Journal of Applied Physics

Image-driven discriminative and generative machine learning algorithms for establishing microstructure–processing relationships

Cite as: J. Appl. Phys. **128**, 134901 (2020); https://doi.org/10.1063/5.0013720
Submitted: 13 May 2020 • Accepted: 04 September 2020 • Published Online: 02 October 2020

W. Ma, E. J. Kautz, A. Baskaran, et al.

COLLECTIONS

Paper published as part of the special topic on Machine Learning for Materials Design and Discovery

View Online Export Citation CrossMark

ARTICLES YOU MAY BE INTERESTED IN

Machine learning for materials design and discovery
Journal of Applied Physics **129**, 070401 (2021); https://doi.org/10.1063/5.0043300

Machine learning approaches for the prediction of materials properties
APL Materials **8**, 080701 (2020); https://doi.org/10.1063/5.0018384

Perspective: Materials informatics and big data: Realization of the "fourth paradigm" of science in materials science
APL Materials **4**, 053208 (2016); https://doi.org/10.1063/1.4946894

© 2020 Author(s).

Journal of Applied Physics ARTICLE scitation.org/journal/jap

Image-driven discriminative and generative machine learning algorithms for establishing microstructure–processing relationships

Cite as: J. Appl. Phys. **128**, 134901 (2020); doi: 10.1063/5.0013720
Submitted: 13 May 2020 · Accepted: 4 September 2020 ·
Published Online: 2 October 2020

View Online Export Citation CrossMark

W. Ma,[1] E. J. Kautz,[2,a)] A. Baskaran,[3] A. Chowdhury,[4] V. Joshi,[5] B. Yener,[1] and D. J. Lewis[3,a)]

AFFILIATIONS

[1]Computer Science Department, Rensselaer Polytechnic Institute, Troy, New York 12180, USA
[2]National Security Directorate, Pacific Northwest National Laboratory, Richland, Washington 99352, USA
[3]Materials Science and Engineering Department, Rensselaer Polytechnic Institute, Troy, New York 12180, USA
[4]Artificial Intelligence, GE Research, Niskayuna, New York 12309, USA
[5]Energy and Environment Directorate, Pacific Northwest National Laboratory, Richland, Washington 99352, USA

Note: This paper is part of the special collection on Machine Learning for Materials Design and Discovery
[a)]**Authors to whom correspondence should be addressed:** elizabeth.kautz@pnnl.gov and lewisd2@rpi.edu

ABSTRACT

We investigate the methods of microstructure representation for the purpose of predicting processing condition from microstructure image data. A binary alloy (uranium–molybdenum) that is currently under development as a nuclear fuel was studied for the purpose of developing an improved machine learning approach to image recognition, characterization, and building predictive capabilities linking microstructure to processing conditions. Here, we test different microstructure representations and evaluate model performance based on the F1 score. A F1 score of 95.1% was achieved for distinguishing between micrographs corresponding to ten different thermo-mechanical material processing conditions. We find that our newly developed microstructure representation describes image data well, and the traditional approach of utilizing area fractions of different phases is insufficient for distinguishing between multiple classes using a relatively small, imbalanced original dataset of 272 images. To explore the applicability of generative methods for supplementing such limited datasets, generative adversarial networks were trained to generate artificial microstructure images. Two different generative networks were trained and tested to assess performance. Challenges and best practices associated with applying machine learning to limited microstructure image datasets are also discussed. Our work has implications for quantitative microstructure analysis and development of microstructure–processing relationships in limited datasets typical of metallurgical process design studies.

© 2020 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/). https://doi.org/10.1063/5.0013720

I. INTRODUCTION

Microstructure image data are rich in information regarding morphology and implied composition of constituent phases and can provide unique insight into the pathways leading to microstructure formation and mechanisms responsible for material behavior and performance. Thus, the analysis of micrographs (i.e., microstructure image data) is central to several materials science studies establishing processing–structure–property relationships and for the design of new material systems. Despite the ubiquity of micrographs in material science research, significant challenges exist related to consistent and accurate recognition and analysis of image data. Such challenges arise from the domain knowledge and skill required to obtain micrographs, the diverse types of image data possible (e.g., optical and electron microscopy), domain-specific challenges to image analysis techniques, and more. With the advancing application of artificial intelligence (AI) (i.e., machine learning) in a wide range of fields, we find that the application of established AI methods to microstructure recognition and analysis opens up an opportunity for computationally guided experiments and objective, repeatable analysis of image data. To address the need for improved

© Author(s) 2020

microstructure quantification via image-driven machine learning using small, imbalanced datasets, we investigate microstructure–processing relationships in a model binary uranium–molybdenum (U–Mo) alloy. The U–Mo system is of particular interest due to the alloy's applicability as a nuclear fuel for research reactors and the need to understand microstructure–processing relationships for improved fabrication design and fuel qualification.

Uranium (U) alloyed with 10 weight percent (wt. %) molybdenum (Mo), referred to here as U-10Mo, is currently under development as a new metallic nuclear fuel for application in research and radioisotope production facilities. U-10Mo is a candidate for low enriched U (LEU) fuel, designed to replace currently used highly enriched U (HEU) fuels with the aim of reducing proliferation and safety risks associated with HEU handling and operation.[1–4] A monolithic, plate-type design for U-10Mo has been selected due to the high U densities achievable while meeting the low enrichment specification, where the fuel must have $\leq$20% ^{235}U relative to all U isotopes. In order to fabricate fuel plates to meet dimensional requirements, the U-10Mo alloy must be subjected to several thermo-mechanical processing steps, leading to microstructural evolution during fabrication (e.g., hot rolling and hot isostatic pressing). To design a fuel with microstructure that meets performance requirements and to enable future materials processing design, the microstructure–processing relationship must be well-established.

The equilibrium phase of pure U at room temperature (α-U) has an orthorhombic crystal structure. α-U is known to experience non-uniform thermal expansion in a high temperature, irradiation environment, thus 10 wt. % Mo is added to stabilize the high-temperature BCC γ-U phase at room temperature. During processing, U-10Mo is exposed to temperatures below 560−575 °C during hot isostatic pressing (HIP). Below these temperatures, a eutectoid decomposition of the metastable γ-UMo matrix phase into α-U and γ' (U_2Mo) is expected, based on the equilibrium binary phase diagram.[5] Prior work has demonstrated that this eutectoid decomposition occurs via a discontinuous precipitation (DP) mechanism. The decomposition involves the γ-UMo matrix phase transforming to α-U and Mo-enriched γ-UMo products with lamellar morphology. Our previous work showed that this transformation was initiated primarily at grain boundaries and interphase interfaces, where Si segregation was observed.[6–9]

Significant prior work has been performed using machine learning for a range of materials science applications.[10–19] A rapidly growing area in machine learning in materials science is in image data quantification. Previous studies demonstrated success of convolutional neural networks (CNNs) in microstructure recognition tasks without significant development time and state-of-the-art performance for a wide range of microstructures (e.g., forged titanium, perlitic steel, metal powder, and ceramics).[11–15,20,21] Additionally, the application of machine learning methods to large image datasets such as those available through ImageNet is routinely done.[22,23] The ImageNet database includes over 14×10^6 natural images that can be used for training and testing of machine learning models. Application of machine learning methods to limited datasets is still a frontier in the machine learning community[16,20,24–26] and is of interest to materials science data analysis problems, where only limited, unbalanced, or historic datasets are available, and the cost/time associated with obtaining very large datasets is prohibitive.

The present study explores the applicability of image-driven machine learning methods[12] to developing microstructure–processing relationships. Specifically, we seek to understand the role of several thermomechanical processing steps in the microstructure evolution observed in the U-10Mo system. An improved approach to determining microstructure–processing relationships is developed and presented here, involving feature extraction, segmentation, and classification using a random forest model. Microstructure image data are segmented to identify microstructural features of interest and quantify area fraction of these features, including the γ-UMo matrix, uranium carbide, and DP reaction transformation products. The application of generative adversarial networks (GANs)[27,28] is also discussed as an emerging method for microstructure image generation. Our work has broad implications for machine learning applications in microstructure image analysis and the development of quantitative microstructure–processing relationships in a wide range of alloy systems.

II. EXPERIMENTAL AND COMPUTATIONAL METHODS

A. Image data

Image data used in this work are from two scanning electron microscopes (SEMs): a FEI Quanta dual beam Focused Ion Beam/Scanning Electron Microscope (FIB/SEM) and a JEOL JSM-7600F Field Emission SEM. The backscatter electron detector was used for improved atomic number (Z) contrast. Two different microscope operators took the images. Thus, the image data analyzed here were diverse in terms of resolution, contrast, focus, and magnifications selected. The idea in using a variety of images taken by different operators using different microscopes (but all of the same samples) was to develop a more robust model that can distinguish between different material processing conditions.

All images used in this work are of a depleted U-10Mo alloy fabricated and prepared according to the details presented elsewhere.[29,30] Images were taken over a range of magnifications from 250× to 500×. Image data were labeled based on the processing condition, detailed in Fig. 1. Ten different image classes were studied, where each class corresponds to a different processing history that generates a unique microstructure. The processing conditions detailed in Fig. 1 include two different homogenization annealing treatments (900 °C−48 h and 1000 °C−16 h) and several thermo-mechanical processing steps such as cold and hot rolling. Each image class is therefore labeled by the homogenization treatment (HT) and processing condition (C) numbers, where 900 °C−48 h is referred to as HT1 and 100 °C−16 h is HT2. Processing conditions are indicated by C followed by the number in the list of all possible conditions in Fig. 1. For example, HT1-C1 is a U-10Mo sample that is in the as-cast and homogenized condition, where homogenization was done at 900 °C for 48 h. Representative micrographs from each class are given in Fig. 2.

Original images vary in size. Different image sizes used in this work (in pixels) include 2048 by 2560, 1448 by 2048, 1428 by 2048, and 1024 by 1280. All images included a scale bar region which was removed prior to training and testing by cropping the image. The dataset analyzed here consists of a total of 272 original images from 10 classes. Bilateral filters were applied to each image for

© Author(s) 2020

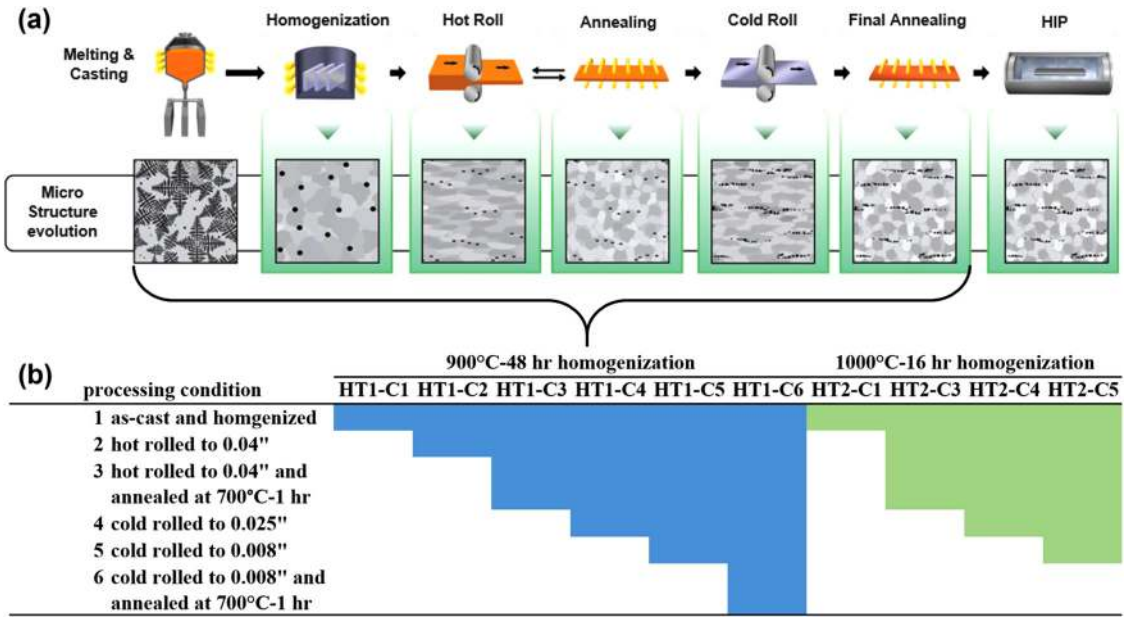

FIG. 1. (a) Schematic of U-10Mo fuel fabrication, where the steps shown in the bracket were used to generate microstructure imaged and analyzed in this work. (b) Sample matrix of sample processing conditions analyzed. For each condition, SEM micrographs were analyzed. Two different homogenization temperatures were utilized, and thus samples are grouped based on the homogenization treatment performed. Sample conditions are indicated by homogenization treatment (HT) number, where HT-1 refers to 900 °C−48 h and HT-2 refers to 1000 °C−16 h. The processing condition (C) the micrographs represent is indicated by C followed by a number that corresponds to the number processing condition (e.g., C1 is a sample in the as-cast condition).

noise removal while keeping edges sharp. In our study, we chose the diameter of each pixel neighborhood as 15 while keeping all other default parameters.

B. Discriminative methods

1. Feature extraction

In order to determine how to best quantify microstructure image data on the U-10Mo system (as a function of thermo-mechanical processing parameters), different methods of feature extraction were developed and tested. Here, each microstructure image is described by a feature vector, and how that feature vector is derived either depends on area fraction of different regions or spatial relationships between microstructural features of interest. These two types of features are referred to here as area and spatial features, respectively. These two different feature types are extracted from each image after segmentation. Area features are simply the area fractions of each phase or region (γ-UMo matrix, UC, and lamellar transformation products). U-10Mo microstructures have been described by the area fractions of these regions in prior work.[8,29,31] Spatial features are computed by first measuring the following for each region (matrix, carbide, lamellar transformation products): the x and y coordinates of the centroid, area (in square pixels), and the ratio of area of the region to the area of its bounding box. The spatial feature is simply a concatenation the following measures: the number of regions, the mean and standard deviation of the areas, the standard deviation of the centroid coordinates, and the mean and standard deviation of area ratios.

C. Approach and machine learning model

All experimentation was carried out with Python version 3.6.9 with the help of various open-source libraries. The opencv, scipy, skimage, numpy, and sklearn packages (compatible with Python version 3.6.9) were used for training, testing, and validation. All relevant model parameters are summarized in Table I.

The approach to image recognition and characterization developed here is schematically described in Fig. 3(a) and involves the following steps: (1) image segmentation, (2) extracting interpretable features from the image data, and (3) classifying microstructures from different classes based on extracted features.

The image segmentation algorithm used here is based on our prior work where k-means was applied to classify image pixels based on the grayscale values.[31] This method is built upon the assumption that pixels correspond to different grayscale values and the differences between clusters are significant. However, in our dataset not every image comes with three different phases (dictated by processing condition), which leaves us with the question whether k is 2 or 3 for each image. While there are some well-known methods for choosing k, such as the elbow method and the silhouette method, they do not work well in our experiments. A reason for why these methods do not work here is that the grayscale shades are typically spread out evenly on a γ-compressed nonlinear scale, which

© Author(s) 2020

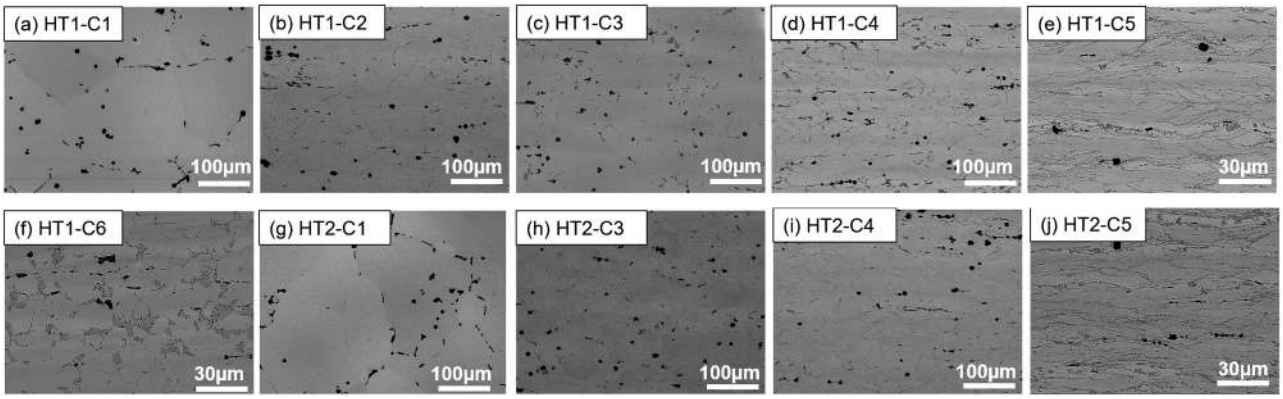

FIG. 2. Example micrographs from the different material conditions studied in this work. Micrographs from homogenization temperature 1 (HT1) are given in (a)–(f) for conditions 1–6 (C1–C6), and for homogenization temperature 2 (HT2) in (g)–(j) for conditions 1–5 (C1–C5). A total of ten classes were assigned as part of this effort. In the area of classification problems, this is a large number of class labels; furthermore, the availability of microstructures and the subtle differences between structures further complicates the problem of successful classification.

means there are insignificant differences in grayscale values, even though they are noticeable to the human eye. Thus, for the image data used here, a specific k value needs to be hard coded for each class. However, this hard coding requires ground truth knowledge about the image processing condition (i.e., class label).

To overcome the limitations associated with applying k-means clustering to our image data, we developed a two-stage segmentation method that combines the k-means clustering and the image morphology. This approach is schematically described in Fig. 3(b). In the first stage, we apply k-means clustering based on the grayscale values of each pixel with $k = 2$. The purpose of this step is to segment the γ-UMo matrix phase from the rest of the image. In the second stage, we apply morphological opening and closing (i.e., dilation and erosion),[33] to remove the fine-scale lamellae in the transformed region (so that this region is considered as a single grayscale value), and smooth the border of UC inclusions. These morphological operations aid in improving segmentation results via k-means. It is noted that for the purpose of this work, it is desirable that transformation products that appear as fine lamellae are treated as one region, where distinguishing between lamellae is not needed.

K-means was selected over other popular clustering methods (e.g., Agglomerative Hierarchical Clustering, Spectral clustering, DBSCAN, Single Linkage, and DeBaCl Geom Tree) due to numerous pixels classified in our work, and the faster time to classify pixels and complete the clustering process in comparison to these other algorithms.

TABLE I. Parameters selected for model specification, compilation, and cross validation for U-10Mo image analysis.

	Parameter	Value
Preprocessing	Noise-reducing	Bilateral
	Diameter	15
	Sigma color	75
	Sigma space	75
Segmentation	k-means	k-means++ center initialization[32]
	Lamellar	(9,9) closing then (9,9) opening
	UC	(9,9) opening then (3,3) closing

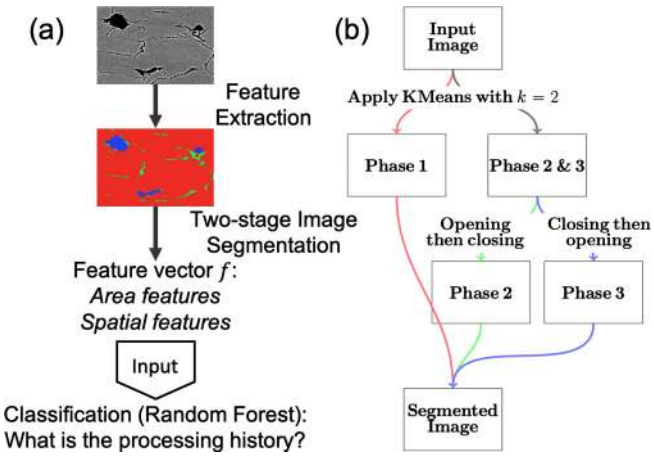

FIG. 3. (a) Schematic describing the approach developed for the ten class classification experiment performed in this work. A two-stage segmentation approach was used and is schematically described in (b). The two stage segmentation is described as follows: (1) stage 1 involved the use of multiple features and (2) stage 2 utilized the area-based features. A random forest model was then used for classification and five-fold cross validation was used to validate results.

© Author(s) 2020

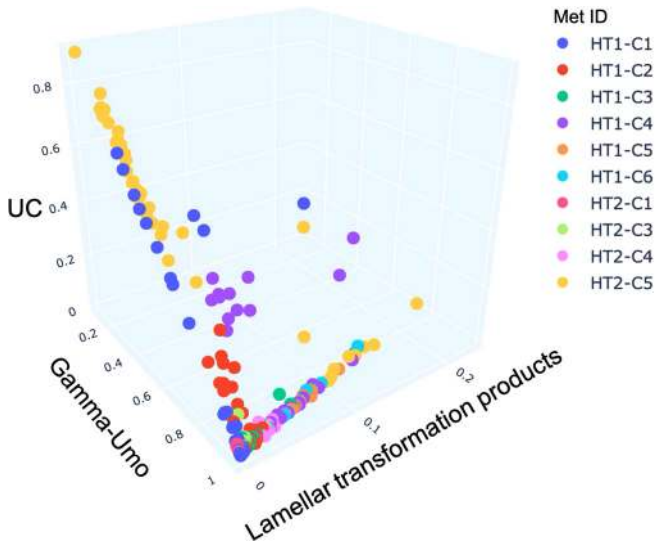

FIG. 4. Visualization of area feature for each phase.

III. RESULTS AND DISCUSSION

A. Developing microstructure–processing relationships using discriminative learning methods

Developing understanding of microstructure–processing relationships and improving predictive capability become more difficult as processing complexity increases. In our case study on the U-10Mo system, several steps are performed during fuel fabrication, and the ability to recognize what processing parameters lead to a given microstructure can allow for improved process design and quality control. However, the question of how to quantitatively describe microstructure image data in order to predict processing condition from microstructure images remains unanswered. Significant prior work has been performed in which area fraction of different phases with varying gray scale serves as a proxy for volume fraction and is thus used as the primary quantitative microstructure descriptor.[8,29,31] Yet, the choice of area fraction may not be the best metric when several phenomena are changing with varying processing conditions, such as extent of phase transformations, distribution or fragmentation of inclusions, and change in grain size and morphology. To measure how accurately different features can represent microstructures, we use features (area and spatial, described above) as inputs to train a random forest model to predict the corresponding processing history. Images are segmented and area and spatial features are extracted. In addition, we collect other texture features, such as the Haralick features and the local binary patterns (LBPs), which have previously been demonstrated to represent microstructure image data well.[12,31] The following four experiments were considered to explore metrics of microstructure representation:

1. Characterization of micrographs using area features only. For the two tasks below, we train two separate random forest models for classification and fivefold cross-validation is applied to evaluate the model performance:
   (a) A 10-class classification to predict microstructure processing history (HT1-C1, HT1-C2, HT2-C1, etc.);
   (b) A binary classification to predict the homogenization temperatures (900 °C−48 h or 1000 °C−16 h).
2. Characterization of micrographs based on area, and spatial features, in an effort to increase predictive power of our model. Similar to Experiment 1 (above), we train two random forest models for the two tasks listed in item 1(a) and (b).
3. Characterization of microgrpahs using area, spatial, and texture features. All features are concatenated as a single feature vector to represent a microstructure image. A model is trained to learn and predict the processing history (HT1-C1, HT1-C2, HT2-C1, etc.) of an image.
4. Binary classification for each possible pair of processing histories (HT1-C1, HT1-C2, HT2-C1, etc.) based on area features only. This experiment provides a detailed investigation of how well area features represent micrographs.

Training results from these three experiments are summarized in Table II. The model performance is measured in F1 score, defined as follows:

$$\mathrm{F1} = \frac{2 \times \mathrm{precision} \times \mathrm{recall}}{\mathrm{precision} + \mathrm{recall}}$$

In Experiment 1, the F1 scores of the 10-class classification and the binary classification are 62.4% and 68.5%, respectively. In Experiment 2, spatial features are added to the area features to help improve model classification results. The performance of models is improved significantly to 78.9% and 65.1% for Experiments 2a and 2b, respectively. This increased performance indicates that spatial features are correlated with the processing histories. In Experiment 3, we used all the features available (both interpretable area and spatial features, and texture features) and reach an F1 score of 95.1% for the 10-class classification task. This result serves as a benchmark for this dataset and allows us to evaluate the predictive power of other models. While the area features have long been regarded as a strong indication of the microstructure processing history, from the microstructure representation experiments detailed here, we find that the predictive power of area features is actually very limited. This limitation can be visualized in Fig. 4, where there are multiple overlapping data

TABLE II. Summary of experiments, features used to represent microstructure image data, metric, and performance.

Experiment	Features	Metric	Performance
1a	Area features	F1	62.4%
1b	Area features	F1	68.5%
2a	Area and spatial features	F1	78.9%
2b	Area and spatial features	F1	65.1%
3	All features	F1	95.1%
4	Area features	F1	See Fig. 5

© Author(s) 2020

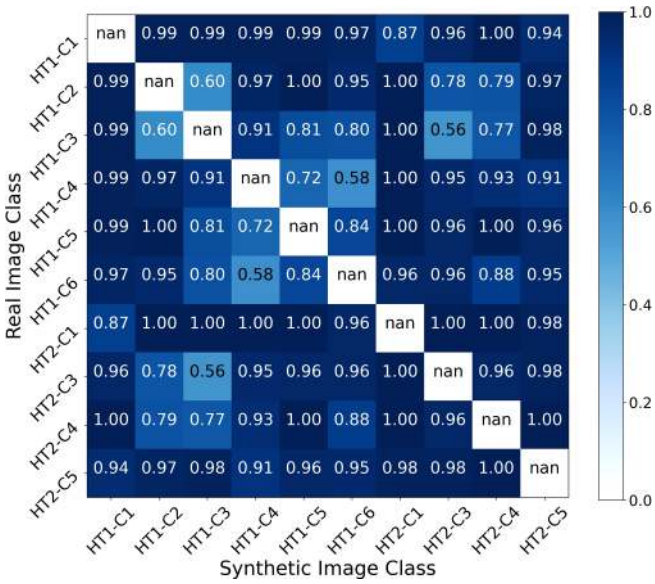

FIG. 5. Model performance of binary classification for each pair of processing histories for the U-10Mo microstructure. Here, the F1 scores are reported on a 0–1 scale, where 1 indicates 100% correct predictions.

points, and thus micrographs from different processing histories are difficult to separate. Based on a trained random forest model from Experiment 3, the feature importance of the area feature corresponding to UC is 0.09, which is higher than the other 40 features. However, there are many other features from spatial and texture features with a feature importance of approximately 0.06. This conclusion can also be verified in Experiment 4 (results given in Fig. 5), where we find that for binary classifications between two specific processing histories, the area features do not always result in high F1 scores. This finding is highlighted by very poor classification performance listed in the matrix, for example, a F1 score of 61% for the following two conditions which were both homogenized at 900 °C−48 h: HT1-C4 (cold rolled to 0.025 in.) and HT1-C6 (cold rolled to 0.008 in. and annealed at 700 °C−1 h).

B. Synthetic microstructure generation using generative adversarial networks

Generative adversarial networks (GANs) have been proven successful for many image synthesis and unsupervised learning tasks.[28] It is a popular framework for representation learning, such as disentangle pose from lighting in 3D rendered images,[34] and image completion, where large missing regions are synthesized utilizing the surrounding image features.[35] Variants of GANs have surpassed many other generative models in the quality of samples as well as their underlying representation. Recently, GANs have emerged as a promising methodology for application in computational materials design, for the purpose of developing structure-property and structure–performance relations via physical simulations.[36,37] GANs are implemented by deep neural networks and thus are able to capture complex microstructural characteristics. Hence, we investigate different GAN architectures here for the specific material system of U-10Mo and the task of generating realistic artificial micrographs that could be useful in supplementing real datasets or used in an effort to predict microstructure from processing parameters.

A GAN framework consists of a generator, G, that generates samples from a noise variable, z, and a discriminator, D, that aims to distinguish between samples from the real data distribution and those from the synthetic data distribution (from the generator). The training of a GAN can be summarized as a two-player minimax game:

$$\min_G \max_D V(D, G) = \mathbb{E}_{x\sim p_{\text{data}}(x)}[\log D(x)] + \mathbb{E}_{z\sim p_z(z)}[\log(1 - D(G(z))],$$

where p_{data} is the underlying distribution of real images and p_z is some noise distribution.

Although the objective of the training is straightforward, the actual training can be quite unstable because of the non-convex cost functions and the high-dimensional parameter space.[38] In practice, the model could encounter many problems, such as vanishing gradients, where the discriminator gets too good and the generator fails to make progress, and the mode collapses (i.e., the generator collapses to a state where it always outputs the same sample). Specifically in cases where we want the GAN to learn a disentangled representation of the training data and output high-quality samples, the training can be extremely hard to converge.

In this work, we make use of multiple variants of GANs to generate artificial microstructure images and demonstrate how GANs can be used to synthesize realistic images with varying resolution. In this small case study, the same set of original SEM-BSE U-10Mo micrographs described previously are used as the training set. Images are cropped into 1024 by 1024 and resized to 512 by 512 square pixels. We choose 512 by 512 as the size of training images and output samples for several reasons, including the following:

1. High-resolution images are needed for characterization. Phases such as the lamellar transformation products may not be represented well if images are too low in resolution.
2. The microstructural area contained within the image should be large enough to reflect the processing history. This would help to keep the variance of the training images small enough so that the GAN synthesized images represent the different classes well.
3. The most recent GAN technology is capable of higher resolution images up to 1024 squared and in this study we wished to explore this higher resolution capability.

Two different GAN architectures were trained and compared on the basis of artificial image quality: (1) a progressively growing GAN,[39] where samples are generated from latent noise in original micrographs, and (2) a Pix2Pix generative model,[40] where a segmentation label map is provided as input.

1. Progressive growing GAN

Progressively growing GAN (pg-GAN) is an adversarial network variant that helps to stabilize the training of a high-resolution GAN. The generator is initialized with low resolution images, over which

© Author(s) 2020

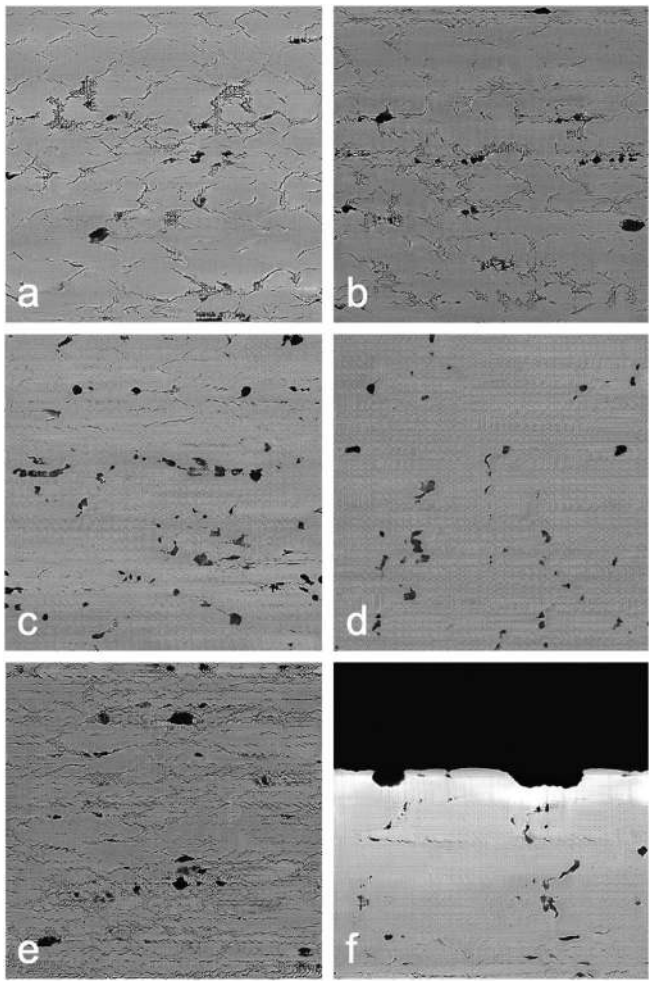

FIG. 6. Example synthetic images generated by the trained progressive growing GAN. Images given in (a)–(f) show varying microstructural features, specifically different extent of lamellar transformation products, and distribution of carbides. The micrograph in (f) shows the edge of a fuel plate.

new layers are added progressively to capture finer spatial details. Each of the new layers is treated as a residual block that smoothly blends into the network when the resolution of the GAN is doubled.

Data augmentation is applied to the original set of 272 SEM-BSE images in order to increase the number of images available for training. The data augmentation utilized here involves cropping original images into smaller squares with a horizontal shift, rotation by 90° and horizontal/vertical flipping. Finally, images are resized to 512 by 512 square pixels. Using these methods, a total of 10 880 images were available for training.

We follow the model specification from the original paper[39] with Python 3.6.9 and TensorFlow-GPU 1.13.1. We use the Adam optimizer with the default learning rate scheduling algorithm. Both training images and output samples are 512 by 512 square pixels, and the training length is set to 1 000 000 images. To measure the model performance, we sampled 1000 images generated by the model, with examples given in Fig. 6. The images sampled from the generated set are qualitatively close to the real data distribution although some images contain visual artifacts, such as image (f) in Fig. 6 (the boundary between the microstructure and the background should be a straight line).

Lastly, the approximated distribution is entangled, meaning that the image data are encoded in a complicated manner and the input noise variables are not interpretable (see Fig. 7). Although the synthetic images are visually nice, we cannot interpret the role of the input noise vectors in the generation of synthetic images. This blocks us from understanding how samples from different processing histories are distributed in the learned space of microstructure images or possibly revealing their underlying connections. Additionally, from visual examination of the Fig. 6 artificial images, we find some spatial patterns that are not visible in training images. Visually, we see the texture of artificial images is different, and such texture anomalies are discussed in further detail in Sec. III B 4.

To better assess the GAN results, we turn to automated methods such as the sliced Wasserstein distance (SWD)[41] to measure how similar artificial images are to the training set over different scales. We measure the SWD with the checkpoint images during training and plot the distances with respect to the number of training images fed into the model, with results given in Fig. 8. We find that the SWD at different resolutions generally decreases

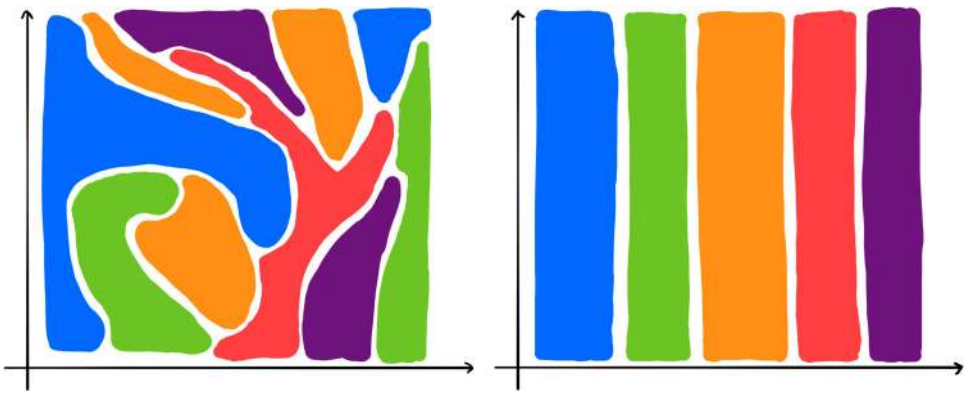

FIG. 7. Schematic of an entangled representation (left) and disentangled representation (right). In the entangled representation, data are encoded in a complicated manner. In the disentangled representation, the independent variables are interpretable.

© Author(s) 2020

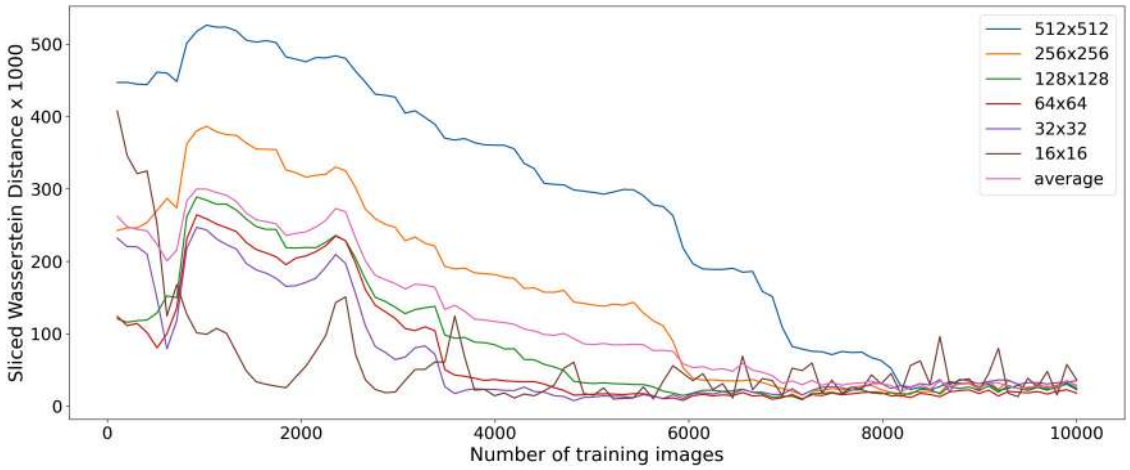

FIG. 8. We use the sliced Wasserstein distance (SWD) between the training images and generated images to evaluate the model performance. This figure shows the SWD under different resolutions as the training progress.

as the training proceeds. The model converges after approximately 8000 thousand training images. Even with limited micrographs for training, the progressive growing GAN can still learn the real data distribution well, as demonstrated by the realistic synthetic images shown in Figs. 6(a)–6(f).

2. Pix2Pix generative model

Unlike the progressive growing GAN, Pix2Pix GAN is a conditional GAN[42] variant that learns a mapping from some extra given information y ("condition") and input noise z to output images. The objective of a conditional GAN is given by

$$\min_G \max_D V(D, G) = \mathbb{E}_{x\sim p_{\text{data}}(x)}[\log D(x \mid y)] + \mathbb{E}_{z\sim p_z(z)}[\log(1 - D(G(z \mid y)))].$$

Image-to-image translation is the task of learning a mapping between an input image and an output image. Several examples of this image-to-image translation exist in the literature, including Zhu *et al.* used CycleGAN to transfer images of one style to another, such as between landscape images in summer and in winter and between photographs and paintings of Monet.[43] Park *et al.* used GauGAN to create photorealistic images from segmentation maps, which are labeled sketches that depict the layout of a scene.[44] Isola *et al.* used Pix2Pix on many applications of image-to-image translation, such as mapping from aerial photos to maps and mapping from edges to photos.[40] In this work, we use the labels generated from the segmentation algorithm detailed above as style A and realistic microstructure images as style B. Overall, the Pix2Pix generative model takes a labeled image as input and generates a realistic microstructure image, as shown in Fig. 9. We note here that the segmented image given as the input includes some noise (due to charging from the sample in the SEM) that was segmented as a separate phase. Although from a segmentation point of view, this is not an accurate representation of the microstructure (i.e., the noise is not an important microstructure feature of the image), it does mean that the charging artifact in the original image is accurately captured in the synthetic image, thereby making the synthetic image realistic when compared to the original image data. The same set of 272 original images is used, after removing the scale bar and cropping each original image into squares. These images are then used for training and are referred to as real B, which serve as the ground-truth images (style

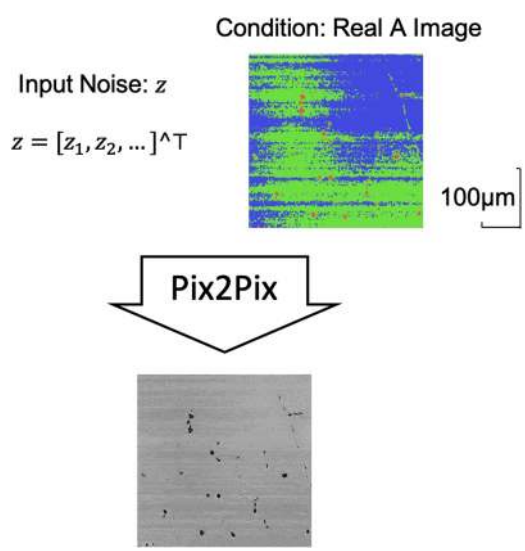

FIG. 9. Pix2Pix generative model takes a label image as input and outputs a synthetic microstructure image. In this example, the Real A image is generated from a sample with the HT1-C1 processing history.

© Author(s) 2020

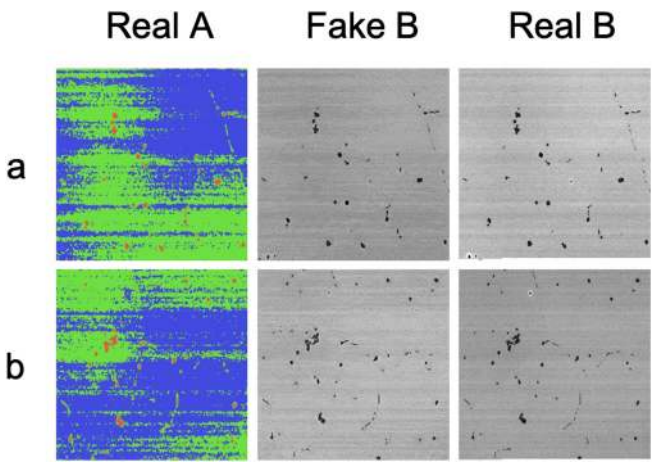

FIG. 10. Randomly sampled images from the trained Pix2Pix model, where rows (a) and (b) are different samples. From left to right, in each row (a,b) there are the following images: real A, fake B (model output), and real B (ground-truth).

B). Prior to training, real image labels are prepared, where the real images are referred to as real A (the model input). Image segmentation, such as the algorithm suggested in Ref. 31 or the one described here in Sec. II B 1, can be applied so that for each image in real B, we have a label image in real A.

We use the default model specification described in the Pix2Pix model paper.[40] After the training is finished, 50 synthetic images are randomly sampled and some of them are displayed in Fig. 10. From the sampled images, we can tell that the synthetic images generated by the Pix2Pix model are visually close to the ground-truth. With sufficient information from the label images, they are more realistic than those sampled from the progressive growing GAN. While spatial patterns are not visible in these sampled images, we apply the same measurements as we did on the progressive growing GAN as a comparison, which can be found in Sec. III B 4.

While the outputs from the Pix2Pix model are visually more realistic, they require additional information as inputs. They ignore the distribution of microstructures in the image and focus on the learning and simulating the textures in real images. Training a high-resolution GAN with interpretable conditions remains a challenge. For future studies, training a high-resolution GAN could be split into two steps in which we first focus on synthesizing the underlying representation of the microstructures (such as the label image), and then in a second step, adding texture to the image.

3. Analysis of microstructure distribution learned by the pg-GAN

In this section, we measure the differences between the microstructure represented by the real images and synthetic images. We apply the characterization pipeline introduced in Sec. II C. Considering the visual differences (image "sharpness" and local patterns) between real images and synthetic images and that original images are resized to a smaller size before being used for GAN training, we use a slightly different parameter setting (given in Table III) from the previously described to ensure the quality of image segmentation.

TABLE III. Parameters selected for microstructure characterization of synthetic images.

	Parameter	Value
Preprocessing	Sharpen	(3, 3)
	Noise-reducing	Bilateral
	Diameter	15
	Sigma color	75
	Sigma space	75
Segmentation	k-means	k-means++ center initialization[32]
	Lamellar	(7,7) closing then (7,7) opening
	UC	(5,5) opening then (3,3) closing
GPC	Kernel	Radial-basis function kernel
	Length scale	1.0
	Optimizer	L-BFGS-B algorithm

We generate 300 synthetic images by random sampling from the trained pg-GAN model. Synthetic images are segmented; area features, spatial features, and texture features are collected. The processing histories of the synthetic images are predicted by the trained model from Experiment 4 in Sec. III. The area features of Lamellar Transformation Products and UC are plotted in Fig. 11, along with the predicted processing histories. It can be found that the area features collected from real images and area features collected from synthetic images come from similar distributions.

To quantitatively measure whether the synthetic image is from the same distribution as the real images, or how well the pg-GAN has learned to represent the microstructure, we carry out the two experiments below.

From the previous experiments, we have collected area features, spatial features, and texture features from the 272 original images and the 300 synthetic images. Two models are trained to classify whether a specific feature is collected from a real image or a synthetic image. The first model uses the area features as the only input, while the second model takes all the features (area, spatial, texture) as the input. Both models use the Gaussian Process Classifier (GPC) to learn two distributions for the features from real images and synthetic images (the model specifications are summarized in Table III). With fivefold cross-validation, the accuracies for the two models are 52.6% and 52.4%, respectively. The fact that both models fail to distinguish between features collected from real images and synthetic images supports our assumption that the pg-GAN managed to learn the underlying distribution of microstructure well.

Here, we consider the area features collected from the 272 original images and the 300 synthetic images. For each pair of processing histories (e.g., HT1-C1 and HT1-C2), we train a binary classification model to classify between the area features collected from real HT1-C1 images and the area features collected from synthetic HT1-C2 images. Again, the Gaussian Process Classifier (GPC) is used as the model classifier and the classification results

© Author(s) 2020

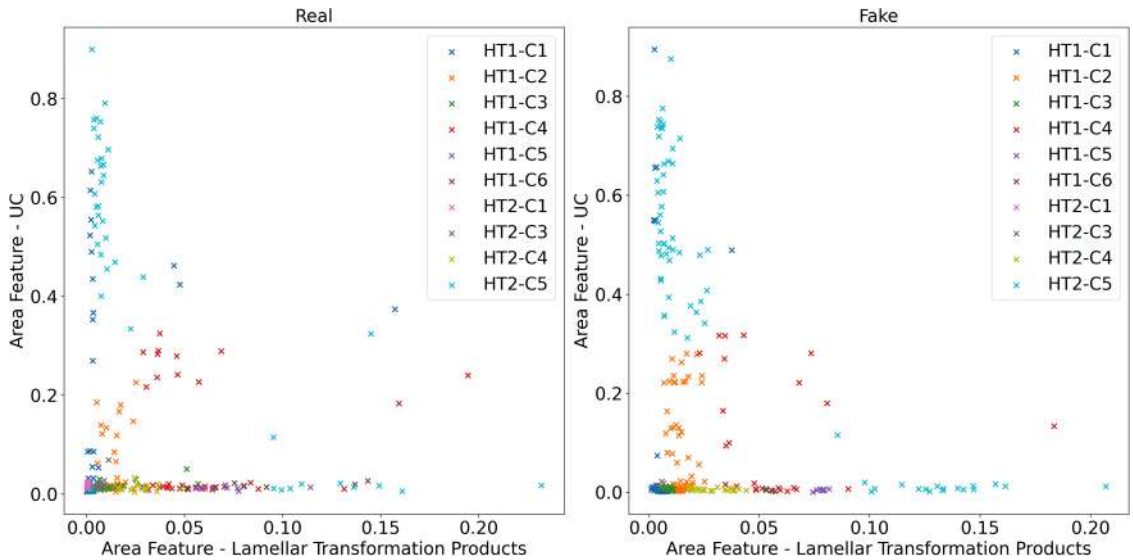

FIG. 11. Comparison between area features from real images and synthetic images. The processing histories of the synthetic images are predicted by a random forest model trained on features from real images. For interpretation of the references to color in this figure legend, the reader is referred to the online version of this article.

are reported in Fig. 12 as F1 scores with fivefold cross-validation. It is noted here that since there are no synthetic images predicted as HT2-C1 in the 300 synthetic images, the column corresponding to HT2-C1 is left empty in the matrix.

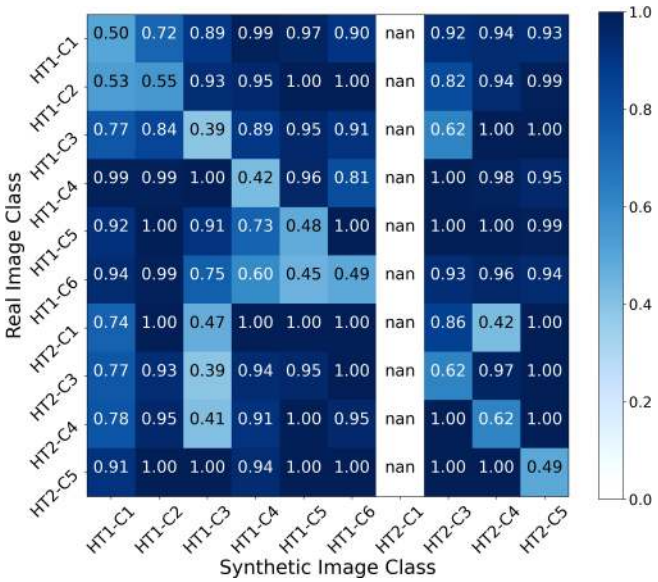

FIG. 12. Results of binary classification between area features collected from real class A images and area features collected from synthetic class B images. Model performance is measured in F1 scores with fivefold cross-validation. Notice that there are no synthetic images predicted as HT2-C1 in the 300 synthetic images.

It can be found that on the main diagonal matrix, where the real area features and synthetic area features are from the same processing histories, the F1 scores are generally low. This finding is consistent with our assumption that the pg-GAN managed to learn the underlying distribution of microstructure well. Also, the classification performance in other cells off the main diagonal is relatively high with a few exceptions. For those off diagonal cells with low F1 scores, such as the classification between HT2-C3 and HT1-C3, they are quite consistent with the classification results reported in Fig. 5. The poor classification performance for this particular example of HT2-C3 and HT1-C3 may be attributed to the very similar microstructure generated after few processing steps, where the only difference between these two conditions is the homogenization temperature and time.

By evaluating classification performance between features from real and synthetic images and between area features for different processing histories, we can compare the area features, spatial features, and texture features collected from real images and synthetic images. In previous experiments, we have shown that these features are quite good at characterizing the microstructure. Specifically for the random forest model that takes all features as the input, processing history of microstructure can be predicted with an F1 score of 95.1%. However, these features are not explicitly considered during the GAN training. From the fact that features collected from real or synthetic images are not distinguishable indicates that the pg-GAN model managed to learn the underlying distribution of microstructure well. Although the results from binary classification are highly based on the processing histories that are predicted by a random forest model and not originally given by the pg-GAN, the results can serve as a comparison with the experiment results in Sec. III and suggest that the underlying

© Author(s) 2020

representation learned by the pg-GAN is similar to the microstructure representation given by the real images.

4. Texture differences in real vs synthetic micrographs

Sub-regions of images from real microstructures and images generated by the GAN were subjected to the Discrete Fourier Transform (DFT) operation. This was performed with the objective of studying the spatial patterns exhibited by these two classes of images.

DFT samples a discrete set of frequencies corresponding to the size of the image in the spatial domain. In the Fourier domain, the intensity at frequency point (k,l) is calculated by

$$F(k,l) = \sum_{i=0}^{N-1}\sum_{j=0}^{N-1} f(i,j)\, e^{-\frac{2\pi}{N}(ki+lj)}, \tag{1}$$

where $f(i,j)$ is the pixel intensity at position (i,j) in the real space. Images exhibiting geometric or spatial patterns tend to amplify specific frequencies in the Fourier domain, and hence the Fourier transform of an image can be used to highlight spatial patterns present in the image. For example, when an image exhibits horizontal patterns seperated by a pixel distance of WIDTH/2, its Fourier transform exhibits a local maxima at $(k, l) = (2, 0)$.

In this work, DFT was implemented in Python using the OpenCV[45] and Numpy[46] libraries. The "dft" method in OpenCV was used to perform the transform, and the "fftshift" method in Numpy was used to shift the zero-frequency (DC frequency) to the center of the transform. The insets of Figs. 13 and 14 are magnitudes of the transforms that have been subjected to a log filter, in order to visualize the local maximas effectively.

Figure 13 shows a characteristic example of an image of a real microstructure and the Fourier transforms performed on the different sub-regions of the image. The sub-regions selected for this analysis are all 200 pixels × 200 pixels in size. It can be observed that the horizontal texture manifests itself in thin vertical frequency lines on the transforms.

Figure 14 shows a characteristic example of a synthetic image generated from an adversarial network and the Fourier transforms performed on the different sub-regions of the image. It can be seen that the synthetic images are characterized by strong patterns, which manifest as local maximas at several frequency points on the transformed images. The Fourier transforms of a synthetic image exhibit more local maximas than the corresponding transforms of a typical image of a real microstructure.

In order to quantitatively analyze how the magnitudes of discrete Fourier transform can be used to characterize the texture in real and synthetic images, as well as microstructure from different processing histories. We collect the magnitudes of the transform as a feature vector, referred to as DFT features. More specifically, after the DFT process, we take the magnitudes from the 9 × 9 region in the center and flatten them into a one-dimensional feature vector. We collect the DFT features from the 272 real images and the 300 synthetic images, which are prepared in Sec. III B 3. Two experiments are carried out as follows:

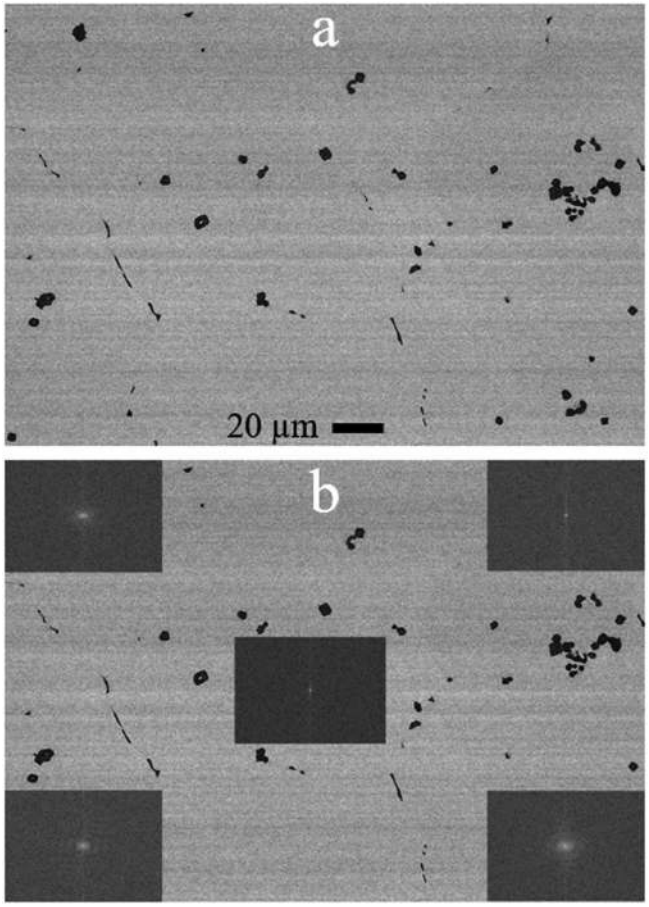

FIG. 13. (a) A characteristic image of a real microstructure, showing visible texture in the horizontal direction. (b) Fourier transforms of different sub-regions of size 200 × 200, superimposed upon the original image at the corresponding region. The background pattern is reflected in the thin vertical lines on the transform plots.

1. Using the DFT features only, we train a Gaussian process classifier model to predict whether a DFT feature vector is collected from real microstructure images or synthetic microstructure images. With fivefold cross-validation, the accuracy of the model is 99.4%.
2. For each pair of processing histories (e.g., HT1-C1 and HT1-C2), we train a Gaussian process classifier model to predict if a DFT feature vector is collected from real HT1-C1 images and real HT1-C2 images. Similar experiments are conducted between real HT1-C1 images and synthetic HT1-C2 images and between synthetic HT1-C1 images and synthetic HT1-C2 images. Model performance is measured in F1 scores, as reported in Fig. 15.

It can be concluded that as a feature vector, the DFT magnitudes can discriminate between real and synthetic images well, but fail to characterize microstructure from different processing

© Author(s) 2020

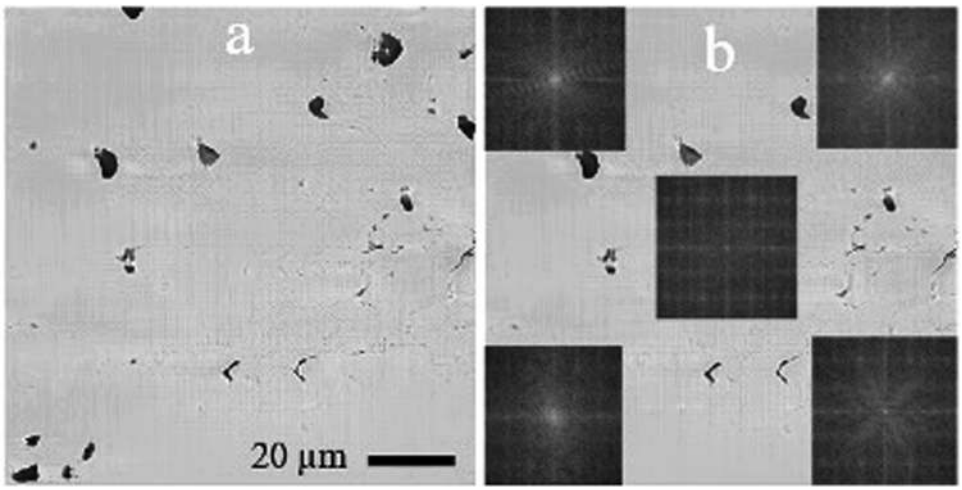

FIG. 14. (a) A characteristic image of a microstructure generated by pg-GAN. (b) Fourier transforms of different sub-regions of size 200 × 200, superimposed upon the original image at the corresponding region. Though these synthetic images show strong background features, the presence of these features does not bias the discriminator.

histories. Considering the synthetic images generated from the pg-GAN model, the DFT features are good indicators that the synthetic images are not perfect.

It should be noted that the synthetic image used for the analysis above was generated after the adversarial network had reached a steady state (i.e., at a point when the network can no longer distinguish between a real and a synthetic image). Thus, it is reasonable to conclude that though the images generated by GANs exhibit spatial patterns that are not present in the training images, the presence of these spurious patterns is not significant enough to bias the discriminator. The likelihood of the occurrence of such patterns must be taken into consideration for classification problems such as texture detection in a microstructure dataset.

Further, artifacts may be introduced into images from sample preparation (i.e., scratches from polishing media), imaging (e.g., blurring or periodic patterns in the image due to a lack of a conducting path for electrons, i.e., charging), and up-sampling in GAN pipelines (i.e., checkerboard patterns in generated images[47,48]). Such artifacts due to up-sampling for the purposes of this work are considered minor, as they would not affect final classification accuracy of micrographs.

IV. CHALLENGES AND BEST PRACTICES

Several challenges associated with recognition and quantification of microstructure image data exist due to various limitations inherent in materials science studies. The two major challenges are limited size of datasets and imbalanced datasets. Many machine learning (particularly deep learning) models require large datasets for training. A deep learning model generalizes to test datasets

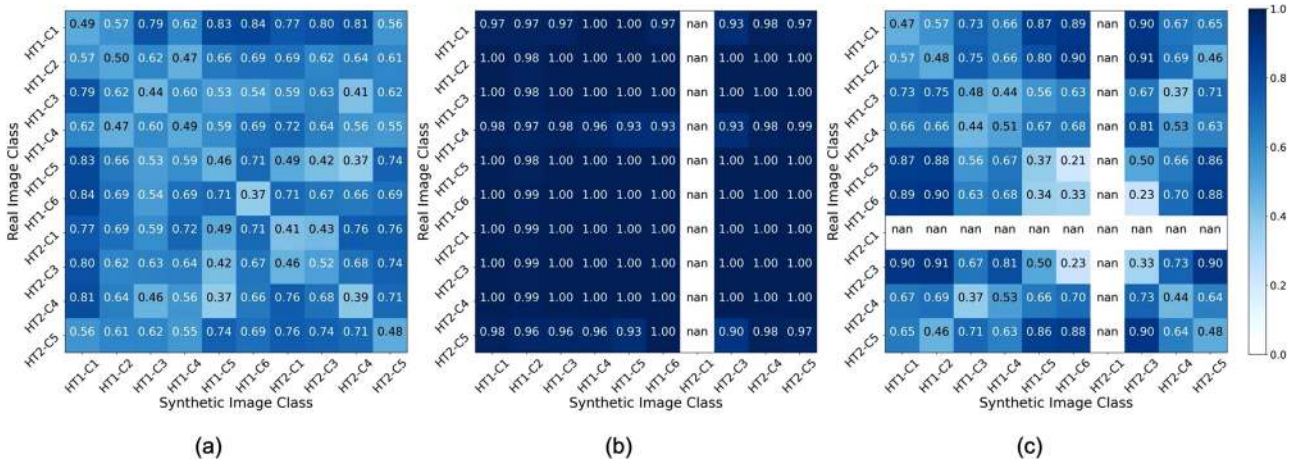

FIG. 15. Binary classification between microstructure images from different processing histories based on DFT features. The model performance is measured in F1 scores. From left to right, the matrices measure the model performance between (a) real images and real images, (b) real images and synthetic images, and (c) synthetic images and synthetic images.

© Author(s) 2020

better as the size of training set is increased. In a typical environment for microstructure imaging, generating high quality micrographs in large numbers is dependent upon metallographic sample preparation, microscope operator skill, facilities, and time. For objectives such as multi-class classification, training from real data sets could lead to a situation where few classes have a disproportionate number of images, highlighted in prior work,[21] and in the U-10Mo dataset studied here. Imbalanced datasets can result in a reduced classification accuracy for the class with the disproportionately lower number of images, even if the overall accuracy is within acceptable tolerance. With respect to the case study of GAN seen earlier in the study, an imbalanced dataset may also result in a reduced significance for the spatial patterns present in the disproportionate class among the patterns exhibited by the synthetically generated images. Given these challenges, the authors suggest several best practices, in addition to the current work, to produce meaningful results from an image driven machine learning approach to microstructure recognition and quantification:

1. Shallow learning and conventional learning techniques: Convolutional neural networks and other algorithms such as SVMs or ensemble classifiers are capable of performance within an acceptable tolerance for simple classification problems.
2. Semantic segmentation: DeCost *et al.*[13] demonstrated the use of the semantic segmentation algorithm to isolate objects in the microstructure dataset.
3. Serialization of techniques through a task pipeline: Prior work has demonstrated a method for quantitative feature extraction by automating the algorithm selection process through a task pipeline.[21]
4. Dataset augmentation: The user may artificially populate the training dataset by methods such as cropping, rotating, and adding uncorrelated noise to the original images in order to ensure that the training process generalizes reasonably well to test datasets.
5. Automated algorithm selection and hyperparameter optimization in machine learning: Given the size and scale of datasets in material science, automatic methods for selection of algorithms and hyperparameters[49–54] can prove to be a viable option for fine tuning the parameters of algorithms and improve generalization performance as much as possible given the scarcity of data to learn from.

V. CONCLUSIONS

In this work, we perform multi-class classification for the purpose of linking microstructure to processing condition. The original dataset consists of micrographs for ten different thermo-mechanical processing conditions of a U-10Mo alloy. We evaluate the classification model performance for different microstructure reperesentations, and the results reveal that area, spatial, and texture information are needed for accurately describing image data. Using this newly developed microstructure representation, an F1 score of 95.1% was achieved for distinguishing between micrographs corresponding to ten different thermo-mechanical material processing conditions. Generative adversarial networks were also explored to better understand if synthetic image data could be used to supplement small, imbalanced original image datasets. Two different networks were trained and tested to assess the performance: progressive growing GAN and Pix2Pix GAN. We find that the progressive growing GAN introduces spatial patterns that are not present in original image data. Texture detection in a microstructure dataset might be adversely affected by the presence of such spurious patterns. Our work highlights that semantically meaningful segmentation alone may be insufficient in representing image data, particularly as the complexity of material processing and the resultant microstructure increase. Hence, the need for predictive or generative methods is a frontier in materials science and engineering and should be leveraged in future studies to accelerate the materials design and characterization process.

AUTHORS' CONTRIBUTIONS

W.M. performed all machine learning methods development. E.J.K. generated a portion of image data, assisted in interpretation of results, and led manuscript writing with W.M. V.J. contributed to discussions of results and significance of microstructure evolution in the U-Mo system. A.B. analyzed synthetic image data provided by W.M. A.C. contributed to discussion of machine learning methods and results. B.Y. and D.J.L. conceptualized and directed research performed. All authors contributed to manuscript preparation.

ACKNOWLEDGMENTS

Data analysis work was conducted at Rensselaer Polytechnic Institute. Experimental work was conducted at Pacific Northwest National Laboratory (PNNL) operated by Battelle for the United States Department of Energy under Contract No. DE-AC05-76RL01830. A portion of this work was funded by the U.S. Department of Energy National Nuclear Security Administration's Office of Material Management and Minimization. The authors also wish to acknowledge personnel at PNNL for experimental work done to generate data used in this work, in particular, the following individuals: Shelly Carlson, Mark Rhodes, and Jesse Lang (PNNL) for metallographic sample preparation, and Alan Schemer-Kohrn (PNNL) for expertise in electron microscopy and image acquisition of image data used in this work. The authors also gratefully acknowledge Ms. Sarah Reehl for helpful discussions and review of the manuscript.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

[1] *Reducing the Use of Highly Enriched Uranium in Civilian Research Reactors* (The National Academies Press, Washington, DC, 2016).
[2] S. Van Den Berghe and P. Lemoine, "Review of 15 years of high-density low-enriched UMo dispersion fuel development for research reactors in Europe," Nucl. Eng. Technol. **46**, 125–146 (2014).
[3] J. Snelgrove, G. Hofman, M. Meyer, C. Trybus, and T. Wiencek, "Development of very-high-density low-enriched-uranium fuels," Nucl. Eng. Des. **178**, 119–126 (1997).
[4] M. Meyer, J. Gan, J. Jue, D. Keiser, E. Perez, A. Robinson, D. Wachs, N. Woolstenhulme, G. Hofman, and Y. Kim, "Irradiation performance of u-mo monolithic fuel," Nucl. Eng. Technol. **46**, 169–182 (2014).
[5] A. E. Dwight, "The uranium-molybdenum equilibrium diagram below 900 °C," J. Nucl. Mater. **2**, 81–87 (1960).

© Author(s) 2020

[6]A. Devaraj, L. Kovarik, E. Kautz, B. Arey, S. Jana, C. Lavender, and V. Joshi, "Grain boundary engineering to control the discontinuous precipitation in multi-component U10Mo alloy," Acta Mater. **151**, 181–190 (2018).
[7]A. Devaraj, E. Kautz, L. Kovarik, S. Jana, N. Overman, C. Lavender, and V. V. Joshi, "Phase transformation of metastable discontinuous precipitation products to equilibrium phases in U10Mo alloys," Scr. Mater. **156**, 70–74 (2018).
[8]S. Jana, N. Overman, T. Varga, C. Lavender, and V. Joshi, "Phase transformation kinetics in rolled U-10 wt. % Mo foil: Effect of post-rolling heat treatment and prior γUMo grain size," J. Nucl. Mater. **496**, 215–226 (2017).
[9]E. Kautz, D. Burkes, V. Joshi, C. Lavender, and A. Devaraj, "Nanoscale spatially resolved mapping of uranium enrichment," Sci. Rep. **9**, 12302 (2019).
[10]R. Ramprasad, R. Batra, G. Pilania, A. Mannodi-Kanakkithodi, and C. Kim, "Machine learning in materials informatics: Recent applications and prospects," NPJ Comput. Mater. **3**, 60 (2017).
[11]J. Ling, M. Hutchinson, E. Antono, B. DeCost, E. A. Holm, and B. Meredig, "Building data-driven models with microstructural images: Generalization and interpretability," Mater. Discovery **10**, 19–28 (2017).
[12]A. Chowdhury, E. Kautz, B. Yener, and D. Lewis, "Image driven machine learning methods for microstructure recognition," Comput. Mater. Sci. **123**, 176–187 (2016).
[13]B. L. DeCost, T. Francis, and E. A. Holm, "High throughput quantitative metallography for complex microstructures using deep learning: A case study in ultrahigh carbon steel," Microsc. Microanal. **25**, 21–29 (2019).
[14]B. L. DeCost and E. A. Holm, "A computer vision approach for automated analysis and classification of microstructural image data," Comput. Mater. Sci. **110**, 126–133 (2015).
[15]B. L. DeCost, H. Jain, A. D. Rollett, and E. A. Holm, "Computer vision and machine learning for autonomous characterization of am powder feedstocks," JOM **69**, 456–465 (2017).
[16]K. T. Butler, D. W. Davies, H. Cartwright, O. Isayev, and A. Walsh, "Machine learning for molecular and materials science," Nature **559**, 547–555 (2018).
[17]S. V. Kalinin, B. G. Sumpter, and R. K. Archibald, "Big–deep–smart data in imaging for guiding materials design," Nat. Mater. **14**, 973–980 (2015).
[18]A. J. Ballard, R. Das, S. Martiniani, D. Mehta, L. Sagun, J. D. Stevenson, and D. J. Wales, "Perspective: Energy landscapes for machine learning," Phys. Chem. Chem. Phys. **19**, 12585–12603 (2017).
[19]G. Pilania, C. Wang, X. Jiang, S. Rajasekaran, and R. Ramprasad, "Accelerating materials property predictions using machine learning," Sci. Rep. **3**, 419 (2013).
[20]R. Kondo, S. Yamakawa, Y. Masuoka, S. Tajima, and R. Asahi, "Microstructure recognition using convolutional neural networks for prediction of ionic conductivity in ceramics," Acta Mater. **141**, 29–38 (2017).
[21]A. Baskaran, G. Kane, K. Biggs, R. Hull, and D. Lewis, "Adaptive characterization of microstructure dataset using a two stage machine learning approach," Comput. Mater. Sci. **177**, 109593 (2020).
[22]A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems 25*, edited by F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger (Curran Associates, Inc., 2012), pp. 1097–1105.
[23]J. Deng, W. Dong, R. Socher, L. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition* (IEEE, 2009), pp. 248–255.
[24]B. M. Lake, R. Salakhutdinov, and J. B. Tenenbaum, "Human-level concept learning through probabilistic program induction," Science **350**, 1332–1338 (2015).
[25]E. J. Kautz, A. R. Hagen, J. M. Johns, and D. E. Burkes, "A machine learning approach to thermal conductivity modeling: A case study on irradiated uranium-molybdenum nuclear fuels," Comput. Mater. Sci. **161**, 107–118 (2019).
[26]E. K. Mace, J. D. Ward, and C. E. Aalseth, "Use of neural networks to analyze pulse shape data in low-background detectors," J. Radioanal. Nucl. Chem. **318**, 117–124 (2018).
[27]I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Advances in Neural Information Processing Systems 27*, edited by Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger (Curran Associates, Inc., 2014), pp. 2672–2680.
[28]Z. Pan, W. Yu, X. Yi, A. Khan, F. Yuan, and Y. Zheng, "Recent progress on generative adversarial networks (GANs): A survey," IEEE Access **7**, 36322–36333 (2019).
[29]S. Jana, A. Devaraj, L. Kovarik, B. Arey, L. Sweet, T. Varga, C. Lavender, and V. Joshi, "Kinetics of cellular transformation and competing precipitation mechanisms during sub-eutectoid annealing of U10Mo alloys," J. Alloys Compd. **723**, 757–771 (2017).
[30]R. Prabhakaran, V. V. Joshi, M. A. Rhodes, A. L. Schemer-Kohrn, A. D. Guzman, and C. A. Lavender, "U-10mo sample preparation and examination using optical and scanning electron microscopy," Tech. Rep. PNNL-25308 (Pacific Northwest National Laboratory (PNNL), Richland, WA, 2016).
[31]E. Kautz, W. Ma, S. Jana, A. Devaraj, V. Joshi, B. Yener, and D. Lewis, "An image-driven machine learning approach to kinetic modeling of a discontinuous precipitation reaction," Mater. Charact. **166**, 110379 (2019).
[32]D. Arthur and S. Vassilvitskii, "K-means++: The advantages of careful seeding," in *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA-07 (Society for Industrial and Applied Mathematics, 2007), pp. 1027–1035.
[33]R. M. Haralick, S. R. Sternberg, and X. Zhuang, "Image analysis using mathematical morphology," IEEE Trans. Pattern Anal. Mach. Intell. **9**, 532–550 (1987).
[34]X. Chen, Y. Duan, R. Houthooft, J. Schulman, I. Sutskever, and P. Abbeel, "Infogan: Interpretable representation learning by information maximizing generative adversarial nets," in *Advances in Neural Information Processing Systems 29*, edited by D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, and R. Garnett (Curran Associates, Inc., 2016), pp. 2172–2180.
[35]J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu, and T. S. Huang, "Generative image inpainting with contextual attention," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition* (IEEE, 2018), pp. 5505–5514.
[36]Z. Yang, X. Li, L. Catherine Brinson, A. N. Choudhary, W. Chen, and A. Agrawal, "Microstructural materials design via deep adversarial learning methodology," J. Mech. Des. **140**, 111416 (2018).
[37]X. Li, Z. Yang, L. C. Brinson, A. Choudhary, A. Agrawal, and W. Chen, "A deep adversarial learning methodology for designing microstructural material systems," in *ASME 2018 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference* (American Society of Mechanical Engineers Digital Collection, 2018).
[38]T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen, "Improved techniques for training GANs," in *Proceedings of the 30th International Conference on Neural Information Processing Systems*, NIPS'16 (Curran Associates Inc., Red Hook, NY, 2016), pp. 2234–2242.
[39]T. Karras, T. Aila, S. Laine, and J. Lehtinen, "Progressive growing of GANs for improved quality, stability, and variation," arXiv:1710.10196 [cs.NE] (2017).
[40]P. Isola, J. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (IEEEE, 2017), pp. 5967–5976.
[41]N. Bonneel, J. Rabin, G. Peyré, and H. Pfister, "Sliced and Radon Wasserstein barycenters of measures," J. Math. Imaging Vis. **51**, 22–45 (2014).
[42]M. Mirza and S. Osindero, "Conditional generative adversarial nets," arXiv:1411.1784 [cs.LG] (2014).
[43]J. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *2017 IEEE International Conference on Computer Vision* (IEEE, 2017), pp. 2242–2251.
[44]T. Park, M. Liu, T. Wang, and J. Zhu, "Semantic image synthesis with spatially-adaptive normalization," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition* (IEEE, 2019), pp. 2332–2341.
[45]G. Bradski, "The OpenCV library," Dr. Dobb's J. Software Tools 120, 122–125 (2000).
[46]T. E. Oliphant, *A Guide to NumPy* (Trelgol Publishing, 2006), Vol. 1.
[47]X. Zhang, S. Karaman, and S. Chang, "Detecting and simulating artifacts in gan fake images," in *2019 IEEE International Workshop on Information Forensics and Security (WIFS)* (IEEE, 2019), pp. 1–6.

© Author(s) 2020

[48]A. Odena, V. Dumoulin, and C. Olah, "Deconvolution and checkerboard artifacts," Distill (2016). doi:10.23915/distill.00003.
[49]A. Chowdhury, M. Magdon-Ismail, and B. Yener, "Quantifying error contributions of computational steps, algorithms and hyperparameter choices in image classification pipelines," arXiv preprint arXiv:1903.02521 (2019).
[50]F. Hutter, L. Kotthoff, and J. Vanschoren, *Automated Machine Learning* (Springer, 2019).
[51]L. Kotthoff, C. Thornton, H. H. Hoos, F. Hutter, and K. Leyton-Brown, "Auto-WEKA 2.0: Automatic model selection and hyperparameter optimization in WEKA," J. Mach. Learn. Res. **18**, 826–830 (2017).
[52]Y. Bengio, F. Bastien, A. Bergeron, N. Boulanger-Lewandowski, T. Breuel, Y. Chherawala, M. Cisse, M. Côté, D. Erhan, J. Eustache, X. Glorot, X. Muller, S. Pannetier Lebeuf, R. Pascanu, S. Rifai, F. Savard, and G. Sicard, "Deep learners benefit more from out-of-distribution examples," J. Mach. Learn. Res. Proc. Track **15**, 164–172 (2011).
[53]J. S. Bergstra, R. Bardenet, Y. Bengio, and B. Kégl, "Algorithms for hyper-parameter optimization," in *Advances in Neural Information Processing Systems* (Neural Information Processing Systems Foundation, Inc., 2011), pp. 2546–2554.
[54]L. Li, K. Jamieson, G. DeSalvo, A. Rostamizadeh, and A. Talwalkar, "Hyperband: A novel bandit-based approach to hyperparameter optimization," J. Mach. Learn. Res. **18**, 6765–6816 (2017).

© Author(s) 2020