

Optimizing convolutional neural networks to perform semantic segmentation on large materials imaging datasets: X-ray tomography and serial sectioning



Tiberiu Stan*, Zachary T. Thompson, Peter W. Voorhees

Department of Materials Science and Engineering, Northwestern University, 2220 Campus Drive, Cook Hall, Evanston, IL 60208, USA

ARTICLE INFO

Keywords:
 Machine learning
 Convolutional neural network
 Semantic segmentation
 X-ray tomography
 Serial sectioning
 Solidification

ABSTRACT

Machine learning was used to segment large materials science datasets resulting from synchrotron-based x-ray computed tomography (XCT) images of dendrite growth, and serial sectioning (SS) images of dendrite coarsening. Both neural networks (NNs) yielded quantitatively more accurate outputs than conventional segmentation techniques using only 30 XCT or 6 SS training images. We show that performance can be improved if NNs are trained using a large number of small images that are sampled from the fixed amount of training data. The optimal image size and number of training images was identified for the XCT and SS datasets. NN transferability was also tested by applying the highest performing XCT and SS NNs to related datasets. While the initial segmentations were successful, applying simple transformations to the raw images further improved NN performance. These results show the great predictive ability and promising future of using machine learning for segmentation of large materials science datasets.

1. Introduction

The mechanical properties of structural alloys are largely controlled by the microstructures left behind after solidification [1–3]. Thus, properly characterizing these features is important to producing high-quality materials in a controllable fashion [4,5]. Common microstructural imaging techniques include: optical microscopy, atomic force microscopy, ultrasound, magnetic resonance imaging, scanning electron microscopy (SEM), x-ray computed tomography (XCT), field electron microscopy, and transmission electron microscopy. The images produced through these methods must be interpreted to extract useful information. A major challenge for the materials science and engineering community is that modern instruments have large detectors and are capable of recording multimodal or time-dependent data. This leads to large terabyte (TB) to petabyte datasets that need to be analyzed. As illustration of the challenge, we employ images created by synchrotron-based four-dimensional XCT, and automated serial sectioning (SS). Both techniques result in 3D visualizations of dendrite sizes, shapes, and their spatial distribution. However, image segmentation (the intermediate step of accurately separating dendrites from background features) is often the most difficult and time-consuming part of the analysis [6].

The goal of our 4D XCT in-situ solidification experiments is to

record the formation of dendritic structures while an alloy is being cooled from the liquidus. Samples are placed in the path of the incident synchrotron x-ray beam and projections of the continuously-changing microstructure are captured. These projections are then grouped and reconstructed into a series of volumetric datasets which show the full 3D alloy microstructure at different time steps. The datasets are then split into 2D z-slice images and segmented into the solid and liquid phases. A week-long synchrotron beamtime can produce ~40 dendrite growth experiments, each one containing ~170 time steps, each with ~340 z-slices, for a total of ~2.3 million images and ~12.5 TB of data. Segmenting each image by hand takes a human ~20 min, thus processing all of the beamtime datasets would take ~88 years of continuous work.

SS is a post-mortem characterization technique. In our study the samples are melted, re-solidified, coarsened, and cooled to room temperature prior to being mounted in the automated serial sectioner. A thin slice of the sample surface is mechanically removed, and the underlying microstructure is imaged using an optical camera. This process is repeated until the desired sample volume has been captured. A typical experiment contains 24 samples, each sectioned ~500 times, for a total of ~12,000 images. Segmenting SS data by hand takes a human ~4 h per image, thus requiring a total of ~5.5 years of continuous work.

* Corresponding author.

E-mail address: tiberiu.stan@northwestern.edu (T. Stan).

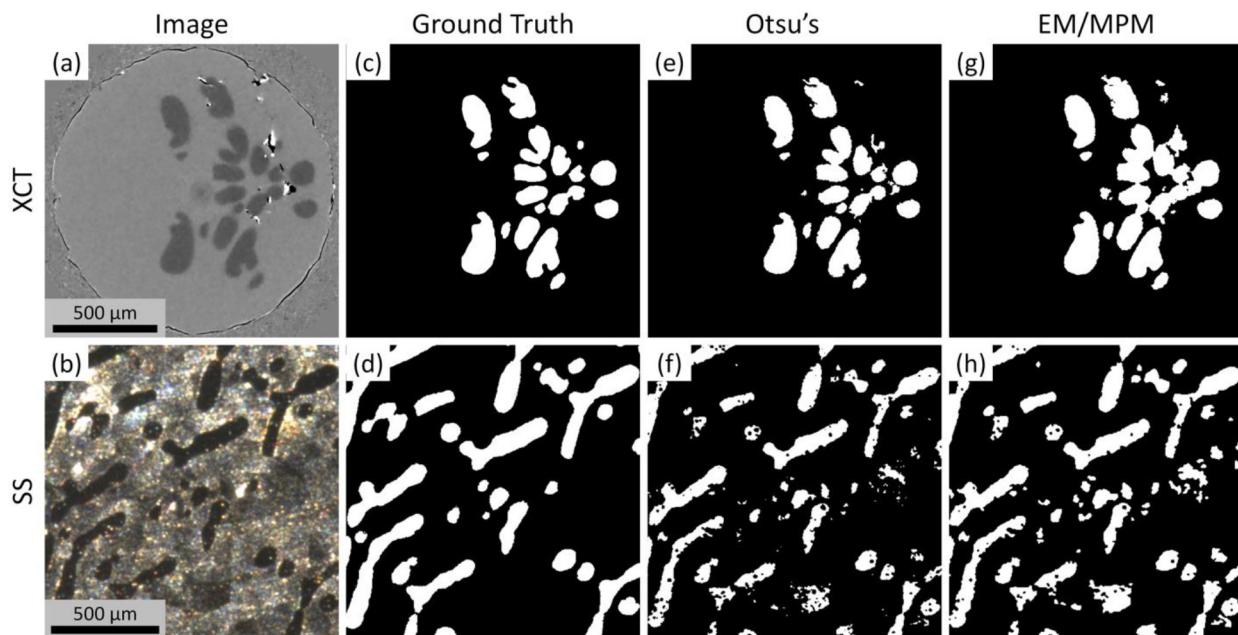


Fig. 1. (a) An XCT image and (b) SS image; (c) the XCT ground truth corresponding with (a), and (d) the SS ground truth corresponding with (b), where the location of dendrites is shown in white; (e) and (f) segmentations produced using Otsu's method; (g) and (h) segmentations produced using EM/MPM.

Both the XCT and SS techniques produce a large number of images which would take a human an impractical amount of time to segment. Therefore, computational approaches must be used. Some traditional segmentation algorithms include thresholding using Otsu's method [7] and Canny edge detection [8]. More advanced programs, such as Expectation-Maximization/Maximization of the Posterior Marginals (EM/MPM) [9], have a variety of optimizable input parameters which aid in segmentation. However, these approaches generally only work well on images where the features of interest have uniform and distinct luminance (pixel intensity values), as well as easily-identifiable edges. Complex images may contain unwanted contrast gradients, random noise, un-focused regions, and a variety of artifacts due to the image collection method. These factors can lead to segmentations with extra or missing features, and poor boundary identification.

Fig. 1a shows a 852×852 pixel XCT image from a solidification experiment on Al-Zn, and **Fig. 1b** shows a cropped portion of an SS micrograph from a coarsening experiment on Pb-Sn (experimental details are discussed in [Section 2.1](#)). The corresponding ground truth segmentations in **Fig. 1c** and d show the location of dendrites in white, and background in black. Segmentation attempts using Otsu's method and EM/MPM are shown in **Fig. 1e-h**. Both techniques are unable to closely match the ground truths. A quantitative analysis of segmentations using Otsu's method and EM/MPM is presented in [Section 3.2](#).

Recently, machine learning methods have been applied for image classification and semantic segmentation. Classification using convolutional neural networks involves passing image data through network layers which apply convolutions, normalizations, non-linear calculations, and max-pooling operations to then decrease the dimensionality of the input [10]. Applying these layers multiple times eventually results in the full image being designated one individual class. For example, convolutional neural networks (NNs) trained to classify steel microstructures take SEM images as input, and output classes such as martensite, bainite, and pearlite [11]. NNs have been used to classify many materials structures including steels [11–13], dendrites [12–14], grains [13,14], and precipitates [13,15,16]. Further, there is large interest in using NNs for medical imaging and detection of anomalies such as fractures, tumors, or diseases [17].

Semantic segmentation is the process of assigning a class label to each pixel of an image, usually using NNs [10,18]. Unlike NNs trained

for classification, architectures designed for semantic segmentation include a decoder network [19]. These upsampling layers project the lower resolution features learnt by the encoder layers and restore the original image dimensionality to allow for each pixel to be labeled. This technique has been widely used to develop software for identifying humans [20], objects [10,19–21], and scenery [10,19–22]. However, there is currently limited literature on semantic segmentation for materials science applications. Kaira et al. used a NN to segment Al₂Cu plate-like precipitates in an Al – 4 wt%Cu alloy [23]. Further, Bulgarevich et al. used a Random Forest statistical algorithm to segment steel microstructures [24]. While machine learning has aided the analysis of several materials science imaging experiments in recent years, the potential for its widespread use in the field has not yet been realized.

NN training requires a set of raw input images along with corresponding ground truth label images. Creating training ground truths is laborious and perhaps the largest barrier to entry for getting started with machine learning. Generally, having more training data leads to improved network accuracy [25,26]. However, the actual number of training images needed to obtain satisfactory NN segmentations depends on many factors such as: network architecture, class distribution throughout the images, number of classes, overall image size, and image complexity [25]. In this study, we quantify image complexity through measurements of luminance (a pixel's intensity value) and contrast (root mean square of all luminance in an image). Higher luminance values indicate brighter pixels, while higher contrast values indicate more variability in pixel intensity.

One of the major advantages of NNs is that training can be done using small images (hundreds of pixels), but the NN can be applied to arbitrarily large images [27]. If the training images are too small, the NN may not detect the class-defining features such as luminance, size, shape, distribution, relation to other classes [28]. If the training images are large but too few in number, there may be insufficient data to fit the large number of internal network parameters, especially since the number of parameters increases with training image size [29]. Thus, for a given amount of training data, there exists an optimal combination of image size and number of images that leads to best NN performance.

The present study is aimed at answering two main questions. The first is: given a set amount of 3D or 4D training data, how can it be best used to maximize the network performance? We show that properly

trained NNs can segment large datasets with higher accuracy than most classical algorithms that perform mathematical operations on the grey levels of the pixels in the image. Thus, NNs are ideal for the large datasets produced by modern instruments. The second question is: how well do NNs trained on one dataset perform on other types of images? The highest performing XCT and SS NNs were applied to different datasets including: XCT images containing ring artifacts, XCT images from a different alloy, and SS images from a different coarsening condition. The resulting outputs and best-practice techniques to improve NN performance are discussed.

2. Materials and methods

2.1. Image acquisition

The XCT dataset was obtained from a solidification experiment. Measurements were carried out at the Argonne National Laboratory, Advanced Photon Source, beamline 2-BM. A 6 mm tall, 1 mm diameter rod was machined out of an Al – 20 wt%Zn slab, and placed into a cylindrical boron nitride holder. The assembly was then elevated into a custom furnace built within the x-ray beam-path, and the temperature was raised to ~ 625 °C to ensure full sample melting. The assembly was then rotated at a slew speed of 720 degrees/s, cooled from ~ 620 °C to ~ 618 °C at a rate of 1 °C/min, and projections were captured every ~ 2 ms. Differences in phase composition provide contrast between the dendritic solid and the liquid background. A total of 115,200 projections were obtained throughout the ~ 3.3 min solidification experiment. The projections were later processed into two hundred 3D reconstructions using the time-interlaced model-based iterative reconstruction (TIMBIR) algorithm [30,31]. The resulting datasets have $1.65 \times 1.65 \times 1.65 \mu\text{m}^3$ voxel size and 1 s temporal resolution. Each reconstruction is composed of 340 grayscale z-slice images of 852×852 pixels, a total of 68,000 images that need to be segmented.

The SS dataset was obtained from a coarsening experiment performed in microgravity to decrease the effect of convection and to isolate the diffusion-controlled processes. The Pb-Sn system was chosen because of the well-understood phase diagram, low eutectic temperature of 183 °C [32], and large optical contrast between the Sn-rich and Pb-rich phases. On Earth, Pb and Sn were mixed and cast into an ingot containing 68.1 wt%Sn (20 vol%Sn). Rods obtained from the ingot were directionally solidified in a Bridgman-type furnace [33] to create a Sn-rich dendritic phase. 0.5 cm long sections were then machined out from the steady state region of the rods in preparation for the microgravity experiments. Upon the International Space Station, samples were isothermally coarsened at 185 °C for 1.6 h. Once returned to Earth, the samples were stored at -80 °C to slow further microstructural evolution prior to sectioning. Following a previously reported serial sectioning general procedure [34], $\sim 4.07 \mu\text{m}$ thin slices of the surface were mechanically removed and the underlying microstructure was optically imaged using a Leica DFC295 camera. No etchant was used between surface removal and imaging. The resulting 1689×985 pixel 3-channel RGB images have voxel dimensions of $3.24 \times 3.24 \times 4.07 \mu\text{m}^3$. A total of 519 images were captured, corresponding to a sample volume of $\sim 37 \text{ mm}^3$.

2.2. Ground truths

The full XCT and SS experiments contain 68,000 and 519 images, respectively. One of the goals of the present study is to report the best way to train a NN given a fixed amount of training data. Images were selected to cover a wide range of microstructural properties such as dendrite size, shape, luminance, and area fraction. To limit the number of ground truth segmentations needed, a total of 42 XCT images and 9 SS images were chosen as representative portions of the microstructures. It is common that 50 to 90% of the available images are used for training [11,19,21], thus 30 XCT and 6 SS images were selected.

These images and their corresponding ground truths were used during NN training to adjust internal fitting parameters. 10 XCT and 2 SS images make up the validation sets which are used during training to periodically test NN performance, and act as a convergence criterion to stop NN training. Finally, 2 XCT images and 1 SS image were used as a standard for comparing the performance of NNs trained using different parameters.

The first step in creating XCT ground truths is to remove the ring artifacts commonly found in tomography studies [35]. The Al-Zn solidification experiment spans a total of 128 time steps, each with 340 z-slices. It was observed that the rings in the first reconstruction (containing no dendritic structure) are nearly identical to the rings in the later time steps. Thus, to better isolate the dendrites, the first reconstruction was subtracted from all subsequent time steps.

Next, the 42 single-channel grayscale XCT images selected for the NN study were converted to three-channel by repeating the pixel values three times. Each image was then smoothed using a moving Gaussian filter, and Canny edge detection was used to outline the dendrite bodies. An edge-link-and-fill procedure was then used to mark the dendrites with white, and the background with black. The resulting segmentations were usually not acceptable because the edge detection does not differentiate between dendrites, contaminant particles throughout the liquid, and noise within the liquid itself. Thus, the segmentations were manually cleaned using Adobe Photoshop. As described in later sections, the dendrites in the XCT dataset are difficult to segment due to the diffuse edges caused by the reconstruction algorithm. The combination of Canny edge detection and manual cleaning gives an estimated ground truth boundary accuracy of ± 2 pixels. A total of ~ 14 h were required to create the 42 XCT ground truths (~ 20 min/image).

The 9 SS images were initially roughly segmented using EM/MPM [9] and parameters were chosen to produce slightly over-segmented results (a higher dendrite area fraction than in a perfect segmentation) as they are empirically easier to manually clean than under-segmented images. The EM/MPM segmentations along with the original RGB images were then uploaded into the GIMP image processing software. To aid the manual segmentation process, the sections above and below the image being segmented were inspected to determine if some dark features were dendrites or artifacts from the imaging and/or the sectioning process. Similar to the XCT dataset, manual segmentation of SS images results in an estimated ground truth boundary accuracy of ± 2 pixels. The entire SS ground truth creation process took ~ 36 h (~ 4 h/image).

Creating a SS ground truth image takes nearly twelve times as long as an XCT ground truth image. This is primarily due to differences in image size (852×852 pixels for XCT and 1689×985 pixels for SS), and dendrite boundary perimeter that must be accurately segmented (~ 2300 pixels/image for XCT and $\sim 40,200$ pixels/image for SS). A more detailed comparison of the XCT and SS datasets is presented in Section 5.1.

2.3. Image sizes and numbers

Using more images for NN training generally leads to improved network performance [25,26]. However, creating a large number of ground truths is laborious and may not be necessary to achieving satisfactory results. Through a systematic study, we show that NN performance can be improved if a large number of small images are sampled from the training data. Square images with side lengths of 400, 224, 100, and 50 pixels were selected from the training and validation images. VGG-16 NNs used for classification are most often trained using 224×224 pixel images [21,36]. Thus, 224×224 pixel images were used instead of 200×200 pixel images.

Fig. 2a and b show a full sized (852×852 pixel) XCT training image, and an 852×852 pixel section of a SS training image, respectively. Fig. 2c and d show random 400×400 pixel cropped

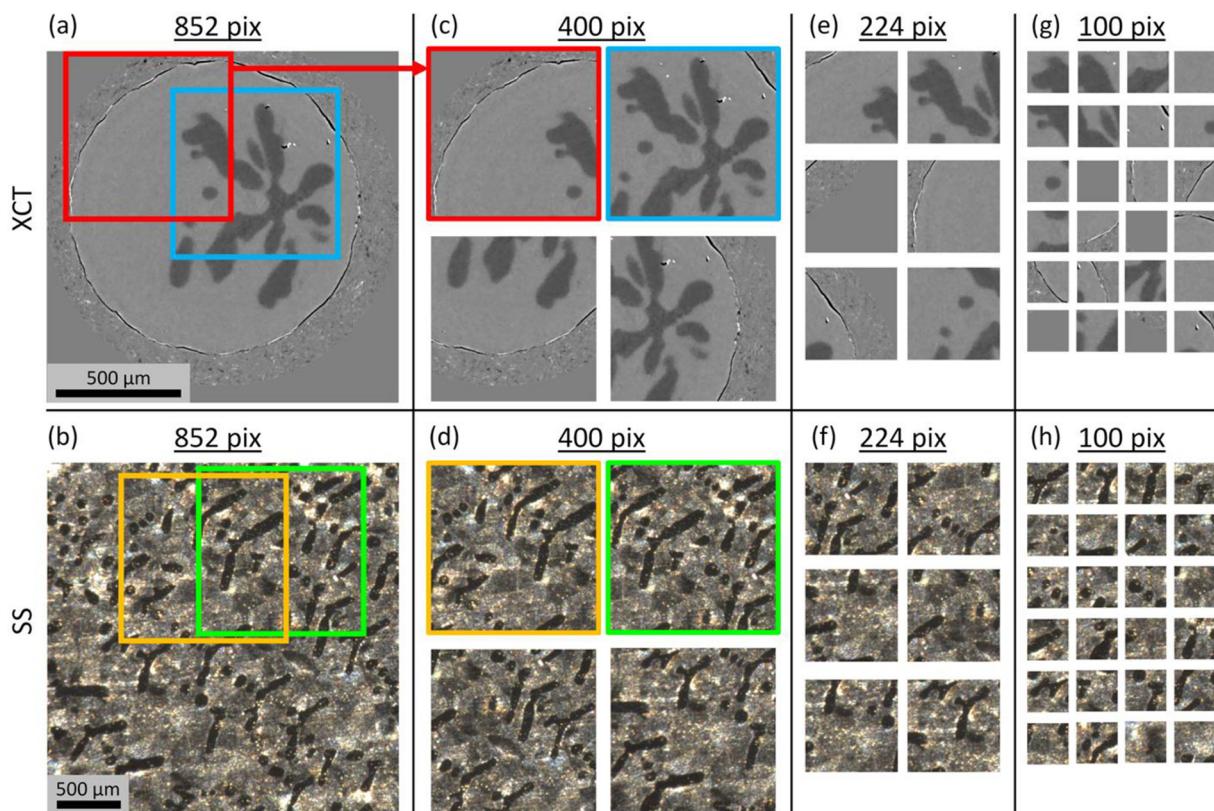


Fig. 2. (a) A 852×852 pixel full-sized training image, and (b) 852×852 pixel section of a SS training image; (c) and (d), 400×400 pixel sections of the XCT and SS images, respectively. The blue and red boxes in (a) and (c) show the same portion of the XCT image, and similarly for the boxes in (b) and (d); (e) and (f) 224×224 pixel sections of the XCT and SS training images, respectively; (g) and (h) 100×100 pixel sections of the XCT and SS training images, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

sections of the XCT and SS training images, respectively. The smaller images were created by taking a full-sized training image and selecting a random portion of specified dimensions. As shown using red and blue boxes in Fig. 2c, and using orange and green boxes in Fig. 2d, the selected images often contain overlapped portions of the full-sized images. Fig. 2e and f show 224×224 pixel sections, while Fig. 2g and h show 100×100 pixel sections. The side-by-side comparisons of four image sizes in Fig. 2 shows the microstructural information present in the smaller sections.

The number of images used for training was also varied. For NNs trained with 50×50 pixel and 100×100 pixel images, the number of images used was: 25, 50, 100, 500, 1000 and 5000. For NNs trained with 224×224 and 400×400 pixel images, the number of images was similarly varied from 25 to 1000. Using 5000 images with 224×224 and 400×400 pixel images would create an unnecessary amount of overlap, and would prolong the training time without adding new data. Also, as discussed in Section 3.3, NN performance does not always increase with more data. For this study on image size versus number of images used, a total of 52 NNs were trained (26 on XCT images, and 26 on SS images).

2.4. NN training

MATLAB was used for NN training and testing. The chosen network architecture (SegNet) consists of an encoder network, a decoder network, and a final pixel-wise classification layer [10]. The encoder network is composed of the first 13 convolutional layers from the VGG-16 classification network [36]. The decoder network has 13 layers, each corresponding to a layer in the encoder network. Finally, a soft-max classifier is used to produce class probabilities for each pixel independently [10].

Image augmentation was employed to create more training data. Each time an image was uploaded during training (example images shown in Fig. 2c-h), it was first randomly augmented using rotations between -45 and $+45$ degrees and translations of any number of pixels from -10 to $+10$ in the vertical and/or horizontal directions. Furthermore, each image had a 50% chance to be vertically and/or horizontally reflected. Class balancing was used to mitigate the effect of having unequal class frequencies (the XCT and SS pixel datasets contain 4.4% and 18.1% dendrite pixels, respectively).

All NNs were trained using desktop Mac Pro computers (2010 or newer), none of which had a graphical processing unit (GPU). Training time ranged from ~ 1.5 h to ~ 280 h (details in Supplementary data), and was dependent on input image size, number of images, number of epochs required for convergence, and the number of other NNs being trained concurrently. Details of how computational resources impact training time are outside the scope of this paper. Common training hyperparameters were used [21,36]: stochastic gradient decent with momentum solver, 0.9 momentum, 0.01 learning rate and 0.0005 L2 Regularization. During training, images were individually augmented, randomly re-ordered, grouped into 4-image mini-batches, and uploaded into the NN. After each epoch (a full pass through all the images), the NN was applied to the validation set, and accuracy and loss were calculated. The loss is a single overall measure of the cost of misidentifying pixels [37]. The training images were then shuffled, re-augmented, and re-used to further adjust NN parameters. Training was automatically stopped if the validation loss did not improve after 20 epochs from the previous minimum loss value, or manually stopped if after 35 epochs the loss was only decreasing by an average of $< 0.01\%$ per 20 epochs.

Because training images are shuffled after each epoch, and combined with the inherent randomness of data augmentation, two NNs trained on the same images and with the same hyperparameters will not

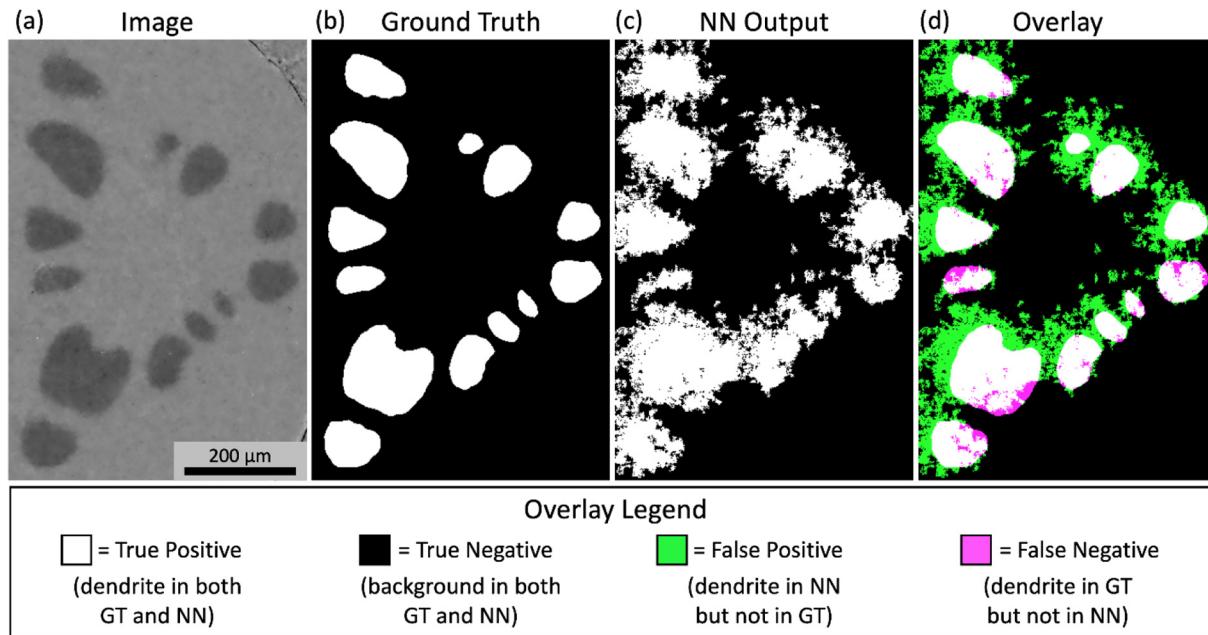


Fig. 3. (a) A section of an XCT test image; (b) the corresponding ground truth segmentation; (c) the output from a poorly-performing NN; and (d) overlay of the ground truth and NN output.

necessarily yield the same results. Some of the best NNs in this study were trained multiple times to ensure reproducibility. For example, five SS NNs were trained using one thousand 224×224 pixel images, and the standard deviation in accuracy was 0.17%. The variation was generally higher for NNs trained with fewer and smaller images.

2.5. Segmentation analysis

After training, each NN was applied to the same two XCT test images (or one SS test image). Using the computer hardware mentioned in Section 2.4, segmenting one 852×852 pixel XCT image takes ~ 7.5 s while segmenting one 1689×985 pixel SS image takes ~ 17.5 s.

The NN outputs and ground truths are overlaid to visually inspect the NN predictions and to show the distribution of correct and misidentified pixels. Fig. 3a shows a portion from one of the XCT test images, while Fig. 3b shows the corresponding ground truth. Fig. 3c shows the output of one of the poorly-performing XCT NNs. The ground truth and NN output are overlaid in Fig. 3d and colored according to the legend at the bottom of Fig. 3. White areas are correctly classified as dendrite (true positive) while black areas are correctly classified as background (true negative). Green pixels are misclassified as dendrite (false positive) while pink pixels are misclassified as background (false negative). Thus, over-segmented regions are shown in green and under-segmented regions are pink.

While the overlays give a visual demonstration of a NN's predictive ability, three objective metrics are also calculated. Global accuracy (Acc) is the fraction of correctly identified pixels divided by the total number of pixels in the image. For the overlay in Fig. 3d, Acc is calculated as the number of true positives (white) and true negatives (black), divided by the total number of positives and negatives together (white + black + green + pink). Acc is often higher in segmentations with a large area fraction of correctly predicted background pixels. Thus, optimizing for Acc alone can lead to NNs which are nominally accurate but perform poorly on the classes of interest (dendrites in the present study).

The intersection over union (IoU) metric, also known as a Jaccard score, is similar to Acc except it does not include true negative (black) pixels, and thus is less sensitive to the amount of background in the image. IoU is the number of true positives divided by the sum of: true

positives, false positives, and false negatives. For the overlay in Fig. 3d, the IoU for the dendrite class is calculated as: white / (white + pink + green). Henceforth, the IoU is calculated for the dendrite class only, as it is of highest interest for this study.

The boundary F-1 score (BF1) [22] is a measure of how close the boundary of segmented bodies in the NN outputs match with the boundaries in the ground truths. The mathematical details are described in reference [22]. A threshold of 2 was used, meaning that a BF1 score of 100% is achieved if all NN predicted boundary pixels are < 2 pixels away from all boundary pixels in the ground truths.

3. Results

3.1. Segmentation cleaning

Before evaluating NN performance and calculating output metrics, the NN segmentations were adjusted to exclude border pixels. It was observed that almost all NNs trained in this study have false positive pixels near the image edges. Fig. 4a and b show an XCT test image and a section of a SS test image, respectively. Fig. 4c shows an overlay from an XCT-trained NN applied to the image in Fig. 4a, while Fig. 4d shows an overlay from an SS-trained NN applied to the image in Fig. 4b. The green features near the overlay edges (some examples are marked with red arrows in Fig. 4c and d) are areas the NNs incorrectly classified as dendrite. The green patch at the bottom right of Fig. 4c (corresponding to the bottom right in Fig. 4a) is clearly an artifact of the NN since all of the corresponding pixels in the original image have the exact same luminance. To further verify that the edge effects are not caused by features in the images themselves, smaller 300×300 pixel center sections of the test images were cropped (red dashed boxes in Fig. 4a and b). The same NNs were applied to the smaller sections, and as shown in Fig. 4e and f, the green edges are still present.

The false-positive edge pixels are an artifact of the image sampling process. However, masking the image edges is computationally straightforward and can be systematically done to every image in an experiment. For the XCT data, it is known that dendrite features can only be present within a 310 pixel radius near the center of the image. Thus, all pixels outside of the near-center region were modified to be correctly segmented as background. XCT segmentations were further

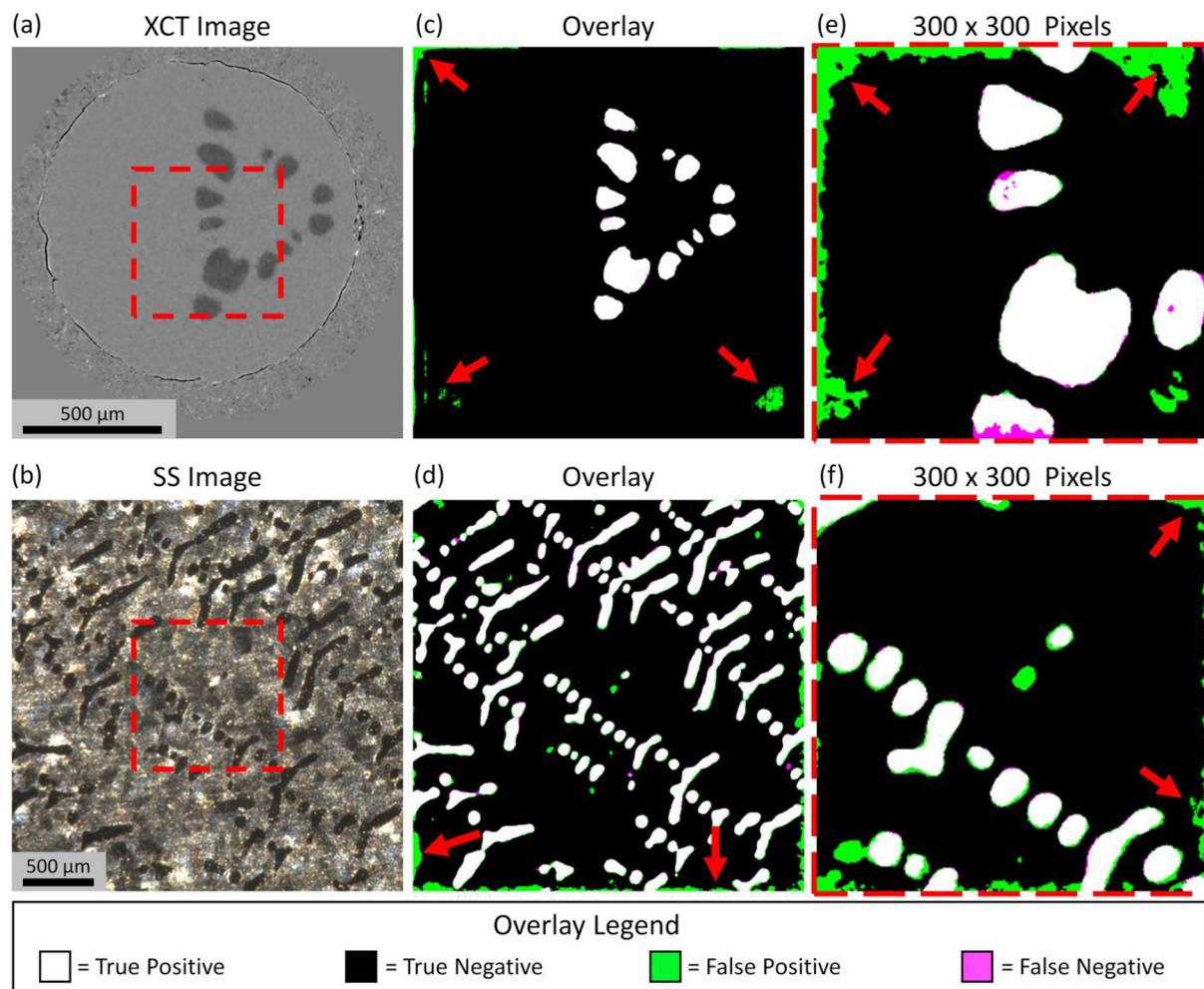


Fig. 4. (a) A full-sized raw XCT test image; (b) a section of a raw SS image; the overlays from a well-performing (c) XCT NN and (d) SS NN; (e) overlay from the 300 × 300 pixel center portion of the XCT image; and (f) from a SS output. The red arrows point to green (false positive) regions at the edges of images. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

cleaned using a small isolated body filter. In the XCT dataset, features with < 25 pixels are too faint to be confidently classified as dendritic. Therefore, any < 25 pixel clusters in the segmentations were also removed. For all NN outputs on the SS test image, the outer 50 pixels were removed from each edge, resulting in an image size of 1589 × 885 pixels. In the SS dataset, bodies with < 10 pixels were considered too small to be reliably classified as dendrite and were removed.

Acc, IoU, and BF1 metrics were calculated only for the cleaned NN outputs. Thus, the scores reported in this study are generally higher than the raw scores reported in other literature. However, using the cleaned versions better shows the final segmentations possible (which are of most interest to researchers), and allows for better comparison with conventional segmentation methods.

3.2. Training image size and number study

Each of the 26 XCT NNs and 26 SS NNs were applied to the test images. All of the corresponding Acc, IoU, and BF1 scores are reported in the Supplementary data. The four heatmaps in Fig. 5 are a visual representation of the NN performances. The XCT and BF1 and IoU scores are shown in Fig. 5a and c, respectively. NNs with higher IoU scores are more yellow than red. Nearly all of the XCT NNs trained with 50 × 50 pixel images under-performed (< 75% IoU and < 75% BF1), regardless of the number of images used. Of the NNs trained with

100 × 100 pixel images, there is no clear trend in score as a function of number of images. The 100 × 100 pixel NNs trained with 200 images out-performed the NNs trained with 100 images and with 500 images. The 224 × 224 pixel and 400 × 400 pixel XCT NNs overall performed better than the NNs trained using smaller images, and exhibit a trend of improvement with increasing number of training images. The NN trained with 224 × 224 pixel and 1000 images was deemed the best as it had the highest Acc (99.9%), IoU (96.9%), and BF1 (98.1%) scores. These scores are higher than segmentations produced using the conventional Otsu's method (99.5% Acc, 92.4% IoU and 85.4% BF1) and EM/MPM (99.1% Acc, 86.0% IoU and 35.4% BF1).

The heatmaps in Fig. 5b and d show the SS IoU and BF1 scores, respectively. NNs with higher BF1 scores are more green than blue. Unlike the XCT NNs, the SS NNs exhibit a trend of increased performance with number of images at all four image sizes. As also observed for the XCT NNs, the 100 × 100 pixel SS NN scores are higher at 1000 images than at 5000 images. There are two high-performing SS NNs with nearly-identical metrics. The NN trained with 1000 images of 100 × 100 pixel has Acc of 97.7%, IoU of 88.7%, and BF1 of 89.8%. The NN trained using 1000 images of 224 × 224 pixel has Acc of 97.8%, IoU of 89.2%, and BF1 of 89.3%. The overlays of both NNs were visually inspected and the 224 × 224 pixel NN was deemed the best. The NN out-performed the conventional Otsu's method (91.59% Acc, 59.57% IoU and 31.99% BF1) and EM/MPM (91.39% Acc, 62.82% IoU and 47.02% BF1).

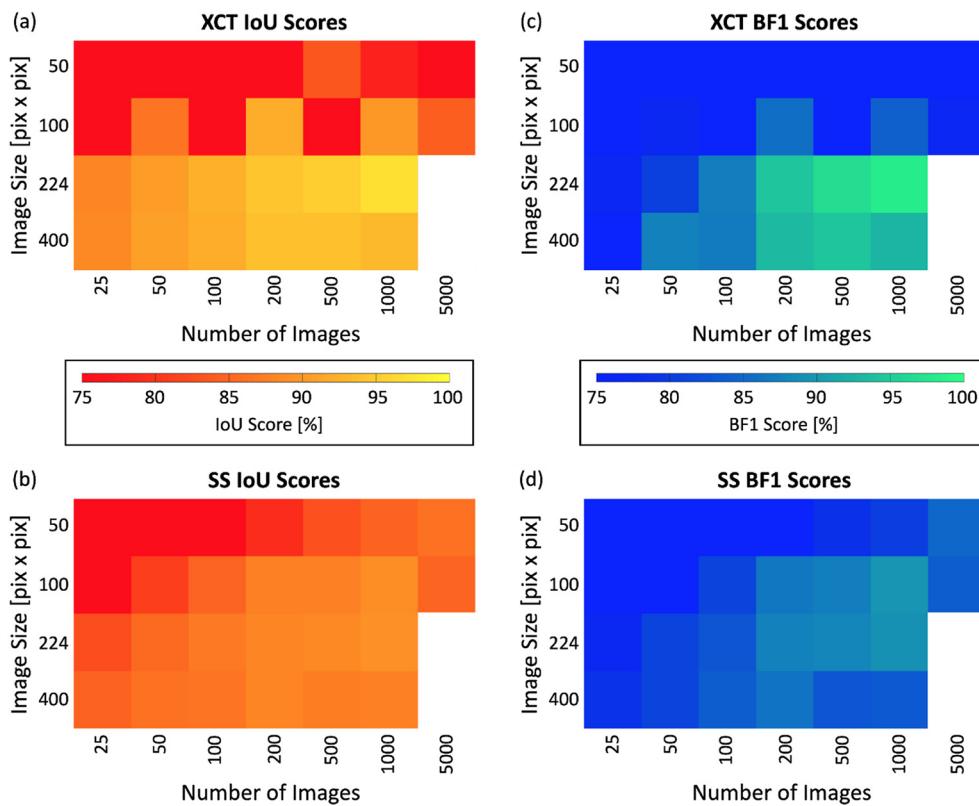


Fig. 5. Heatmaps indicating (a) IoU scores for XCT NNs; (b) IoU scores for SS NNs; (c) BF1 scores for XCT NNs; and (d) BF1 scores for SS NNs. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

3.3. Best NNs

The overlays in Figs. 6 and 7 were output by the best XCT and SS NNs (each trained using one thousand 224×224 pixel images), respectively. These figures highlight the unique challenges faced during segmentation of each dataset. Fig. 6a shows the same XCT image that was poorly segmented by the conventional methods in Fig. 1. The overlay in Fig. 6b shows very few false positive and false negative pixels on the XCT image. Fig. 6c shows the red boxed region from Fig. 6a, while Fig. 6d shows the corresponding overlay. The black and white artifacts in Fig. 6c (red arrow) and diffuse boundaries between dendrite bodies (blue arrow) are challenging to segment using the traditional

methods. However, Fig. 6d shows that the NN ignored the black and white artifacts, and the dendritic features were not welded together.

Fig. 7a shows a portion of the larger SS test image and Fig. 7b shows the NN overlay. The many bright speckles along with diffuse dark regions make the images difficult to segment using standard methods. Fig. 7c and d show the red-boxed region of the raw image and overlay, respectively. The speckled regions (such as the one denoted by the blue arrow) are correctly segmented. The gold arrow indicates a region that the NN misclassifies as a dendrite body. During ground truth creation, the dark smudge was manually identified as background after inspecting the sections above and below this image. The NN did not have access to information from nearby images and the section was

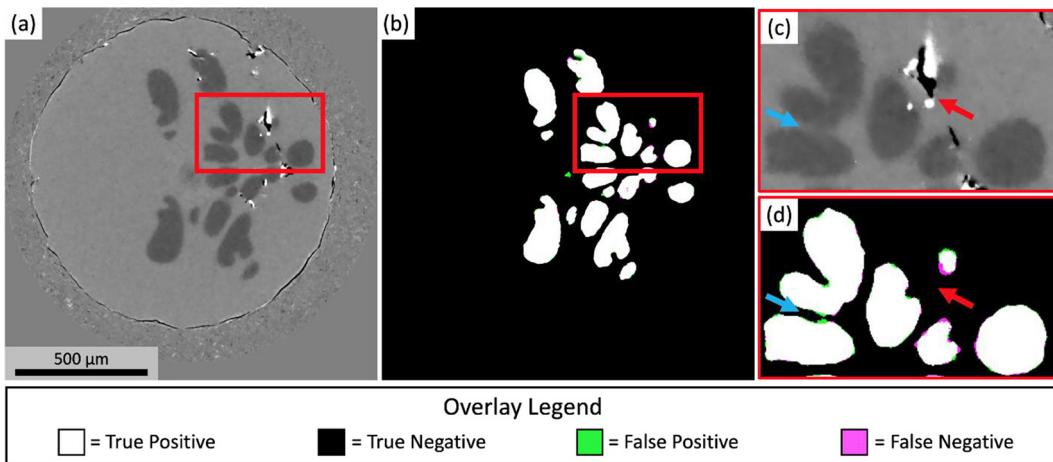


Fig. 6. (a) A full-sized XCT image; (b) overlay from the best-performing XCT NN; (c) higher magnification image of the red-boxed region in (a) showing black and white contaminant regions (red arrow), and diffuse dendrite boundaries (blue arrow); and (d) corresponding overlay. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

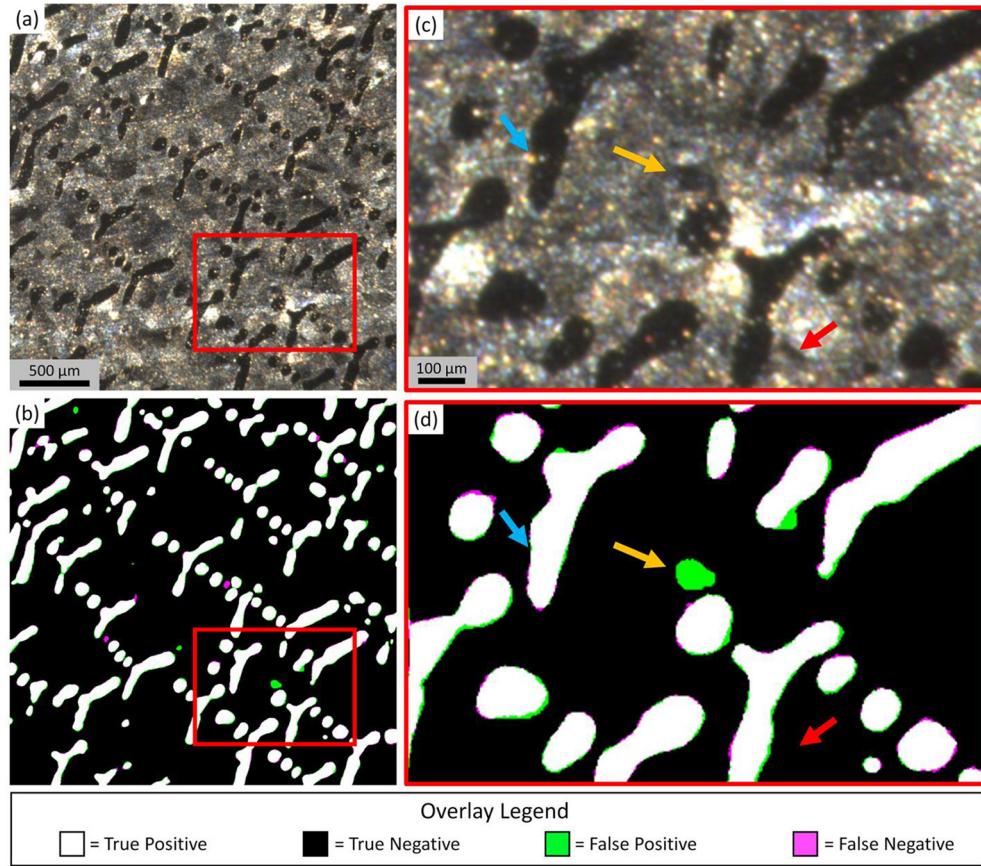


Fig. 7. (a) A portion of the SS test image; (b) a NN overlay of (a) using the best SS NN; (c) the red boxed region from (a) with arrows indicating regions that are difficult to segment; and (d) the corresponding NN overlay. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

misclassified. However, the smudge shown with a red arrow was correctly identified as background.

4. NN transferability

4.1. Overview

The results presented in Section 3 demonstrate the success in using machine learning to segment large materials datasets. NNs were optimized for segmenting one XCT and one SS dataset, and the outputs were more accurate than those possible from other segmentation methods. Creating ground truths and training the NNs is time-consuming, thus, in an effort to get the most use out of the previous work, the NNs were also tested on other materials datasets: XCT images containing ring artifacts, XCT images from an Al-Cu coarsening experiment, and SS images containing larger dendritic features. The results and attempts to improve the transferability of trained NNs to new datasets are reported.

4.2. Ring study

As previously described in Section 2.2, ring artifacts are commonly found in XCT datasets as a result of the reconstruction process. While ring-removal codes exist [35], and background subtraction procedure can sometimes be used (as described in Section 2.2), these techniques may not work well on every XCT dataset. Thus, to test the ability of the previously-trained XCT NNs to remove rings, the best NN (trained using one thousand 224×224 pixel images) was applied to a version of the XCT test image that still contained rings. Also, a new NN was trained using the exact same training images that included rings. All other procedures were identical to those used for the best XCT NN. The NN trained on images containing rings was applied to both types of test images (with and without rings). Segmentation outputs are shown in Fig. 8, and corresponding metrics are summarized in Table 1.

Fig. 8a shows the test image without rings, while Fig. 8b shows the same test image with the rings left in. The overlay from the best XCT NN as applied to Fig. 8a is shown in Fig. 8c. All scores are $> 96\%$, indicating an acceptable segmentation. Fig. 8d shows an overlay of the best XCT NN as applied to the test image containing rings. The pink pixels are mostly found in areas corresponding to the bright rings in Fig. 8b. These false-negative regions are primarily due to rings causing the dendrite luminance to be raised above the range the NN was trained for. The large areas of misidentified pixels indicate that the NN has poor transferability when applied to images with rings.

The new NN (trained using images containing rings) was applied to the test image without rings, and the overlay is shown in Fig. 8e. The output is acceptable, although some dendrites are slightly over-segmented (green pixels around the bodies). This is also reflected as a lower BF1 score (92.7%) compared to the NN in Fig. 8c (98.1%). Overall, the NN trained with rings was largely transferable to images without rings. Fig. 8f shows the overlay from the new NN applied to the test image in Fig. 8b. Most dendrite bodies are well segmented, except for a few pink and green spots. The new output in Fig. 8f is improved compared to the segmentation in Fig. 8d.

The best segmentations were obtained when removing the rings in both the training images and in the entire dataset. The NN trained on images without rings had poor transferability to the image containing rings, showing that NNs are sensitive to dendrite luminance. As shown in Table 1, the new NN (trained with rings) performed acceptably well on both images with and without rings. The new NN could likely be improved if more unique images and corresponding ground truths were used for training such that the NN would have more examples of the large luminance and contrast variation caused by the rings. This study shows that a NN trained using unprocessed images (with rings) can be used to segment XCT datasets that contain ring artifacts.

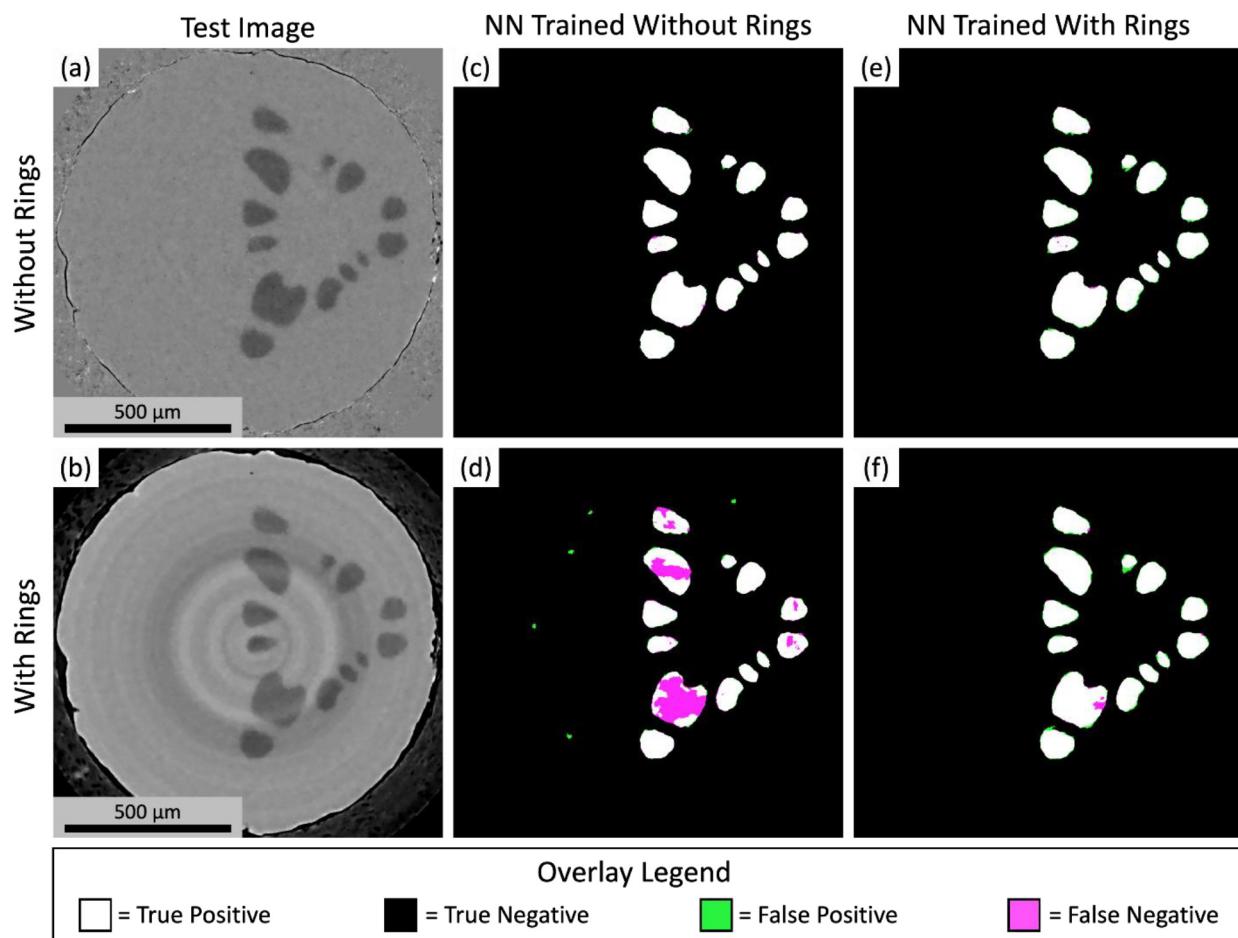


Fig. 8. (a) The XCT test image; (b) the same XCT image but the rings were not removed; (c) and (d) overlays from the best XCT NN (trained without rings) applied to the test image without rings and with rings, respectively; (e) and (f) overlays from the NN trained with rings applied to images (a) and (b), respectively.

Table 1
Performance metrics for the overlays shown in Fig. 8c–f.

	NN trained without rings	NN trained with rings
Test image without rings	99.9% Acc	99.7% Acc
	96.9% IoU	94.0% IoU
	98.1% BF1	92.7% BF1
Test image with rings	98.8% Acc	99.7% Acc
	70.8% IoU	92.2% IoU
	82.3% BF1	88.9% BF1

4.3. Al-Cu coarsening

To further test XCT NN transferability, the best NN (without rings) was applied to images from an Al – 12 wt%Cu isothermal coarsening experiment. The dataset was obtained at the same beamtime as the Al-Zn experiment, but instead of recording the initial dendrite formation, the sample was quickly cooled and held at ~600 °C, thus containing a larger area fraction of coarser dendrites (Fig. 9a). The ground truth was obtained using the same procedure described in Section 2.2, and the best XCT NN (trained using one thousand 224 × 224 pixel Al-Zn XCT images) was applied. The resulting overlay is shown in Fig. 9b. Note that the same mask was applied as for the Al-Zn NNs, thus the pixels far outside of the center region are taken to be correctly classified as background. Although most dendrite bodies close to the image center were segmented (white pixels in Fig. 9b), much of the dendritic structure was missed (pink pixels). This is primarily due to the dendrite luminance gradually increasing (becoming brighter) away from the image center.

The mean luminance and contrast of dendrites in the original Al-Zn dataset are 90.8 & 8.5, while the dendrites in the Al-Cu image have 97.7 & 17.7. In an attempt to more closely match the original NN training data, the luminance of the entire Al-Cu image was first decreased by 7 (on a 0–255 grayscale), and the NN was re-applied. The results did not show much improvement (overlay not shown). However, decreasing the value of each pixel in the raw image by a value of 40 before applying the NN greatly improved the segmentation results, as shown in Fig. 9c with overlay scores: 97.8% Acc, 82.2% IoU, and 85.3% BF1.

This study shows that a NN trained using an Al-Zn solidification dataset can be used to segment Al-Cu coarsening images. Lowering the Al-Cu image luminance to better match the average dendrite luminance from the training data had little impact on segmentation, but greatly darkening the entire Al-Cu image yielded much better results. Although the luminance adjustment technique may not work for other types of datasets, it is computationally straightforward and can be attempted in the quest to improve NN transferability.

4.4. New SS sample

The best SS NN was trained on images from a Pb – 68 wt%Sn sample coarsened for 1.6 h. To test transferability, this network was applied to new SS images from a Pb – 74 wt%Sn sample coarsened for 27 h. The new dataset contains darker images (luminance and contrast of 28.2 & 11.8 compared to 58.5 & 22.9) with dendrites that are on average larger (2062 pixels compared to 1012 pixels) and have a higher area-perimeter ratio (11.5 pixels compared to 7.5 pixels). Area-perimeter ratio is the 2D equivalent of inverse surface area to volume (S_v^{-1}), commonly

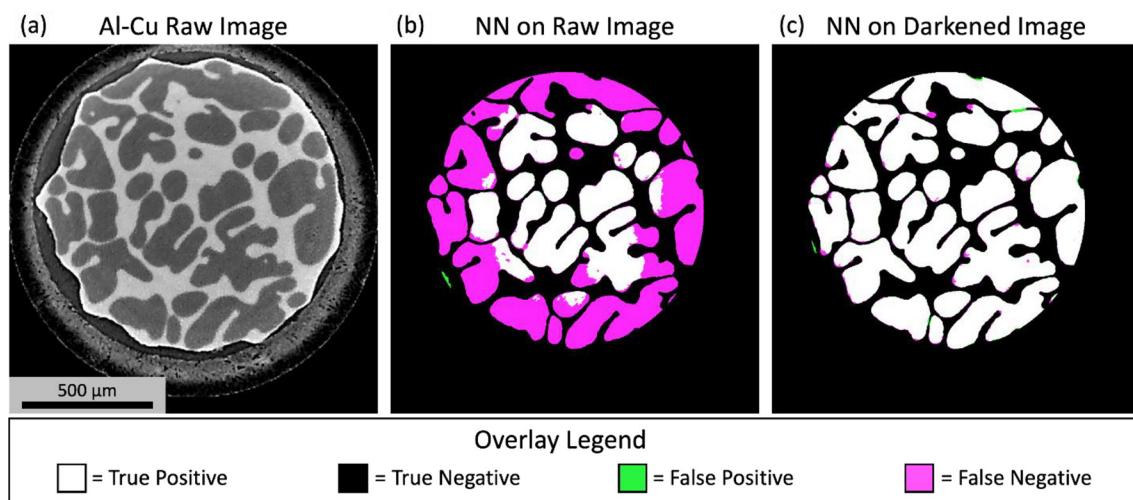


Fig. 9. (a) A raw image from an Al-Cu solidification experiment; (b) overlay using the best NN trained on Al-Zn data on image (a); and (c) overlay from the same NN as applied to a darkened version of image (a).

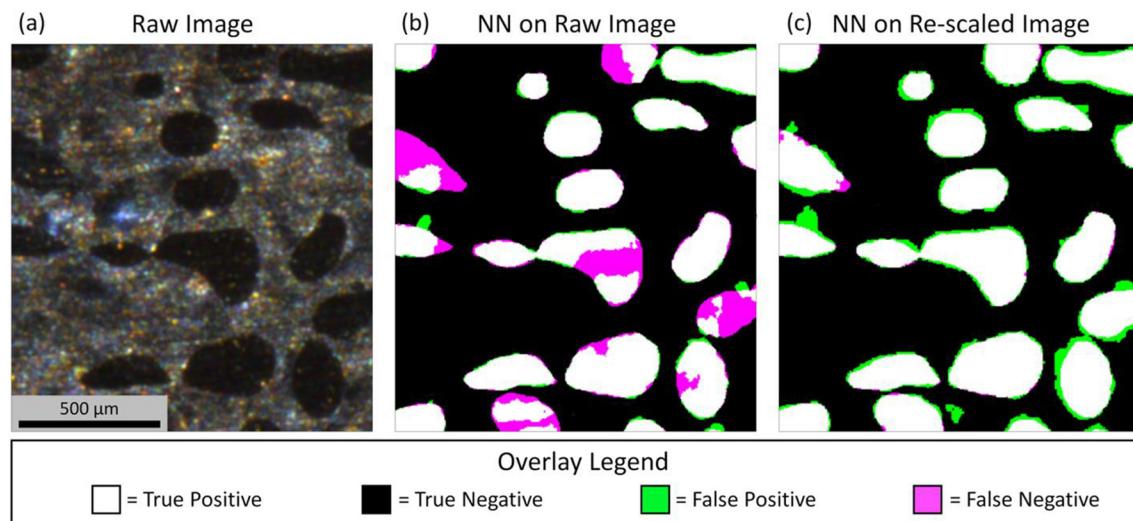


Fig. 10. (a) A raw image from a 27 h coarsened Pb – 74.4 wt%Sn sample; (b) overlay output from applying the best SS NN to (a); and (c) the overlay after scaling down the image to 81%, applying the best NN, and scaling up the segmentation to original dimensions.

used to describe solidification microstructures [1]. The NN performance on an image from the new dataset (Fig. 10a) was visually poor, as seen by the many pink and green pixels in the overlay in Fig. 10b. In an attempt to improve NN performance, the luminance and contrast of the new image were changed to better match those of the training dataset, similar to the technique used in Section 4.3. This, and multiple other attempts at adjusting pixel luminance, failed to improve the results and often led to worse NN segmentations (overlays not shown).

To better match the dendrite size and area-perimeter ratio in the training dataset, the new image was scaled down to 65% of its original size. The NN was applied to the smaller image, and the resulting segmentation was then up-scaled to match the original image dimensions. This procedure showed some improvement (overlay not shown). However, scaling the image down to 81% of its original size (found empirically) yielded the best segmentation. The overlay is shown in Fig. 10c, and the metrics are: 96.5% Acc, 89.0% IoU, and 70.4% BF1. This result illustrates that dendrite size and shape also impact NN transferability, not just pixel luminance. Image re-scaling can introduce artifacts because changing the number of pixels also affects the resolution. This technique might be acceptable in for some datasets and is often more computationally straightforward than training a new NN.

5. Discussion

5.1. Image size and number

We show that a properly trained NN is capable of segmenting complex microstructures. The NN segmentations are quite accurate, far out-perform the conventional segmentation techniques, and require little time to produce once the NN is trained. There are three main observations from the image size and number study. First, using a large number (1000) of smaller (224×224 pixel) images taken from the full-sized training data led to improved NN performance. Despite differences in dendrite shapes and distributions throughout the images, and thus performed poorly, while the NNs trained with 400×400 pixel images

The XCT and SS NNs trained with 50×50 pixel images likely did not capture enough of the relevant dendrite information, such as full dendrite shapes and distributions throughout the images, and thus performed poorly, while the NNs trained with 400×400 pixel images

Table 2

XCT and SS dataset statistics. Luminance (L) and contrast (C) are normalized to a 0 to 255 scale.

	XCT dataset	SS dataset
Total number of dendrite bodies in dataset [#]	450	2688
Number of dendrite bodies in images [#]	0 to 18	286 to 335
Average number of dendrite bodies per image [#]	11 ± 6	299 ± 15
Average dendrite area (total dendrite area / total number of dendrites) [pixels]	2987 ± 3525	1012 ± 1069
Dendrite area-perimeter ratio [pixels]	13.6	7.5
Average L & C in dataset [0–255]	128.5 & 13.0	109.5 & 46.1
Average dendrite L & C [0–255]	90.8 & 8.5	58.5 & 22.9
Average background L & C [0–255]	130.2 & 10.3	120.8 & 42.1

likely did not have enough unique training data to fit the large number of internal parameters. For both XCT and SS datasets, the best NNs used 224×224 pixel images. As previously mentioned, the VGG-16 encoder portion of the SegNet architecture is most often trained with 224×224 pixel images when used for image classification [36]. Although outside the scope of this paper, the authors have also trained SegNet-based NNs to segment images of feline spinal cords (not yet published). The best-performing feline spine NN also used 224×224 pixel images even though the features of interest were roughly twice the size of the dendrites in the XCT studies. This further confirms that the internal VGG-16 portion of the SegNet architecture has some impact on NN performance.

The 100×100 pixel XCT NNs performed much worse than the 224×224 pixel versions, while the 100×100 pixel SS NN trained with 1000 images performed nearly-identically to the 224×224 pixel version. Although there are many factors involved, one is that SS dendrites are smaller than the XCT dendrites. Examples of XCT and SS images are shown in Fig. 4a and b, respectively, and dataset statistics from the training, validation, and test images are summarized in Table 2. The XCT images each contain between 0 and 18 individual dendrite bodies (11 ± 6 per image), while the SS images contain between 286 and 335 dendrite bodies (299 ± 15 per image). Thus, the XCT dataset has a larger relative variability in number of bodies in each image than the SS dataset. The average individual dendrite size (total dendrite area / number of dendrite bodies) is nearly three times larger in the XCT images than in the SS images. Because dendrites in the XCT dataset are overall larger and fewer than in the SS dataset, larger XCT training images are required to capture the essential morphological features and distributions than for the SS dataset. This partially rationalizes why the 100×100 pixel XCT NN performed worse than the 224×224 pixel XCT NN, while there was not as much difference between the 100×100 pixel and the 224×224 pixel SS NNs.

Finally, the XCT NNs overall performed better than the SS NNs. This is likely due to the XCT images being less complex than the SS images. As reported in Table 2, the higher average luminance and lower contrast in the XCT dataset indicates that XCT images are overall brighter and more uniform. By isolating the dendrite and background pixels separately, the luminance and contrast of each class was calculated. Dendrites in the XCT images are overall brighter and have less variability than the SS images. Similarly, the XCT background pixels are less varied than those in the SS images. Overall, the much higher variability in pixel luminance makes the SS images quantitatively more complex than XCT images, and this is likely the main reason why SS NN metrics were lower than XCT NN metrics.

5.2. Transferability

XCT and SS NN transferability was tested by segmenting images containing rings, Al-Cu datasets, and a second SS sample. In general, the best segmentations are obtained when NNs are trained on images from the same datasets that will be segmented by the NN. Using an already-

trained NN on other datasets can be successful, but segmentations are often of lower quality. NNs are most sensitive to object luminance, size, shape, and distribution. Thus, the more similar these characteristics are to the training data, the more successful the NN will be.

There are two primary ways to increase NN transferability. The first method is to generalize the data used for NN training. For example, if the training images contain a large variability in feature size (such as the dendrites in the XCT experiment), the resulting NNs will be more successful at segmenting large and small objects. Conversely, if the NNs are trained using a small and consistent feature size (such as the dendrites in the SS datasets), the NN will be less transferrable to datasets containing larger features (as shown in Section 4.4). Generalizing the training data can also be achieved through image augmentation. In the present study, training images were only translated, rotated, and reflected. However, it is also possible to up- and down-scale the training images, as well as add luminance variations. These techniques would likely have led to improved transferability for both the XCT and SS NNs. The downside to increasing NN transferability by generalizing training data is that more unique training images are required to capture all of the diverse features. Thus, it is likely that more ground truths, training time, and computational resources are needed.

The second way of increasing transferability is to adjust the new dataset to better match the images that were used for NN training. As shown in Sections 4.3 and 4.4, both XCT and SS NNs performed better after the image luminance and/or scaling was changed in the second datasets. These changes are computationally straightforward. However, simply matching the average training image statistics was not enough to obtain accurate segmentations. Finding a combination of adjustments that leads to satisfactory segmentations can be difficult, and even impossible if a NN was highly-tuned to work on a specific dataset. Making changes to the new images instead of training a new NN has the major advantage of not requiring new ground truths.

5.3. Future work

The full potential of using machine learning techniques for segmentation of large materials science datasets has not yet been realized. The present study only used SegNet, but there are many other architectures available, and even more in development. It is possible that different NN architectures could lead to increased segmentation accuracy.

In addition, now that the high-performing XCT and SS NNs presented in this study are available, they can be used to facilitate creating ground truths for new datasets. For example, the false pixels in the overlay shown in Fig. 9c could be easily manually fixed, thus creating a new ground truth image for training a new NN for Al-Cu coarsening datasets. Using NN outputs as a starting point can greatly reduce the time required to create new ground truths.

Finally, the effects of using more varied training images will be studied. For example, better performing and more transferrable NNs could be achieved if the training data was re-scaled in addition to being rotated, translated, and reflected. However, it would be ideal to have a universal NN capable of segmenting any materials science microstructures. Training such a transferrable NN would require a very large number of raw images, as well as high quality ground truths.

6. Conclusions

We find that neural networks (NNs) can be trained to perform highly accurate semantic segmentation on large materials datasets produced by x-ray computed tomography (XCT) and serial sectioning (SS) in a reasonable amount of time and using a limited amount of training data. We show that NN performance is improved if training is done using a large number of small sections that are sampled from the fixed number of training images. While the SS NNs do not match the predictive power of the XCT NNs, the segmented images are

quantitatively superior to what could be obtained using conventional segmentation techniques. In addition to the increase in segmentation accuracy, the amount of user time and computational resources needed to achieve well-performing NNs is not prohibitive. NN transferability was also tested by applying the highest performing XCT and SS NNs to other datasets. Initial segmentations are promising, and applying simple transformations to the raw images greatly improved NN performance. These results show that NNs will likely become the method of choice for segmenting large materials datasets.

Acknowledgements

The authors thank M. Rappaz (EPFL) and P. Jarry (Constellium) for fabricating the Al-Zn alloy, X. Xiao (BNL) for aid in XCT data acquisition, and E. Holm (CMU) for fruitful discussions. The XCT portion of this work was supported by the U.S. Department of Energy's Office of Science under grant number DE-FG02-99ER45782, and by the U.S. Department of Commerce, National Institute of Standards and Technology as part of the Center for Hierarchical Materials Design (CHiMaD) under the financial assistance award 70NANB14H012. The SS portion of this work was sponsored by National Aeronautics and Space Administration under grant number NNX16AR13G.

Data availability

The raw data (XCT and SS images) required to reproduce these findings are available to download from [[http://doi.org/10.18126/M2RM08](https://doi.org/10.18126/M2RM08)]. The processed data (ground truth segmentations) required to reproduce these findings are available to download from [[http://doi.org/10.18126/M2W93J](https://doi.org/10.18126/M2W93J)].

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.matchar.2020.110119>.

References

- [1] L. Ratke, P.W. Voorhees, Growth and Coarsening, Springer-Verlag Berlin Heidelberg, 2002, <https://doi.org/10.1007/978-3-662-04884-9>.
- [2] J.A. Dantzig, M. Rappaz, Solidification, EPFL press, 2016.
- [3] W. Kurz, D.J. Fisher, Fundamentals of Solidification, Third Edit, Trans Tech Publications Ltd, 1992, <https://doi.org/10.1002/crat.2170210909>.
- [4] Y. Leng, Materials Characterization: Introduction to Microscopic and Spectroscopic Methods, 2nd edition, Wiley, 2013, <https://doi.org/10.1002/9783527670772>.
- [5] D. Brandon, W.D. Kaplan, Microstructural Characterization of Materials, 2nd edition, Wiley, 2008, <https://doi.org/10.1002/9780470727133>.
- [6] D.J. Rowenhorst, P.W. Voorhees, Measurement of interfacial evolution in three dimensions, Annu. Rev. Mater. Res. (2012), <https://doi.org/10.1146/annurev-matsci-070511-155028>.
- [7] N. Otsu, A threshold selection method from gray-level histograms, IEEE Trans. Syst. Man. Cybern. 9 (1979) 62–66, <https://doi.org/10.1109/TSMC.1979.4310076>.
- [8] J. Canny, A computational approach to edge detection, IEEE Trans. Pattern Anal. Mach. Intell. PAMI-8 (1986) 679–698, <https://doi.org/10.1109/TPAMI.1986.4767851>.
- [9] M.L. Comer, E.J. Delp, Parameter estimation and segmentation of noisy or textured images using the EM algorithm and MPM estimation, Proc. - Int. Conf. Image Process. ICIP, 1994, <https://doi.org/10.1109/ICIP.1994.413651>.
- [10] V. Badrinarayanan, A. Kendall, R. Cipolla, SegNet: a deep convolutional encoder-decoder architecture for image segmentation, IEEE Trans. Pattern Anal. Mach. Intell. (2017), <https://doi.org/10.1109/TPAMI.2016.2644615>.
- [11] S.M. Azimi, D. Britz, M. Engstler, M. Fritz, F. Mücklich, Advanced steel microstructural classification by deep learning methods, Sci. Rep. (2018), <https://doi.org/10.1038/s41598-018-20037-5>.
- [12] B.L. DeCost, T. Francis, E.A. Holm, Exploring the microstructure manifold: image texture representations applied to ultrahigh carbon steel microstructures, Acta Mater. (2017), <https://doi.org/10.1016/j.actamat.2017.05.014>.
- [13] B.L. Decost, E.A. Holm, A computer vision approach for automated analysis and classification of microstructural image data, Comput. Mater. Sci. (2015), <https://doi.org/10.1016/j.commatsci.2015.08.011>.
- [14] A. Chowdhury, E. Kautz, B. Yener, D. Lewis, Image driven machine learning methods for microstructure recognition, Comput. Mater. Sci. (2016), <https://doi.org/10.1016/j.commatsci.2016.05.034>.
- [15] T.M. Smith, P. Bonacuse, J. Sosa, M. Kulis, L. Evans, A quantifiable and automated volume fraction characterization technique for secondary and tertiary γ' precipitates in Ni-based superalloys, Mater. Charact. (2018), <https://doi.org/10.1016/j.matchar.2018.03.051>.
- [16] J.P. Papa, R.Y.M. Nakamura, V.H.C. De Albuquerque, A.X. Falcão, J.M.R.S. Tavares, Computer techniques towards the automatic characterization of graphite particles in metallographic images of industrial materials, Expert Syst. Appl. (2013), <https://doi.org/10.1016/j.eswa.2012.07.062>.
- [17] J. Ker, L. Wang, J. Rao, T. Lim, Deep Learning Applications in Medical Image Analysis, IEEE Access, (2017), <https://doi.org/10.1109/ACCESS.2017.2788044>.
- [18] H. Noh, S. Hong, B. Han, Learning deconvolution network for semantic segmentation, Proc. IEEE Int. Conf. Comput. Vis., 2015, <https://doi.org/10.1109/ICCV.2015.178>.
- [19] V. Badrinarayanan, A. Kendall, R. Cipolla, SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation, CoRR, abs/1511.00561, 2015. <http://arxiv.org/abs/1511.00561>.
- [20] L.C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, A.L. Yuille, DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs, IEEE Trans. Pattern Anal. Mach. Intell. (2018), <https://doi.org/10.1109/TPAMI.2017.2699184>.
- [21] K. Chatfield, K. Simonyan, A. Vedaldi, A. Zisserman, Return of the Devil in the Details: Delving Deep Into Convolutional Nets, (2014), <https://doi.org/10.5244/C.28.6>.
- [22] G. Csurka, D. Larlus, F. Perronnin, What is a good evaluation measure for semantic segmentation? Proceedings Br. Mach. Vis. Conf. 2013, 2013, <https://doi.org/10.5244/C.27.32>.
- [23] C. Shashank, Kaira, X. Yang, V. De Andrade, F. De Carlo, W. Scullin, D. Gursoy, N. Chawla, Automated correlative segmentation of large Transmission X-ray Microscopy (TXM) tomograms using deep learning, Mater. Charact. 142 (2018) 203–210, <https://doi.org/10.1016/j.matchar.2018.05.053>.
- [24] D.S. Bulgarevich, S. Tsukamoto, T. Kasuya, M. Demura, M. Watanabe, Pattern recognition with machine learning on optical microscopy images of typical metallurgical microstructures, Sci. Rep. (2018), <https://doi.org/10.1038/s41598-018-20438-6>.
- [25] P. Dube, B. Bhattacharjee, E. Petit-Bois, M. Hill, Improving Transferability of Deep Neural Networks, (2018) (doi:arXiv:1807.11459v1).
- [26] J. Cho, K. Lee, E. Shin, G. Choy, S. Do, How Much Data is Needed to Train a Medical Image Deep Learning System to Achieve Necessary High Accuracy? (2015), <https://doi.org/10.1111/j.1471-4159.2007.05072.x>.
- [27] I. Goodfellow, Y. Bengio, A. Courville, Deep Learning, The MIT Press, 2016, <https://doi.org/10.1007/s10710-017-9314-z>.
- [28] J. Shotton, J. Winn, C. Rother, A. Criminisi, TextronBoost for image understanding: multi-class object recognition and segmentation by jointly modeling texture, layout, and context, Int. J. Comput. Vis. (2009), <https://doi.org/10.1007/s11263-007-0109-1>.
- [29] B.K. Spears, J. Brase, P.T. Bremer, B. Chen, J. Field, J. Gaffney, M. Kruse, S. Langer, K. Lewis, R. Nora, J.L. Peterson, J. Jayaraman Thiagarajan, B. Van Essen, K. Humbird, Deep learning: a guide for practitioners in the physical sciences, Phys. Plasmas (2018), <https://doi.org/10.1063/1.5020791>.
- [30] K. Aditya Mohan, S.V. Venkatakrishnan, J.W. Gibbs, E.B. Gulsoy, X. Xiao, M. De Graef, P.W. Voorhees, C.A. Bouman, TIMBIR: a method for time-space reconstruction from interlaced views, IEEE Trans. Comput. Imaging (2015), <https://doi.org/10.1109/TCI.2015.2431913>.
- [31] J.W. Gibbs, K.A. Mohan, E.B. Gulsoy, A.J. Shahani, X. Xiao, C.A. Bouman, M. De Graef, P.W. Voorhees, The three-dimensional morphology of growing dendrites, Sci. Rep. (2015), <https://doi.org/10.1038/srep11824>.
- [32] B. Karakaya, W.T. Thompson, The Pb-Sn (lead-tin) system equilibrium diagram, Bull. Alloy Phase Diagrams 9 (1988), <https://doi.org/10.1007/BF02890552>.
- [33] R. Mendoza, J. Alkemper, P.W. Voorhees, The morphological evolution of dendritic microstructures during coarsening, Metall. Mater. Trans. A. 34 (2003) 481–489, <https://doi.org/10.1007/s11661-003-0084-2>.
- [34] J. Alkemper, P.W. Voorhees, Quantitative serial sectioning analysis, J. Microsc. (2001), <https://doi.org/10.1046/j.1365-2818.2001.00832.x>.
- [35] S. Rashid, S.Y. Lee, M.K. Hasan, An improved method for the removal of ring artifacts in high resolution CT imaging, EURASIP J. Adv. Signal Process. (2012), <https://doi.org/10.1186/1687-6180-2012-93>.
- [36] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, ICLR 2015, 2015.
- [37] C.M. Bishop, Pattern Recognition and Machine Learning, Springer, 2006.