NASA/TM-20210026119

Microstructure Segmentation With Deep Learning Encoders Pre-Trained on a Large Microscopy Dataset

Joshua Stuckner, Bryan Harder, and Timothy M. Smith
Glenn Research Center, Cleveland, Ohio

January 2022

NASA STI Program . . . in Profile

Since its founding, NASA has been dedicated to the advancement of aeronautics and space science. The NASA Scientific and Technical Information (STI) Program plays a key part in helping NASA maintain this important role.

The NASA STI Program operates under the auspices of the Agency Chief Information Officer. It collects, organizes, provides for archiving, and disseminates NASA's STI. The NASA STI Program provides access to the NASA Technical Report Server—Registered (NTRS Reg) and NASA Technical Report Server—Public (NTRS) thus providing one of the largest collections of aeronautical and space science STI in the world. Results are published in both non-NASA channels and by NASA in the NASA STI Report Series, which includes the following report types:

- TECHNICAL PUBLICATION. Reports of completed research or a major significant phase of research that present the results of NASA programs and include extensive data or theoretical analysis. Includes compilations of significant scientific and technical data and information deemed to be of continuing reference value. NASA counter-part of peer-reviewed formal professional papers, but has less stringent limitations on manuscript length and extent of graphic presentations.

- TECHNICAL MEMORANDUM. Scientific and technical findings that are preliminary or of specialized interest, e.g., "quick-release" reports, working papers, and bibliographies that contain minimal annotation. Does not contain extensive analysis.

- CONTRACTOR REPORT. Scientific and technical findings by NASA-sponsored contractors and grantees.

- CONFERENCE PUBLICATION. Collected papers from scientific and technical conferences, symposia, seminars, or other meetings sponsored or co-sponsored by NASA.

- SPECIAL PUBLICATION. Scientific, technical, or historical information from NASA programs, projects, and missions, often concerned with subjects having substantial public interest.

- TECHNICAL TRANSLATION. English-language translations of foreign scientific and technical material pertinent to NASA's mission.

For more information about the NASA STI program, see the following:

- Access the NASA STI program home page at http://www.sti.nasa.gov

- E-mail your question to help@sti.nasa.gov

- Fax your question to the NASA STI Information Desk at 757-864-6500

- Telephone the NASA STI Information Desk at 757-864-9658

- Write to:
  NASA STI Program
  Mail Stop 148
  NASA Langley Research Center
  Hampton, VA 23681-2199

NASA/TM-20210026119

Microstructure Segmentation With Deep Learning Encoders Pre-Trained on a Large Microscopy Dataset

Joshua Stuckner, Bryan Harder, and Timothy M. Smith
Glenn Research Center, Cleveland, Ohio

National Aeronautics and
Space Administration

Glenn Research Center
Cleveland, Ohio 44135

January 2022

Acknowledgments

This work was supported by the NASA Transformational Tools and Technologies (T3) project under the Transformative Aeronautics Concept Program within the Aeronautics Research Mission Directorate. Computational resources were provided by the Scientific Computing and Visualization Team in the Information and Applications Division within the Office of the CIO at NASA Glenn Research Center. In-house MicroNet images, many of which were captured using facilities provided by the NASA Glenn ASG lab, were provided by researchers from the Materials and Structures Division at NASA Glenn.

This work was sponsored by the
Transformative Aeronautics Concepts Program.

Level of Review: This material has been technically reviewed by technical management.

Available from

NASA STI Program
Mail Stop 148
NASA Langley Research Center
Hampton, VA 23681-2199

National Technical Information Service
5285 Port Royal Road
Springfield, VA 22161
703-605-6000

This report is available in electronic form at http://www.sti.nasa.gov/ and http://ntrs.nasa.gov/

Microstructure Segmentation With Deep Learning Encoders Pre-Trained on a Large Microscopy Dataset

Joshua Stuckner, Bryan Harder, and Timothy M. Smith
National Aeronautics and Space Administration
Glenn Research Center
Cleveland, Ohio 44135

Abstract

This study examined the improvement of microscopy segmentation accuracy by transfer learning from a large dataset of microscopy images called MicroNet. Many neural network encoder architectures, including VGG, Inception, and ResNet, were trained on over 100,000 labelled microscopy images from 54 classes. These pre-trained encoders were then embedded into multiple segmentation architectures including U-Net and DeepLabV3+ to evaluate segmentation performance on newly created benchmark microscopy datasets. Compared to ImageNet pre-training, models pre-trained on MicroNet generalized better to out-of-distribution micrographs taken under different imaging and sample conditions and were more accurate with less training data. When training with only a single Ni-superalloy image, pre-training on MicroNet produced a 72.2 percent reduction in relative segmentation error. These results suggest that transfer learning from large in-domain datasets generate models with learned feature representations that are more useful for downstream tasks and will likely improve any microscopy image analysis technique that can leverage pre-trained encoders.

Introduction

Establishing processing-structure-property (PSP) relationships is critical to the design and improvement of materials. Microscopy image segmentation is often the first and hardest step in quantifying material structure, which is the central link in PSP relationships. Traditional microstructure quantification requires numerous manual measurements on a micrograph (e.g., Refs. 1 and 2), is tedious, time-consuming, and prone to bias. Automatic segmentation using classic computer vision techniques such as image thresholding and morphology operations (Refs. 3 and 4) is much faster and repeatable, but difficult to implement and often not robust to slight changes in imaging or sample conditions. Recently, convolutional neural networks (CNN) pre-trained on ImageNet (Ref. 5) have produced superior microscopy segmentation results and are much easier to implement (Refs. 6 to 12). However, segmentation CNNs require expensively labeled training data to operate well and ImageNet pre-training does not adequately alleviate this problem when transferred to microscopy segmentation tasks because many of the learned filters are not applicable (e.g., those adapted to detect dogs). Therefore, we created MicroNet, a large dataset containing over 100,000 labelled microscopy images. Here we show that leveraging transfer learning from classification models pre-trained on MicroNet rather than ImageNet produces segmentation models that perform better with less training data and are more successful at segmenting micrographs with imaging or sample conditions that were not seen in the training data.

Semantic segmentation with CNNs is performed with encoder-decoder type architectures, which offer state-of-the-art performance on benchmark datasets such as the cityscapes dataset (Ref. 13). The encoder uses learned convolutional filters to extract semantic information from the input image, transforming the image data into a latent representation vector. The decoder then maps the extracted information to each pixel location in the image to generate a pixelwise classification prediction of the objects in original image (i.e., semantic segmentation).

Training data for semantic segmentation is expensive and time consuming to create. Transfer learning can be used to supplement small training data by transferring learned parameters from a model trained on another similar task, such as image classification, which is significantly easier to create training data for. To leverage transfer learning with encoder-decoder architectures, a CNN is first trained to perform image classification on a large image dataset as shown in the schematic in Figure 1. The classification model uses convolutional layers to extract semantic information from the image and a classification head, which contains several fully connected neural network layers, to make a classification prediction based on the extracted feature representation. The convolutional layers learn to extract useful features for classification during training by learning useful image filters. These learned filters are likely to be useful for other image analysis tasks such as segmentation. Transfer learning is applied when the trained convolution layers from an image classification model are copied directly to the encoder in an encoder-decoder segmentation model. ImageNet contains images of everyday life and is a common source of pre-training image classification data. The convolutional filters that are used to classify ImageNet images have also been applied very successfully to microscopy segmentation. However, recent

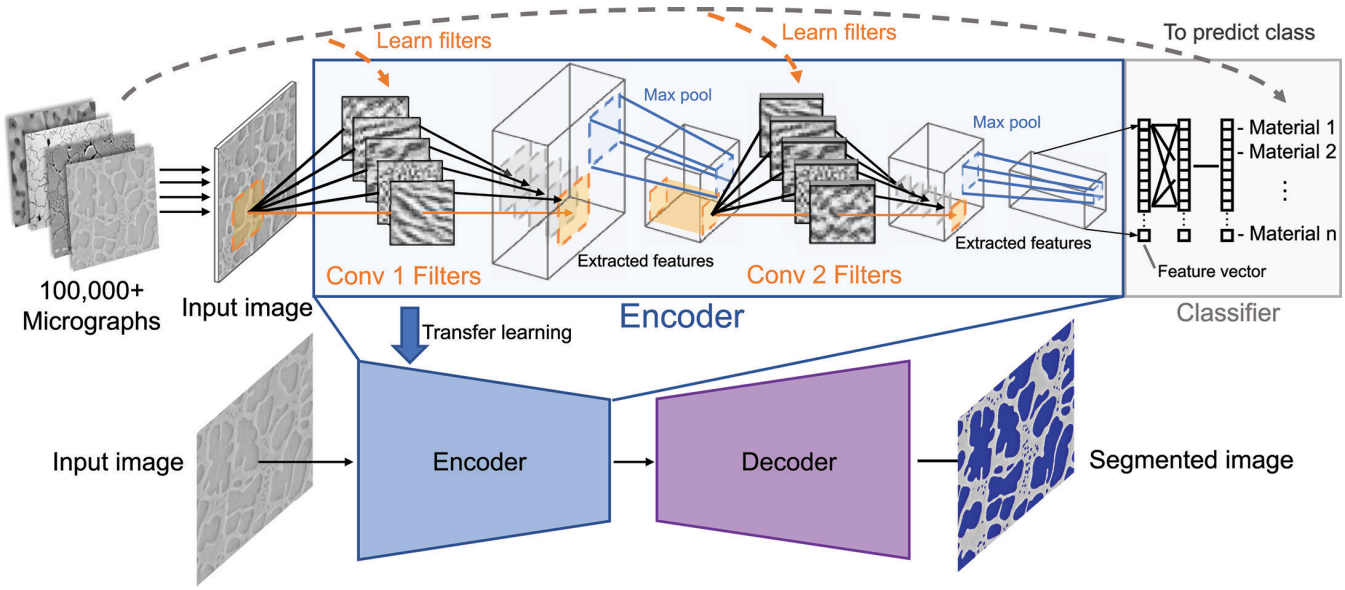

Figure 1.—Schematic of pre-training CNN encoders on MicroNet and embedding in a segmentation model through transfer learning. First, a classification model (top) with a convolutional encoder (blue box) and dense classifier head (gray box) is trained to predict the class of each material by learning filters (Conv filters, orange) which extract relevant features into a feature vector. Through transfer learning (blue arrow) the learned filters are copied into an encoder-decoder segmentation model (bottom) which learns to segment microscopy images with less training data than without transfer learning.

work has shown that the first few layers of VGG models (a powerful early CNN classification model that is still widely used) are highly useful for transfer learning to microscopy segmentation while the deeper layers are not (Ref. 14). This is because the initial layers detect simple features like edges, corners, and simple textures, which are likely to appear in microscopy images, while the deeper layers detect higher level features such as dog ears and car tires, which do not appear in microscopy images.

The central hypothesis of the work presented here is that because higher level feature detectors from models trained on ImageNet do not transfer well to microscopy segmentation, the full advantages of transfer learning are not realized. Therefore, we trained classification models on a large dataset of microscopy images called MircoNet so that the models could learn to detect higher level microstructure features that do not appear in pictures of everyday life such as grain boundaries, precipitates, and oxide layers. We show that transfer learning from models trained on MicroNet rather than ImageNet produces more accurate segmentation results with less training data and is more robust to changing imaging and sample conditions.

Methods

Description of Datasets

A large dataset called MicroNet, containing 110,861 microscopy images, was created to pre-train classification models to be used as encoders in segmentation models. The majority of MicroNet images were sourced inhouse with additional images from the UltraHigh Carbon Steel Micrograph DataBase (Ref. 15), the Aversa Scanning Electron Microscopy (SEM) dataset (Ref. 16), synthetic SEM powder data (Ref. 17), SEM images from the Materials Data Repository hosted by the National Institute of Standards and Technology, and a photovoltaic dataset (Ref. 18). MicroNet contains 54 classes and was split into train/validation sets with 50 images for each class in the validation set. The training set had some class imbalance with several classes containing less than 200 images and one class containing more than 10,000 images. Most classes had over 1,000 images. On average MicroNet images were much larger than ImageNet (1048×741 vs. 469×387 pixels) giving the MicroNet dataset a pixel equivalence of 474,323 ImageNet images for the encoders to learn from.

The segmentation algorithms were tested on two sets of material micrographs: SEM images of a Ni-superalloy and cross-sectional SEM images of a SiC/SiC environmental barrier coating (EBC) with a thermally grown oxide layer. The Ni-superalloy had three classes to segment: a matrix phase, secondary precipitates, and tertiary precipitates. The EBC had two classes: an oxide layer and the background (not oxide layer).

Training Classification Models

Many CNN classification models were trained on MicroNet to use as segmentation encoders through transfer learning. Models for each architecture were initialized with weights from

pre-training on ImageNet. Additional models for most of the classification architectures were also initialized with random weights following Kaiming initialization to evaluate the effect of encoder training on MicroNet from scratch. The Kaiming initialization was designed to reduce the exploding or vanishing gradient problem by encouraging the variance of activations to be similar across network layers when using rectified linear unit (ReLU) activation functions (Refs. 19 and 20). During training, all images were preprocessed by mean centering and normalizing each channel according to the ImageNet statistics in order to best utilize pre-trained weights (Ref. 21). Image transformations were used to augment the training data set including random cropping, random resizing, horizontal and vertical flipping, rotation, photometric distortions, and added noise. Training was performed using the PyTorch Python library (Ref. 22) in a similar fashion to Tan et al. (Ref. 23). Optimization was performed with stochastic gradient decent with a momentum of 0.9 and an initial learning rate of 0.1 that decayed by 10 percent every 30 epochs. Weight decay was 1e-4. A batch size of 1024 was used where possible and reduced for larger models due to hardware memory constraints. Models were trained until there was no improvement to the validation score using early stopping with a patience of 30 epochs. The following encoder architectures were tested in this work: VGG (Ref. 24) (with and without batch normalization (Ref. 25)), DenseNet (Ref. 26), dual path networks (dpn) (Ref. 27), EfficientNet (Ref. 28), ResNet (Ref. 29), Inception-V4 (Ref. 30), Inception-Resnet-V2 (Ref. 30), Xception (Ref. 31), MobileNet-V2 (Ref. 32), ResNeXt (Ref. 33), and SE-Net (Ref. 34).

Training Segmentation Models

Segmentation models were trained using PyTorch (Ref. 22) and the segmentation models library (Ref. 35). Training data images were converted to color and each channel was normalized and mean centered in the same manner as the classification data. Training data augmentation included random cropping to 512×512 pixels; random changes to contrast, brightness, and gamma; and added blur or image sharpening. The superalloy data was also randomly flipped vertically and horizontally and rotated while the EBC data was only horizontally flipped to preserve orientation significance. While not applied here, random resizing could be included when desired to make the models robust to changes in magnification or image resolution. The Adam (Ref. 36) optimizer was used during training with a learning rate of 2e-4 until there was no improvement on the validation dataset for 30 epochs followed by training with a learning rate of 1e-5 until early stopping after another 30 epochs with no validation improvement. The model validation metric to determine early stopping and compare different models was intersection over union (IoU). The loss function was a weighted sum of balanced cross entropy (BCE) and dice loss (Ref. 37) with a 70 percent weighting towards BCE. Initial testing showed better results with the combined loss function than either independently or using IoU as the loss function directly. This is likely because BCE has more stable gradients while dice loss is more robust to imbalanced classes and similar to the real objective of maximizing IoU. The following decoder architectures were tested: Unet (Ref. 38), Unet++ (Ref. 39), Linknet (Ref. 40), FPN (Ref. 41), PSPNet (Ref. 42), PAN (Ref. 43), and DeepLabV3+ (Refs. 44 and 45).

Results and Discussion

Classification Results

Seventy-six models were trained to classify MicroNet images into one of 54 classes. 40 models were initially pre-trained on ImageNet before training on MicroNet and 36 were randomly initialized with Kaiming initialization and trained from scratch (VGG-11, VGG-13, EfficientNet-b6, and EfficientNet-b7 architectures were not trained from scratch). Pre-training on ImageNet allows useful features learned from classifying natural images to be reused for microscopy classification through transfer learning. The classification accuracy of these models on the MicroNet validation set are shown in Figure 2. The best classification model was EfficientNet-b4 pre-trained on ImageNet, which achieved a top 1 classification accuracy of 94.5 percent. SENet-154 achieved the highest accuracy of the models trained from scratch with an accuracy of 94.0 percent. The EfficientNet, ResNet, and VGG models showed a strong benefit from pre-training on ImageNet. However, it is interesting to note that some architectures, including the squeeze and excitation (SE) and inception families of models, showed no benefit from initial pre-training on ImageNet. For the DenseNet and MobileNet architectures, pre-training was detrimental. There is no obvious trend between model size and the benefit of ImageNet pre-training for classification accuracy.

An open question is whether progress in developing vision models, which are optimized for natural images, will transfer well to microscopy images or if architecture design is overfit to natural images. Figure 2(b) shows that in general, models that performed better on ImageNet tended to perform better on MicroNet. However, a notable exception to the trend is the EfficientNet family. A significant amount of testing was done during the development of EfficientNet to optimize the depth, width, resolution scaling, and other hyperparameters to develop an architecture that performed well on ImageNet (Ref. 28). A microscopy dataset of comparable size to ImageNet would be helpful to study the full extent to which progress on natural image processing transfers to microscopy image processing.

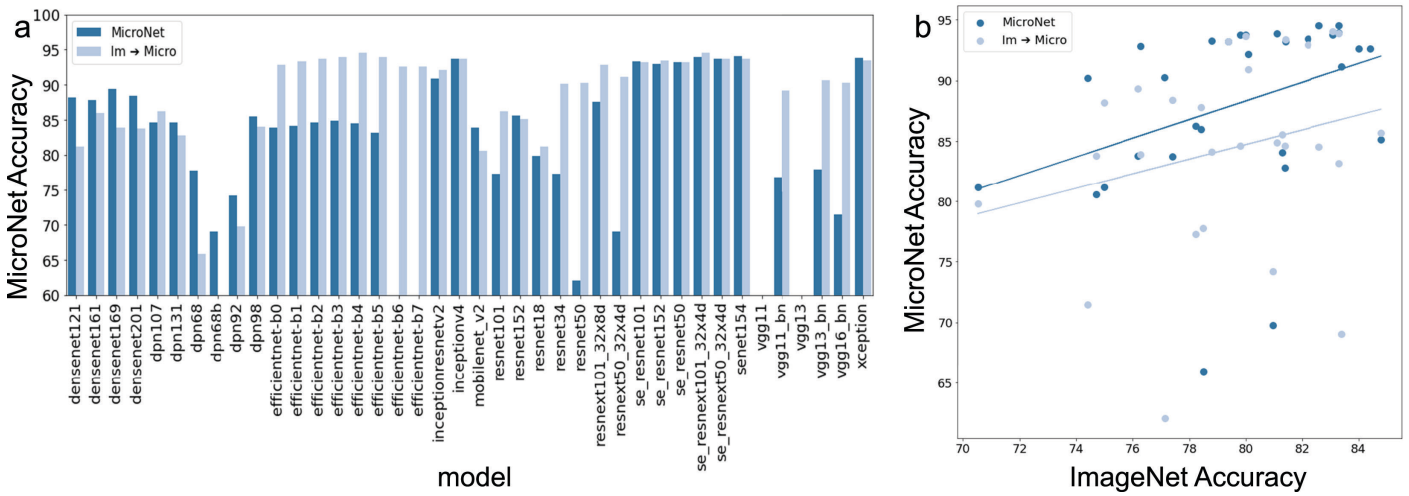

Figure 2.—Accuracy of classification models. (a) Top 1 accuracy of classification models on the MicroNet validation set. The models indicated with dark blue were initialized with Kaiming initialization and trained from scratch while the "Im → Micro" models shown in light blue were pre-trained on ImageNet and then finetuned on the MicroNet data. (b) Comparison of each architecture's accuracy on ImageNet versus MicroNet.

Such a study could help determine whether it would be fruitful to design architectures, scaling rules, and techniques specifically for microscopy analysis instead of largely borrowing best practices from large research efforts on natural images. From our results it seems that there would be value in optimizing the scaling compound coefficient used in EfficientNet for microscopy specific data.

Segmentation Results

The real measure of the value of the trained classification models (i.e., pre-trained encoders) is how well the learned representations transfer to downstream tasks such as segmentation. The pre-trained encoders were applied through transfer learning to seven segmentation tasks which came from two materials: nickel-based superalloys (hereinafter referred to as Super) and environmental barrier coatings (EBC). The number of images in each dataset split is shown in Table 1. Super-1 and EBC-1 contain the full set of labelled data from their respective materials. Super-2 and EBC-2 have only four images in the training set to evaluate the models' performance in few-shot learning. Super-3 and EBC-3 had only one image in the training set to evaluate performance during one-shot learning. Super-4 had test images that were taken under different imaging conditions (Figure 5) to test how well the models' would generalize to unseen out-of-distribution data (e.g., images from different microscopes, microscopists, microscope settings, sample preparation conditions, or different research groups).

Pre-training on MicroNet led to a significant increase in accuracy in during few-shot and one-shot learning on the Super datasets. The segmentation models were trained to segment each pixel into one of three classes: secondary precipitates (large blobs), tertiary precipitates (small blobs), and matrix (background). The training data splits and segmentation accuracy maps for these datasets are shown in Figure 3. The training, validation and test splits had similar looking images and an equal number of dark and light contrast images in each split. The train split for Super-3 had only one image which is outlined in orange in Figure 3(a) and did not contain dark contrast images. The average model performance across all encoder and decoder combinations when initialized with different pre-training weights are shown for each experiment in Table 2. Although the error bars are large because a few models failed to converge during training, it appears that on average when using less training data in Super-2 and Super-3, pre-training with ImageNet followed by MicroNet was slightly better than pre-training with MicroNet or ImageNet alone.

TABLE 1.—NUMBER OF IMAGES IN THE TRAIN, VALIDATION, AND TEST SPLITS OF EACH EXPERIMENTAL DATASET

Experiment	# Training	# Validation	# Test
Super-1	10	4	4
Super-2	4	4	4
Super-3	1	4	4
Super-4	4	4	[a]5
EBC-1	18	3	3
EBC-2	4	3	3
EBC-3	1	3	3

[a]The test images in Super-4 were taken under different imaging conditions than the train and validation sets.

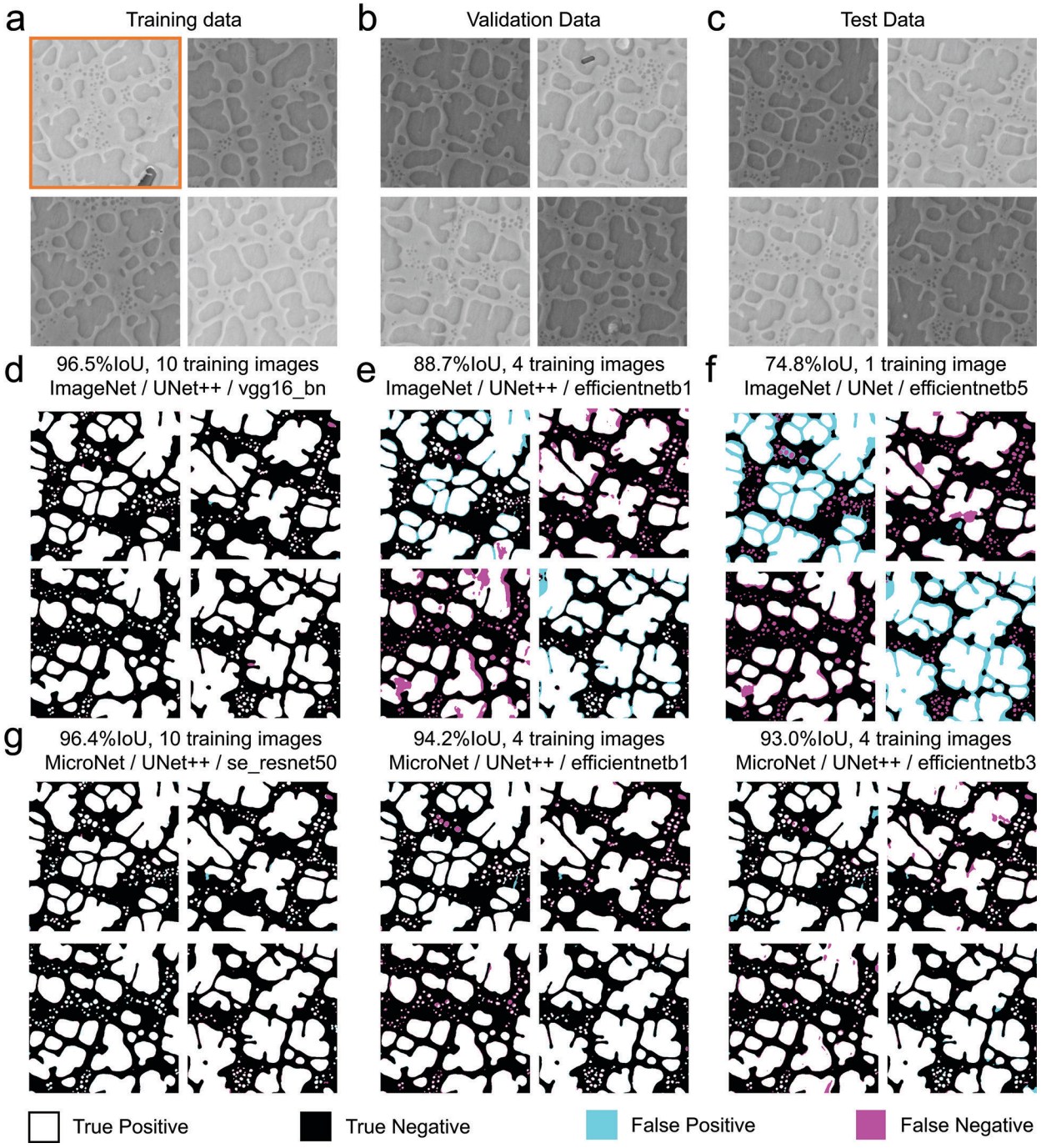

Figure 3.—Segmentation results on Ni-superalloys. (a) shows images from the training data split. Super-1 had ten training images. Super-2 had the four images shown. Super-3 had only one training image which is outlined in orange. (b) and (c) show the validation and test data for the Super datasets respectively. (d)-(i) show the segmentation accuracy masks for the highest accuracy ImageNet and MicroNet models for the first three Super datasets. White pixels indicate true positive predictions, black is true negative, cyan is false positive, and magenta is false negative. (d)-(f) show the models pre-trained on ImageNet. As the number of training images reduce, there is a dramatic reduction in segmentation accuracy. (g)-(i) show the models pre-trained on MicroNet. Even with only one training image, the model accuracy is only slightly reduced when pre-training on MicroNet.

TABLE 2.—AVERAGE IOU PERCENT ACCURACY OF MODELS INITIALIZED WITH DIFFERENT PRE-TRAINING WEIGHT FOR EACH EXPERIMENT

Pre-training	Super-1	Super-2	Super-3	Super-4	EBC-1	EBC-2	EBC-3
None	76.9±22.8	46.2±7.1	48.3±6.2	34.0±7.5	68.0±31.4	48.3±27.7	35.1±10.3
Imagenet	**93.8±7.9**	62.1±12.1	59.7±7.9	47.7±14.1	87.9±19.9	**82.9±17.4**	**43.9±7.0**
MicroNet	93.6±8.7	74.6±14.3	66.9±13.2	52.3±10.5	87.9±18.2	81.6±17.1	40.3±6.2
Im → Micro	85.8±19.2	**74.6±16.2**	**70.0±13.9**	**52.5±16.2**	**88.8±16.6**	81.7±14.7	41.4±8.2

TABLE 3.—AVERAGE IOU PERCENT ACCURACY OF DECODER ARCHITECTURES FOR EACH EXPERIMENT

Decoder	Super-1	Super-2	Super-3	Super-4	EBC-1	EBC-2	EBC-3
DeepLabV3+ (Ref. 44)	----------	----------	----------	----------	86.9±15.3	76.0±22.4	----------
FPN (Ref. 41)	----------	----------	----------	----------	75.6±34.9	----------	----------
LinkNet (Ref. 40)	84.9±20.1	59.3±15.5	53.1±11.9	42.1±12.8	81.6±25.8	65.8±26.2	----------
PAN (Ref. 43)	----------	----------	----------	----------	85.9±17.9	----------	----------
PSPNet (Ref. 42)	----------	----------	----------	----------	72.4±28.5	----------	----------
Unet (Ref. 38)	88.0±15.1	**67.0±17.7**	**62.3±12.3**	48.4±14.9	89.9±13.1	**76.7±22.1**	**40.3±8.4**
Unet++ (Ref. 39)	**89.9±16.3**	66.6±17.8	62.1±11.9	**49.3±14.9**	**90.3±15.2**	76.5±25.5	40.0±8.9

Pre-training with MicroNet showed better performance than ImageNet. With no pre-training (randomly initialized encoder weights), model performance was significantly reduced. Table 3 shows that the UNet and UNet++ decoders were consistently more accurate than LinkNet decoders for Super-1-3. From Table 4, none of the encoder architectures demonstrated clearly superior performance on Super-1-3, although some performed poorly on average. The performance of the best models pre-trained on MicroNet and ImageNet for each of the Super datasets are displayed above the segmentation accuracy maps in Figure 3. The best models pre-trained on ImageNet dropped from 96.5 to 74.8 percent IoU when reducing the number of training images from ten to one. Models pre-trained on MicroNet had only a slight drop in IoU (96.4 to 93.0 percent) when reducing from ten to one training images. This represents a 72.2 percent reduction in relative error in the one-shot case when pre-training on MicroNet versus ImageNet. When training with 10 training images, the segmentation predictions of the top models pre-trained with ImageNet and MicroNet were both highly accurate as shown in Figure 3(d) and (g), respectively. However, when training with only four images, the ImageNet model significantly under-segmented the secondary precipitates in the lighter contrast images and over-segmented them in the darker contrast images (Figure 3(e)). Tertiary precipitates were hardly identified in the lighter contrast image. These problems were exacerbated when trained on only one training image with ImageNet pre-training (Figure 3(f)). For example, the tertiary precipitates were not identified at all and the matrix material between secondary precipitates was misclassified causing many precipitates to be grouped into a single larger one. The MicroNet models were significantly more accurate when training with only four (Figure 3(h)) or one (Figure 3(i)) training images with only some under-segmentation of the tertiary precipitates in the lighter contrast images. Even with only one training image, extracting morphology and size statistics from the predicted segmentation masks would likely be highly accurate. Trying to extract those statistics from the ImageNet one-shot model would be highly misleading and significantly overestimate the size of the secondary precipitates while ignoring the tertiary precipitates. It is interesting to note that in the one-shot case, the MicroNet models produced only slight systematic differences in predictions due to image contrast compared to models pre-trained on ImageNet despite the lack of darker contrast images in the training data. This suggests that pre-training on MicroNet leads to models that are more robust to changes in imaging or sample conditions.

Segmentation accuracy on micrographs with different sample and imaging conditions was greatly improved when pre-training on MicroNet. Figure 4 shows the segmentation accuracy of the Super-4 experiment where the test data was from a different distribution than the training and validation data (shown in Figure 3(a) and (b), respectively). For this experiment the top MicroNet model had an IoU of 78.5 compared to 72.5 percent for the top ImageNet model. Although the accuracy on this extremely out-of-distribution test set is less than ideal, it shows that the MicroNet model has higher usability on a much wider range of sample and imaging conditions without having to label addition training data. For example in Figure 4(b) to (d), the ImageNet model over-segmented and incorrectly aggregated the secondary precipitates (large blobs) while the MicroNet model accurately identified thin gaps between them. In Figure 4(c) tertiary precipitates (small blobs) identified in the MicroNet model were not identified in the ImageNet model. The higher accuracy on out-of-distribution data indicates that pre-trained MicroNet encoders are more general and useful for comparing results between research groups, microscopes, sample preparation conditions, and imaging conditions. When pre-training on MicroNet the top model showed significant improvement for the one-shot learning case on the environmental barrier coating datasets (EBC-3) compared to the top ImageNet model. From

Table 3, the best decoder architecture on average appeared to be UNet or UNet++, although DeepLabV3+ was not evaluated for all datasets and appears promising. The top models for each EBC dataset used the UNet++ decoder except one which used UNet (Figure 5). There was not a clearly best encoder architecture for the EBC datasets as shown in Table 4, although some architectures were clearly inferior. Table 2 shows that on average across all encoders and decoders, ImageNet models performed slightly better for EBC-2 and EBC-3 while pre-training on ImageNet and then MicroNet gave the best average performance on EBC-1. However, it is difficult to determine which pre-training method was clearly superior from the average results because of the wide error bars and the occasional poor performance of a few models that randomly failed to converge. A lack of any pre-training lead to significantly degraded performance. A clearer picture of the best pre-training method is given by the segmentation results of the best ImageNet and MicroNet model for each EBC dataset as shown in Figure 5. On EBC-1 and EBC-2 when training with 18 and 4 images, respectively, there was not a significant difference between pre-training on MicroNet and ImageNet, although ImageNet pre-training was slightly better for the top models. For EBC-3 the top MicroNet model saw a 14.3 percent reduction in relative error compared to the top ImageNet model (65.9 percent IoU vs. 60.2 percent IoU). This resulted in a large over segmentation of the oxide layer with the ImageNet model making it impossible to measure oxide thickness (Figure 5(b) third row). Meanwhile the one-shot MicroNet model (Figure 5(c) third row) is highly useable for oxide thickness measurements after simple morphological operations such as binary opening which is useful for removing small objects.

TABLE 4.—AVERAGE IOU PERCENT ACCURACY OF ENCODER ARCHITECTURES FOR EACH EXPERIMENT

Encoder	Super-1	Super-2	Super-3	Super-4	EBC-1	EBC-2	EBC-3
DenseNet121 (Ref. 26)	----------	----------	----------	----------	89.1±6.2	80.9±10.8	38.3±5.3
DenseNet161 (Ref. 26)	----------	----------	----------	----------	90.9±4.8	86.1±5.5	39.6±2.9
DenseNet169 (Ref. 26)	----------	----------	----------	----------	90.8±4.7	81.5±9.2	----------
DenseNet201 (Ref. 26)	----------	----------	----------	----------	90.3±5.1	**86.5±4.8**	----------
dpn107 (Ref. 27)	----------	----------	----------	----------	90.7±5.6	80.7±22.1	----------
dpn131 (Ref. 27)	----------	----------	----------	----------	86.7±18.0	79.8±16.8	----------
dpn68 (Ref. 27)	----------	----------	----------	----------	74.6±29.2	61.9±27.1	40.1±5.2
dpn68b (Ref. 27)	----------	----------	----------	----------	69.6±29.6	56.9±29.4	36.5±8.5
dpn92 (Ref. 27)	----------	----------	----------	----------	84.1±17.8	78.3±12.3	----------
dpn98 (Ref. 27)	----------	----------	----------	----------	87.1±17.9	72.9±26.9	----------
EfficientNet-b0 (Ref. 28)	77.7±23.9	56.9±14.3	51.1±12.6	53.7±15.8	59.0±41.5	55.9±32.7	34.7±16.9
EfficientNet-b1 (Ref. 28)	66.9±26.4	62.9±19.5	52.8±19.7	58.7±13.9	60.0±40.7	57.1±33.7	40.7±7.1
EfficientNet-b2 (Ref. 28)	65.1±26.0	66.9±21.6	51.1±18.4	**59.4±14.6**	68.6±36.0	57.6±33.6	----------
EfficientNet-b3 (Ref. 28)	69.4±26.9	68.1±22.0	62.4±19.1	59.2±11.4	71.6±33.9	60.0±33.0	----------
EfficientNet-b4 (Ref. 28)	74.8±26.0	67.9±18.7	58.9±15.2	57.2±13.9	70.3±35.3	61.4±33.0	----------
EfficientNet-b5 (Ref. 28)	76.3±23.0	70.5±22.5	61.3±17.2	57.7±17.3	73.4±35.6	64.6±31.3	36.4±17.0
Inception-ResNet-V2 (Ref. 30)	94.5±3.1	64.4±12.2	63.0±11.0	43.1±11.1	87.9±10.4	75.6±25.7	38.2±5.2
Inception-V4 (Ref. 30)	91.6±8.8	76.2±19.9	**69.1±13.3**	48.6±12.1	84.0±26.4	73.2±26.9	39.3±6.3
MobileNet-V2 (Ref. 32)	----------	----------	----------	----------	71.3±34.5	60.5±34.2	45.7±6.5
ResNet-101 (Ref. 29)	----------	----------	----------	----------	86.4±18.1	76.1±15.7	38.5±8.6
ResNet-152 (Ref. 29)	----------	----------	----------	----------	86.2±18.7	75.5±23.7	----------
ResNet-18 (Ref. 29)	----------	----------	----------	----------	89.8±5.9	82.1±7.3	----------
ResNet-34 (Ref. 29)	94.3±2.1	57.9±11.7	59.5±7.2	35.8±11.3	91.1±5.3	80.2±9.7	----------
ResNet-50 (Ref. 29)	91.2±7.2	48.9±11.9	55.3±5.9	32.2±9.4	82.5±24.2	81.2±9.6	38.2±3.4
ResNeXt-101_32x8d (Ref. 33)	95.3±1.5	57.5±12.1	57.3±4.4	39.8±8.8	----------	----------	----------
Resnext-50_32x4d (Ref. 33)	90.9±8.9	54.7±11.6	54.6±7.8	39.0±9.6	81.9±23.9	69.9±27.0	33.4±5.1
SE_ResNet-101 (Ref. 34)	93.3±8.4	66.4±13.1	57.8±9.4	47.7±11.2	93.3±4.0	84.6±7.8	42.4±6.6
SE_ResNet-152 (Ref. 34)	95.2±1.9	60.6±12.5	57.4±6.0	38.6±9.3	93.0±4.4	82.7±19.2	40.2±4.1
SE_ResNet-50 (Ref. 34)	94.8±4.3	63.1±15.3	57.3±5.4	39.8±9.2	88.9±18.2	81.6±14.6	**47.3±8.2**
SE_ResNeXt-101_32x4d (Ref. 33)	95.4±1.3	67.3±19.3	60.8±13.5	44.2±8.6	91.8±6.0	79.5±18.7	45.6±5.3
SE_ResNeXt-50_32x4d (Ref. 33)	**96.0±0.3**	67.3±17.4	59.6±5.1	44.4±9.4	**92.6±4.0**	83.3±14.0	41.6±2.5
SENet-154 (Ref. 34)	94.5±4.2	**76.6±17.5**	63.2±12.7	51.0±15.0	91.9±6.7	75.9±30.0	42.2±4.5
VGG-13_bn (Ref. 24)	94.2±6.6	65.2±17.8	64.3±14.8	46.6±15.0	----------	----------	----------
VGG-16_bn (Ref. 24)	95.2±2.1	66.6±19.0	66.2±14.7	40.2±17.0	87.5±18.7	84.0±5.1	33.8±8.4
Xception (Ref. 31)	93.8±4.5	61.3±10.8	54.1±7.8	39.3±9.9	92.0±5.1	74.6±33.9	42.5±9.7

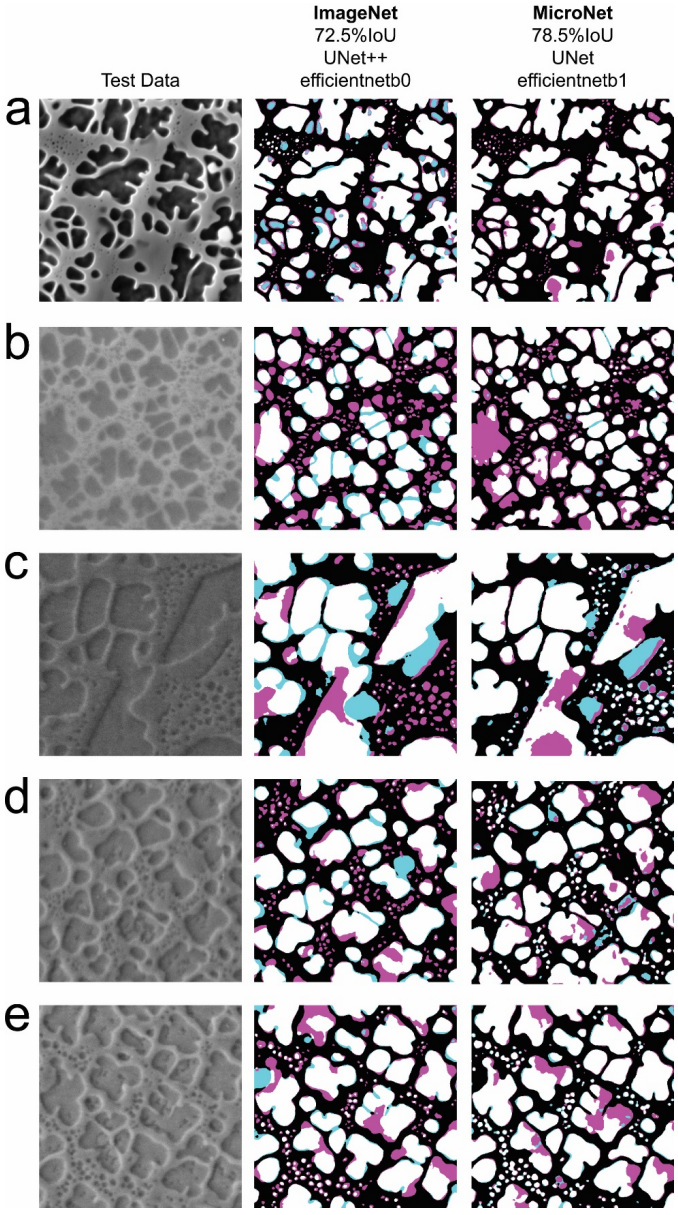

Figure 4.—Accuracy of Super-4 segmentation models evaluated on test data with unseen imaging conditions. (a) shows from left to right: the test image, the results for the top ImageNet model, and the results for the top MicroNet model. (b)–(e) show the same for each of the remaining test set images. The colors represent the same as in Figure 3.

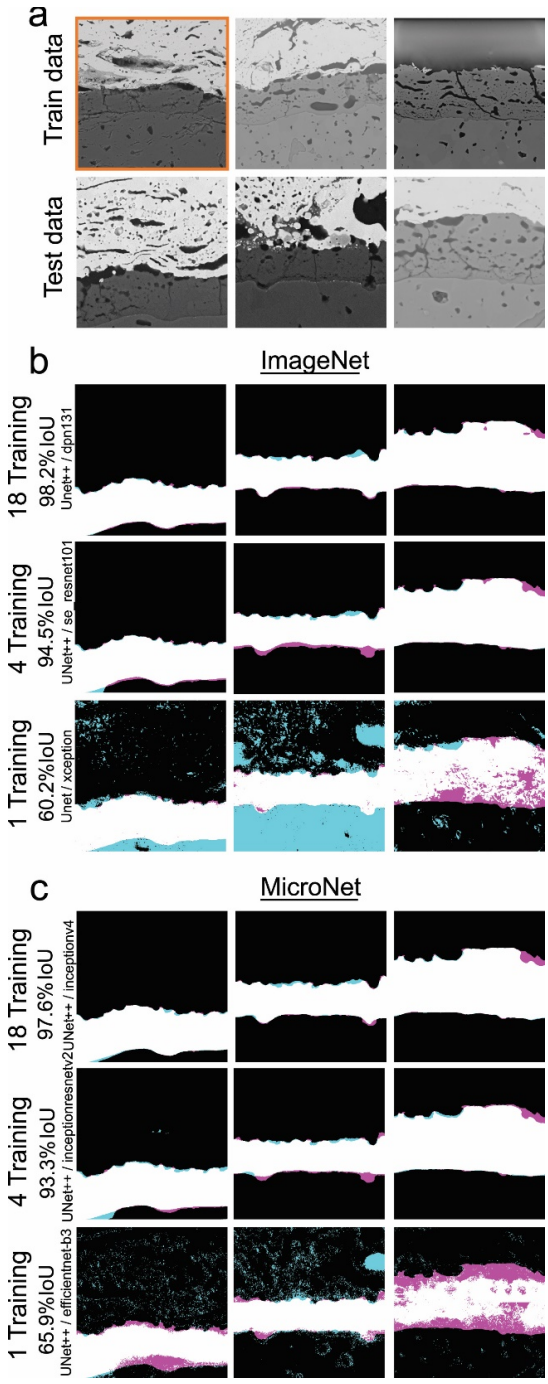

Figure 5.—Results of EBC segmentation. (a) shows examples from the train and test splits of the EBC datasets. The single training image for EBC-3 outlined in yellow. (b) shows the segmentation results for the top ImageNet model for each EBC experiment. (c) shows the segmentation results for the top MicroNet model for each EBC experiment. The colors represent the same as in Figure 3.

Conclusion

Transfer learning from CNN encoders pre-trained on MicroNet produced accurate segmentation models with significantly less training data than pre-training on ImageNet and the models generalized to better to unseen data with different imaging or sample conditions. This is significant because labeling training data for segmentation tasks is expensive and the labeled data cannot account for all possible imaging and sample conditions that the model should be expected to perform accurately on. By generalizing better to out-of-distribution microscopy images, this technique produces segmentation results that are more accurate and comparable between microscopes, microscopists, and research groups, thus increasing the utility and shareability of the trained models. These results suggest that the MicroNet pre-trained encoders generate superior microstructure feature representations and will likely improve the accuracy of other deep learning microscopy analysis tasks that commonly utilize pre-trained ImageNet encoders. The pre-trained MicroNet encoders have been made readily available and can be used in existing projects with only a couple lines of code. Accurate microscopy segmentation and quantitative microstructure feature extraction is critical to quantitatively linking the processing-structure-property relationships of materials. A quantitative understanding of these relationships is required to accelerate materials discovery and design through traditional or data-driven methods such as Bayesian active learning.

Data Availability

Data that supports the findings of this study, including labelled segmentation data and pre-trained encoders trained on MicroNet, are available at https://github.com/nasa/pretrained-microscopy-models.

Code Availability

All code necessary to apply this technique and supports the findings in this study is available at https://github.com/nasa/pretrained-microscopy-models.

References

1. ASTM E112, "Standard Test Methods for Determining Average Grain Size E112-10," Astm E112-10, vol. 96, no. 2004, pp. 1–27, 2010, doi: 10.1520/E0112-10.
2. A. E45-18a, "Standard Test Methods for Determining the Inclusion Content of Steel." ASTM International West Conshohocken, PA, USA, 2018.
3. J. Stuckner, K. Frei, I. McCue, M.J. Demkowicz, and M. Murayama, "AQUAMI: An open source Python package and GUI for the automatic quantitative analysis of morphologically complex multiphase materials," Comput. Mater. Sci., vol. 139, 2017, doi: 10.1016/j.commatsci.2017.08.012.
4. T.M. Smith et al., "Characterization of nanoscale precipitates in superalloy 718 using high resolution SEM imaging," Mater. Charact., vol. 148, no. December 2018, pp. 178–187, 2019, doi: 10.1016/j.matchar.2018.12.018.
5. J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in 2009 IEEE conference on computer vision and pattern recognition, 2009, pp. 248–255.
6. B.L. DeCost, B. Lei, 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., vol. 25, no. 1, pp. 21–29, 2019, doi: 10.1017/S1431927618015635.
7. G. Roberts, R. Sainju, B. Hutchinson, M.B. Toloczko, D.J. Edwards, and Y. Zhu, "DefectNet – A Deep Convolutional Neural Network for Semantic Segmentation of Crystallographic Defects in Advanced Microscopy Images," Microsc. Microanal., vol. 25, no. S2, pp. 164–165, 2019, doi: 10.1017/s1431927619001557.
8. N.M. Senanayake and J. L.W. Carter, "Computer Vision Approaches for Segmentation of Nanoscale Precipitates in Nickel-Based Superalloy IN718," Integr. Mater. Manuf. Innov., vol. 9, no. 4, pp. 446–458, 2020, doi: 10.1007/s40192-020-00195-z.
9. R. Cohn, I. Anderson, T. Prost, J. Tiarks, E. White, and E. Holm, "Instance Segmentation for Direct Measurements of Satellites in Metal Powders and Automated Microstructural Characterization from Image Data," JOM, pp. 1–14, 2021.
10. T. Stan, Z. Thompson, and P. Voorhees, "Building towards a universal neural network to segment large materials science imaging datasets," no. September, p. 49, 2019, doi: 10.1117/12.2525290.
11. C. Groschner, C. Choi, D. Nguyen, C. Ophus, and M. Scott, "Machine Learning for High Throughput HRTEM Analysis," Microsc. Microanal., vol. 25, no. S2, pp. 150–151, 2019, doi: 10.1017/s143192761900148x.
12. P. Potocek, P. Trampert, P. Trampert, M. Peemen, R. Schoenmakers, and T. Dahmen, "Sparse Scanning Electron Microscopy Data Acquisition and Deep Neural Networks for Automated Segmentation in Connectomics," Microsc. Microanal., vol. 26, no. 3, pp. 403–412, 2020, doi: 10.1017/S1431927620001361.

13. M. Cordts et al., “The cityscapes dataset for semantic urban scene understanding,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 3213–3223.
14. Q. Luo, E.A. Holm, and C. Wang, “A transfer learning approach for improved classification of carbon nanomaterials from TEM images,” Nanoscale Adv., vol. 3, no. 1, pp. 206–213, 2021.
15. B.L. DeCost, M.D. Hecht, T. Francis, B.A. Webler, Y.N. Picard, and E.A. Holm, “UHCSDB: UltraHigh Carbon Steel Micrograph DataBase: Tools for Exploring Large Heterogeneous Microstructure Datasets,” Integr. Mater. Manuf. Innov., vol. 6, no. 2, pp. 197–205, 2017, doi: 10.1007/s40192-017-0097-0.
16. R. Aversa, M. H. Modarres, S. Cozzini, R. Ciancio, and A. Chiusole, “Data descriptor: The first annotated set of scanning electron microscopy images for nanoscience,” Sci. Data, vol. 5, 2018, doi: 10.1038/sdata.2018.172.
17. B.L. DeCost and E.A. Holm, “A large dataset of synthetic SEM images of powder materials and their ground truth 3D structures,” Data Br., vol. 9, pp. 727–731, 2016.
18. A.M. Karimi et al., “Automated Pipeline for Photovoltaic Module Electroluminescence Image Processing and Degradation Feature Classification,” IEEE J. Photovoltaics, vol. 9, no. 5, pp. 1324–1335, 2019, doi: 10.1109/JPHOTOV.2019.2920732.
19. K. He, X. Zhang, S. Ren, and J. Sun, “Delving deep into rectifiers: Surpassing human-level performance on imagenet classification,” in Proceedings of the IEEE international conference on computer vision, 2015, pp. 1026–1034.
20. V. Nair and G.E. Hinton, “Rectified linear units improve restricted boltzmann machines,” in Proceedings of the 27th international conference on machine learning (ICML-10), 2010, pp. 807–814.
21. C. Szegedy et al., “Going deeper with convolutions,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 1–9.
22. A. Paszke et al., “Pytorch: An imperative style, high-performance deep learning library,” arXiv Prepr. arXiv1912.01703, 2019.
23. M. Tan et al., “Mnasnet: Platform-aware neural architecture search for mobile,” in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 2820–2828.
24. K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” arXiv Prepr. arXiv1409.1556, 2014.
25. S. Ioffe and C. Szegedy, “Batch normalization: Accelerating deep network training by reducing internal covariate shift,” in International conference on machine learning, 2015, pp. 448–456.
26. G. Huang, Z. Liu, L. Van Der Maaten, and K.Q. Weinberger, “Densely connected convolutional networks,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 4700–4708.
27. Y. Chen, J. Li, H. Xiao, X. Jin, S. Yan, and J. Feng, “Dual path networks,” arXiv Prepr. arXiv1707.01629, 2017.
28. M. Tan and Q. V Le, “Efficientnet: Rethinking model scaling for convolutional neural networks,” arXiv Prepr. arXiv1905.11946, 2019.
29. K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.
30. C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, “Inception-v4, inception-resnet and the impact of residual connections on learning,” in Proceedings of the AAAI Conference on Artificial Intelligence, 2017, vol. 31, no. 1.
31. F. Chollet, “Xception: Deep learning with depthwise separable convolutions,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 1251–1258.
32. M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, “Mobilenetv2: Inverted residuals and linear bottlenecks,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 4510–4520.
33. S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, “Aggregated residual transformations for deep neural networks,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 1492–1500.
34. J. Hu, L. Shen, and G. Sun, “Squeeze-and-excitation networks,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 7132–7141.
35. P. Yakubovskiy, “Segmentation Models Pytorch,” GitHub Repos., 2020, [Online]. Available: https://github.com/qubvel/segmentation_models.pytorch.
36. D.P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” arXiv Prepr. arXiv1412.6980, 2014.
37. C.H. Sudre, W. Li, T. Vercauteren, S. Ourselin, and M.J. Cardoso, “Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations,” in Deep learning in medical image analysis and multimodal learning for clinical decision support, Springer, 2017, pp. 240–248.

39. O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in International Conference on Medical image computing and computer-assisted intervention, 2015, pp. 234–241.
40. Z. Zhou, M.M.R. Siddiquee, N. Tajbakhsh, and J. Liang, “Unet++: Redesigning skip connections to exploit multiscale features in image segmentation,” IEEE Trans. Med. Imaging, vol. 39, no. 6, pp. 1856–1867, 2019.
41. A. Chaurasia and E. Culurciello, “Linknet: Exploiting encoder representations for efficient semantic segmentation,” in 2017 IEEE Visual Communications and Image Processing (VCIP), 2017, pp. 1–4.
42. A. Kirillov, K. He, R. Girshick, and P. Dollár, “A unified architecture for instance and semantic segmentation.” 2017.
43. H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, “Pyramid scene parsing network,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 2881–2890.
44. H. Li, P. Xiong, J. An, and L. Wang, “Pyramid attention network for semantic segmentation,” arXiv Prepr. arXiv1805.10180, 2018.
45. L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, “Rethinking atrous convolution for semantic image segmentation,” arXiv Prepr. arXiv1706.05587, 2017.
46. L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, “Encoder-decoder with atrous separable convolution for semantic image segmentation,” in Proceedings of the European conference on computer vision (ECCV), 2018, pp. 801–818.

39. O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in International Conference on Medical image computing and computer-assisted intervention, 2015, pp. 234–241.
40. Z. Zhou, M.M.R. Siddiquee, N. Tajbakhsh, and J. Liang, “Unet++: Redesigning skip connections to exploit multiscale features in image segmentation,” IEEE Trans. Med. Imaging, vol. 39, no. 6, pp. 1856–1867, 2019.
41. A. Chaurasia and E. Culurciello, “Linknet: Exploiting encoder representations for efficient semantic segmentation,” in 2017 IEEE Visual Communications and Image Processing (VCIP), 2017, pp. 1–4.
42. A. Kirillov, K. He, R. Girshick, and P. Dollár, “A unified architecture for instance and semantic segmentation.” 2017.
43. H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, “Pyramid scene parsing network,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 2881–2890.
44. H. Li, P. Xiong, J. An, and L. Wang, “Pyramid attention network for semantic segmentation,” arXiv Prepr. arXiv1805.10180, 2018.
45. L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, “Rethinking atrous convolution for semantic image segmentation,” arXiv Prepr. arXiv1706.05587, 2017.
46. L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, “Encoder-decoder with atrous separable convolution for semantic image segmentation,” in Proceedings of the European conference on computer vision (ECCV), 2018, pp. 801–818.