput. Vis.* 128 (2020) 261–318.
doi:10.1007/s11263-019-01247-4.
- [15] K. He, X. Zhang, S. Ren, J. Sun, Deep Residual Learning for Image Recognition, in: 2016 IEEE Conf. Comput. Vis. Pattern Recognit., 2016: pp. 770–778.
doi:10.1109/CVPR.2016.90.
- [16] K. Simonyan, A. Zisserman, Very Deep Convolutional Networks for Large-Scale Image Recognition, *Int. Conf. Learn. Represent.* (2015) 1–14. doi:10.1016/j.infsof.2008.09.005.
- [17] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, L. Fei-Fei, ImageNet: A Large-Scale Hierarchical Image Database, 2009 IEEE Conf. Comput. Vis. Pattern Recognit. (2009).
- [18] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollar, C.L. Zitnick, Microsoft COCO: Common Objects in Context, *Eur. Conf. Comput. Vis.* (2014) 740–755.
- [19] S. Ren, K. He, R. Girshick, J. Sun, Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks, *IEEE Trans. Pattern Anal. Mach. Intell.* 39 (2017) 1137–1149.
doi:10.1109/TPAMI.2016.2577031.
- [20] K. He, G. Gkioxari, P. Dollar, R. Girshick, Mask R-CNN, *Int. Conf. Comput. Vis.* (2017).
- [21] Z.Q. Zhao, P. Zheng, S.T. Xu, X. Wu, Object Detection with Deep Learning: A Review, *IEEE Trans. Neural Networks Learn. Syst.* 30 (2019) 3212–3232.
doi:10.1109/TNNLS.2018.2876865.
- [22] O. Ronneberger, P. Fischer, T. Brox, U-Net : Convolutional Networks for Biomedical Image Segmentation, in: *Int. Conf. Med. Image Comput. Comput. Interv.*, 2015: pp. 234–241. doi:10.1007/978-3-319-24574-4.
- [23] J. Beal, E. Kim, E. Tzeng, D.H. Park, A. Zhai, D. Kislyuk, Toward Transformer-Based Object

- Detection, (2020). <http://arxiv.org/abs/2012.09958>.
- [24] 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*. 51A (2020) 1–22. doi:<https://doi.org/10.1007/s11661-020-06008-4>.
 - [25] G.B. Goh, N.O. Hodas, A. Vishnu, Deep learning for computational chemistry, *J. Comput. Chem.* 38 (2017) 1291–1307. doi:10.1002/jcc.24764.
 - [26] D.M. Dimiduk, E.A. Holm, S.R. Niezgoda, Perspectives on the Impact of Machine Learning, Deep Learning, and Artificial Intelligence on Materials, Processes, and Structures Engineering, *Integr. Mater. Manuf. Innov.* 7 (2018) 157–172. doi:10.1007/s40192-018-0117-8.
 - [27] W. Nash, T. Drummond, N. Birbilis, A review of deep learning in the study of materials degradation, *Npj Mater. Degrad.* 2 (2018) 1–12. doi:10.1038/s41529-018-0058-x.
 - [28] D. Morgan, R. Jacobs, Opportunities and Challenges for Machine Learning in Materials Science, *Annu. Rev. Mater. Res.* 50 (2020) 71–103. doi:10.1146/annurev-matsci-070218-010015.
 - [29] K. Choudhary, B. DeCost, C. Chen, A. Jain, F. Tavazza, R. Cohn, C. Woo Park, A. Choudhary, A. Agrawal, S.J.L. Billinge, E. Holm, S.P. Ong, C. Wolverton, Recent advances and applications of deep learning methods in materials science, *Npj Comput. Mater.* 8 (2022). doi:<https://doi.org/10.1038/s41524-022-00734-6>.
 - [30] K.P. Treder, C. Huang, J.S. Kim, A.I. Kirkland, Applications of deep learning in electron microscopy, *Microscopy*. 71 (2022) 100–115.
 - [31] W. Li, K.G. Field, D. Morgan, Automated defect analysis in electron microscopic images, *Npj Comput. Mater.* 4 (2018) 1–9. doi:10.1038/s41524-018-0093-8.
 - [32] H. Zheng, X. Lu, K. He, In situ transmission electron microscopy and artificial intelligence enabled data analytics for energy materials, *J. Energy Chem.* 68 (2022) 454–493. doi:10.1016/j.jechem.2021.12.001.
 - [33] M. Ge, F. Su, Z. Zhao, D. Su, Deep learning analysis on microscopic imaging in materials science, *Mater. Today Nano.* 11 (2020). doi:10.1016/j.mtnano.2020.100087.

- [34] L. Liu, W. Ouyang, X. Wang, P. Fieguth, J. Chen, X. Liu, M. Pietikäinen, Deep Learning for Generic Object Detection: A Survey, (2018).
- [35] X. Zhu, W. Su, L. Lu, B. Li, X. Wang, J. Dai, Deformable DETR: Deformable Transformers for End-to-End Object Detection, ArXiv:2010.04159. (2020) 1–16.
<http://arxiv.org/abs/2010.04159>.
- [36] M. Shen, G. Li, D. Wu, Y. Yaguchi, J.C. Haley, K.G. Field, D. Morgan, O. Ridge, O. Ridge, A deep learning based automatic defect analysis framework for In-situ TEM ion irradiations, *Comput. Mater. Sci.* 197 (2021) 110560.
[doi:10.1016/j.commatsci.2021.110560](https://doi.org/10.1016/j.commatsci.2021.110560).
- [37] M. Shen, G. Li, D. Wu, Y. Liu, J. Greaves, W. Hao, N.J. Krakauer, L. Krudy, J. Perez, V. Srrenivasan, B. Sanchez, O. Torres-Velazquez, 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>.
- [38] 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.
[doi:10.1016/j.ultramic.2020.113068](https://doi.org/10.1016/j.ultramic.2020.113068).
- [39] 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).
[doi:10.1038/s41598-019-49105-0](https://doi.org/10.1038/s41598-019-49105-0).
- [40] 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 Reports Phys. Sci.* (2022) 100876. [doi:10.1016/j.xcrp.2022.100876](https://doi.org/10.1016/j.xcrp.2022.100876).
- [41] A. Seeger, J. Diehl, S. Mader, H. Rebstock, Work-hardening and work-softening of face-centred cubic metal crystals, *Philos. Mag.* 2 (1957) 323–350.
[doi:10.1080/14786435708243823](https://doi.org/10.1080/14786435708243823).
- [42] D. Morgan, G. Pilania, A. Couet, B.P. Uberuaga, C. Sun, J. Li, Machine learning in nuclear materials research, *Curr. Opin. Solid State Mater. Sci.* 26 (2022) 100975.
[doi:10.1016/j.cossms.2021.100975](https://doi.org/10.1016/j.cossms.2021.100975).

- [43] K.G. Field, X. Hu, K.C. Littrell, Y. Yamamoto, L. Snead, Radiation tolerance of neutron-irradiated model Fe-Cr-Al alloys, *J. Nucl. Mater.* 465 (2015) 746–755. doi:10.1016/j.jnucmat.2015.06.023.
- [44] K.G. Field, S.A. Briggs, X. Hu, Y. Yamamoto, R.H. Howard, K. Sridharan, Heterogeneous dislocation loop formation near grain boundaries in a neutron-irradiated commercial FeCrAl alloy, *J. Nucl. Mater.* (2017). doi:10.1016/j.jnucmat.2016.10.050.
- [45] K.G. Field, S.A. Briggs, K. Sridharan, Y. Yamamoto, R.H. Howard, Dislocation loop formation in model FeCrAl alloys after neutron irradiation below 1 dpa, *J. Nucl. Mater.* (2017). doi:10.1016/j.jnucmat.2017.07.061.
- [46] R. Cohn, I. Anderson, T. Prost, J. Tiarks, E. White, E. Holm, Instance Segmentation for Direct Measurements of Satellites in Metal Powders and Automated Microstructural Characterization from Image Data, *Jom.* 73 (2021) 2159–2172. doi:10.1007/s11837-021-04713-y.
- [47] A.B. Oktay, A. Gurses, Automatic detection, localization and segmentation of nano-particles with deep learning in microscopy images, *Micron.* 120 (2019) 113–119. doi:10.1016/j.micron.2019.02.009.
- [48] C.K. Groschner, C. Choi, M.C. Scott, Machine Learning Pipeline for Segmentation and Defect Identification from High-Resolution Transmission Electron Microscopy Data, *Microsc. Microanal.* 27 (2021) 549–556. doi:10.1017/S1431927621000386.
- [49] A. V Nartova, M.Y. Mashukov, R.R. Astakhov, V.Y. Kudinov, Particle Recognition on Transmission Electron Microscopy, *Catalysts.* 12 (2022).
- [50] M. Ziatdinov, O. Dyck, A. Maksov, X. Li, X. Sang, K. Xiao, R.R. Unocic, R. Vasudevan, S. Jesse, S. V. Kalinin, Deep Learning of Atomically Resolved Scanning Transmission Electron Microscopy Images: Chemical Identification and Tracking Local Transformations, *ACS Nano.* 11 (2017) 12742–12752. doi:10.1021/acsnano.7b07504.
- [51] M. Ge, H.L. Xin, Deep Learning Based Atom Segmentation and Noise and Missing-Wedge Reduction for Electron Tomography, *Microsc. Microanal.* 24 (2018) 504–505. doi:10.1017/s143192761800301x.
- [52] J. Wei, B. Blaiszik, D. Morgan, P. Voyles, Benchmark tests of atom-locating CNN models

- with a consistent dataset, *Microsc. Microanal.* 27 (2021) 2518–2520.
doi:10.1017/s1431927621008989.
- [53] C. Zhang, J. Feng, L.R. DaCosta, P.M. Voyles, Atomic resolution convergent beam electron diffraction analysis using convolutional neural networks, *Ultramicroscopy*. 210 (2020) 112921. doi:10.1016/j.ultramic.2019.112921.
 - [54] M. Ragone, V. Yurkiv, B. Song, A. Ramsubramanian, R. Shahbazian-Yassar, F. Mashayek, Atomic column heights detection in metallic nanoparticles using deep convolutional learning, *Comput. Mater. Sci.* 180 (2020) 109722. doi:10.1016/j.commatsci.2020.109722.
 - [55] C.H. Lee, A. Khan, D. Luo, T.P. Santos, C. Shi, B.E. Janicek, S. Kang, W. Zhu, N.A. Sobh, A. Schleife, B.K. Clark, P.Y. Huang, Deep learning enabled strain mapping of single-atom defects in two-dimensional transition metal dichalcogenides with sub-picometer precision, *Nano Lett.* 20 (2020) 3369–3377. doi:10.1021/acs.nanolett.0c00269.
 - [56] M.L. Taheri, E.A. Stach, I. Arslan, P.A. Crozier, B.C. Kabius, T. LaGrange, A.M. Minor, S. Takeda, M. Tanase, J.B. Wagner, R. Sharma, Current status and future directions for in situ transmission electron microscopy, *Ultramicroscopy*. 170 (2016) 86–95.
doi:10.1016/j.ultramic.2016.08.007.
 - [57] J. Redmon, S. Divvala, R. Girshick, A. Farhadi, You Only Look Once : Unified , Real-Time Object Detection, in: *IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016: pp. 779–788.
 - [58] R. Sainju, W.-Y. Chen, S. Schaefer, G. Roberts, M. Toloczko, D. Edwards, M. Li, Y. Zhu, Deep Learning-based Computer Vision for Radiation Defect Analysis: from Static Defect Segmentation to Dynamic Defect Tracking, *Microsc. Microanal.* 27 (2021) 1464–1465.
doi:10.1017/s1431927621005419.
 - [59] J.E. Nathaniel, P.K. Suri, E.M. Hopkins, J. Wen, P. Baldo, M. Kirk, M.L. Taheri, Grain boundary strain as a determinant of localized sink efficiency, *Acta Mater.* 226 (2022) 117624. doi:10.1016/j.actamat.2022.117624.
 - [60] V. Stanev, C. Oses, A.G. Kusne, E. Rodriguez, J. Paglione, S. Curtarolo, I. Takeuchi, Machine learning modeling of superconducting critical temperature, *Npj Comput. Mater.* 4 (2018). doi:10.1038/s41524-018-0085-8.
 - [61] H.-J. Lu, N. Zou, R. Jacobs, B. Afflerbach, X.-G. Lu, D. Morgan, Error assessment and

- optimal cross-validation approaches in machine learning applied to impurity diffusion, *Comput. Mater. Sci.* 169 (2019). doi:10.1016/j.commatsci.2019.06.010.
- [62] L. Ward, S.C.O. Keeffe, J. Stevick, G.R. Jelbert, M. Aykol, C. Wolverton, A machine learning approach for engineering bulk metallic glass alloys, *Acta Mater.* 159 (2018) 102–111. doi:10.1016/j.actamat.2018.08.002.
- [63] I.J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial nets, in: *Adv. Neural Inf. Process. Syst.*, 2014. doi:10.3156/jsoft.29.5_177_2.
- [64] X. Yi, E. Walia, P. Babyn, Generative adversarial network in medical imaging: A review, *Med. Image Anal.* 58 (2019). doi:10.1016/j.media.2019.101552.
- [65] A.J. Hawkins, Welcome to Simulation City, the Virtual World where Waymo Tests its Autonomous Vehicles, July 6, 2021. (2021). <https://www.theverge.com/2021/7/6/22565448/waymo-simulation-city-autonomous-vehicle-testing-virtual> (accessed January 25, 2022).
- [66] Z. Yang, Y. Chai, D. Anguelov, Y. Zhou, P. Sun, D. Erhan, S. Rafferty, H. Kretzschmar, SurfGAN: Synthesizing realistic sensor data for autonomous driving, *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.* (2020) 11115–11124. doi:10.1109/CVPR42600.2020.01113.
- [67] E.J. Kirkland, *Advanced Computing in Electron Microscopy*, Third Edit, 2020.
- [68] C. Zhang, R. Han, A.R. Zhang, P.M. Voyles, Denoising atomic resolution 4D scanning transmission electron microscopy data with tensor singular value decomposition, *Ultramicroscopy.* 219 (2020) 113123. doi:10.1016/j.ultramic.2020.113123.
- [69] G.D. Förster, A. Castan, A. Loiseau, J. Nelayah, D. Alloyeau, F. Fossard, C. Bichara, H. Amara, A deep learning approach for determining the chiral indices of carbon nanotubes from high-resolution transmission electron microscopy images, *Carbon N. Y.* 169 (2020) 465–474. doi:10.1016/j.carbon.2020.06.086.
- [70] P. Kukelhan, A. Beyer, C. Fuchs, M.J. Weseloh, S.W. Koch, W. Stolz, K. Volz, Atomic structure of ‘W’-type quantum well heterostructures investigated by aberration-corrected STEM, *J. Microsc.* 268 (2017) 259–268. doi:10.1111/jmi.12647.

- [71] B. Yao, D.J. Edwards, R.J. Kurtz, G.R. Odette, T. Yamamoto, Multislice simulation of transmission electron microscopy imaging of helium bubbles in Fe, *J. Electron Microsc.* (Tokyo). 61 (2012) 393–400. doi:10.1093/jmicro/dfs065.
- [72] K.G. Field, R. Jacobs, M. Shen, M. Lynch, P. Patki, C. Field, D. Morgan, Development and Deployment of Automated Machine Learning Detection in Electron Microscopy Experiments, *Microsc. Microanal.* 27 (2021) 2136–2137. doi:10.1017/s1431927621007704.
- [73] A. Radford, L. Metz, S. Chintala, Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks, *ArXiv:1511.06434*. (2016) 1–16.
- [74] J. Zhu, T. Park, A.A. Efros, B. Ai, U.C. Berkeley, Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks, in: *Proc. IEEE Int. Conf. Comput. Vis.*, 2017: pp. 2223–2232.
- [75] P. Isola, J. Zhu, A.A. Efros, B. Ai, U.C. Berkeley, Image-to-Image Translation with Conditional Adversarial Networks, in: *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017: pp. 1125–1134.
- [76] J. Lee, N.H. Goo, W.B. Park, M. Pyo, K. Sohn, Virtual microstructure design for steels using generative adversarial networks, *Eng. Reports*. 3 (2021) 1–14. doi:10.1002/eng2.12274.
- [77] 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). doi:10.1038/s41524-020-00392-6.
- [78] M. Arjovsky, S. Chintala, Wasserstein Generative Adversarial Networks, in: *Proc. 34th Int. Conf. Mach. Learn. PMLR*, 2017: pp. 214–223.
- [79] T. Hsu, W.K. Epting, H. Kim, H.W. Abernathy, G.A. Hackett, A.D. Rollett, P.A. Salvador, E.A. Holm, Microstructure Generation via Generative Adversarial Network for Heterogeneous, Topologically Complex 3D Materials, *JOM*. 73 (2021) 90–102. doi:10.1007/s11837-020-04484-y.
- [80] Blender Foundation, Home of the Blender Project - Free and Open 3D Creation Software, (2018). <https://blender.org>.

- [81] A. Cid-Mejías, R. Alonso-Calvo, H. Gavilán, J. Crespo, V. Maojo, A deep learning approach using synthetic images for segmenting and estimating 3D orientation of nanoparticles in EM images, *Comput. Methods Programs Biomed.* 202 (2021). doi:10.1016/j.cmpb.2021.105958.
- [82] L. Mill, D. Wolff, N. Gerrits, P. Philipp, L. Kling, F. Vollnhals, A. Ignatenko, C. Jaremenko, Y. Huang, O. De Castro, J.N. Audinot, I. Nelissen, T. Wirtz, A. Maier, S. Christiansen, Synthetic Image Rendering Solves Annotation Problem in Deep Learning Nanoparticle Segmentation, *Small Methods*. 5 (2021). doi:10.1002/smtd.202100223.
- [83] P. Trampert, D. Rubinstein, F. Boughorbel, C. Schlinkmann, M. Luschkova, P. Slusallek, T. Dahmen, S. Sandfeld, Deep neural networks for analysis of microscopy images—synthetic data generation and adaptive sampling, *Crystals*. 11 (2021) 1–13. doi:10.3390/cryst11030258.
- [84] O. Vinyals, C. Blundell, T. Lillicrap, Matching Networks for One Shot Learning, in: *Adv. Neural Inf. Process. Syst.* 29 (NIPS 2016), 2016: pp. 1–9.
- [85] 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). doi:10.1038/s41524-021-00652-z.
- [86] L. Fu, H. Yu, X. Li, M. Images, Deep Learning for Object Detection in Materials-Science Images: A tutorial, *Signal Process. Adv. Mater.* (2022) 78–88.
- [87] J.R. Hatrick-Simpers, B. DeCost, A.G. Kusne, H. Joress, W. Wong-Ng, D.L. Kaiser, A. Zakutayev, C. Phillips, S. Sun, J. Thapa, H. Yu, I. Takeuchi, T. Buonassisi, An Open Combinatorial Diffraction Dataset Including Consensus Human and Machine Learning Labels with Quantified Uncertainty for Training New Machine Learning Models, *Integr. Mater. Manuf. Innov.* 10 (2021) 311–318. doi:10.1007/s40192-021-00213-8.
- [88] M. Ziatdinov, A. Ghosh, T. Wong, S. Kalinin, AtomAI: A Deep Learning Framework for Analysis of Image and Spectroscopy Data in (Scanning) Transmission Electron Microscopy and Beyond, *ArXiv*. (2021) 1–25.
- [89] M. Ziatdinov, S. Kalinin, AtomAI : Open-source software for applications of deep learning to microscopy data, *Microsc. Microanal.* 27 (2021) 3000–3002.

- doi:10.1017/S1431927621010436.
- [90] C. Doty, S. Gallagher, W. Cui, W. Chen, S. Bhushan, M. Oostrom, S. Akers, S.R. Spurgeon, Design of a graphical user interface for few-shot machine learning classification of electron microscopy data, *Comput. Mater. Sci.* 203 (2022) 1–6.
 - [91] R. Fernandez-Gonzalez, N. Balaghi, K. Wang, R. Hawkins, K. Rothenberg, C. McFaul, C. Schimmer, M. Ly, A.M. do Carmo, G. Scepanovic, G. Erdemci-Tandogan, V. Castle, PyJAMAS: open-source, multimodal segmentation and analysis of microscopy images, *Bioinformatics*. 38 (2022) 594–596. doi:10.1093/bioinformatics/btab589.
 - [92] L. von Chamier, R.F. Laine, J. Jukkala, C. Spahn, D. Krentzel, E. Nehme, M. Lerche, S. Hernández-Pérez, P.K. Mattila, E. Karinou, S. Holden, A.C. Solak, A. Krull, T.-O. Buchholz, M.L. Jones, L.A. Royer, C. Leterrier, Y. Shechtman, F. Jug, M. Heilemann, G. Jacquemet, R. Henriques, Democratising deep learning for microscopy with ZeroCostDL4Mic, *Nat. Commun.* 12 (2021) 1–18. doi:10.1038/s41467-021-22518-0.
 - [93] K.T. Mukaddem, E.J. Beard, B. Yildirim, J.M. Cole, ImageDataExtractor: A Tool to Extract and Quantify Data from Microscopy Images, *J. Chem. Inf. Model.* 60 (2020) 2492–2509. doi:10.1021/acs.jcim.9b00734.
 - [94] Reactiv IP: <https://www.reactivip.com/>, (n.d.).
 - [95] Dragonfly: <https://www.theobjects.com/dragonfly/index.html>, (n.d.).
 - [96] DeeplImageJ: <https://www.epfl.ch/research/domains/imaging/imaging/support-in-image-analysis/deepimagej/>, n.d.
 - [97] Theia Scientific, LLC, (n.d.).
 - [98] X.G. Li, B. Blaiszik, M.E. Schwarting, R. Jacobs, A. Scourtas, K.J. Schmidt, P.M. Voyles, D. Morgan, Graph network based deep learning of bandgaps, *J. Chem. Phys.* 155 (2021). doi:10.1063/5.0066009.
 - [99] U. of Chicago, U. of Wisconsin-Madison, Foundry Materials Informatics Environment:, (2021). <https://ai-materials-and-chemistry.gitbook.io/foundry/v/docs/>.
 - [100] Z. Cai, N. Vasconcelos, Cascade R-CNN: Delving into High Quality Object Detection, *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.* (2018) 6154–6162.
 - [101] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, S. Zagoruyko, End-to-End Object

Detection with Transformers, in: Eur. Conf. Comput. Vis., 2020: pp. 213–229.

doi:https://doi.org/10.1007/978-3-030-58452-8_13.

- [102] C. Zhang, H. Li, X. Wan, X. Chen, Z. Yang, J. Feng, F. Zhang, TransPicker: a Transformer-based Framework for Particle Picking in cryoEM Micrographs, in: 2021 IEEE Int. Conf. Bioinforma. Biomed., 2022: pp. 1179–1184. doi:[10.1109/bibm52615.2021.9669524](https://doi.org/10.1109/bibm52615.2021.9669524).