Dear Author,

Here are the proofs of your article.

- You can submit your corrections **online**, via **e-mail** or by **fax**.
- For **online** submission please insert your corrections in the online correction form. Always indicate the line number to which the correction refers.
- You can also insert your corrections in the proof PDF and **email** the annotated PDF.
- For fax submission, please ensure that your corrections are clearly legible. Use a fine black pen and write the correction in the margin, not too close to the edge of the page.
- Remember to note the **journal title**, **article number**, and **your name** when sending your response via e-mail or fax.
- **Check** the metadata sheet to make sure that the header information, especially author names and the corresponding affiliations are correctly shown.
- **Check** the questions that may have arisen during copy editing and insert your answers/ corrections.
- **Check** that the text is complete and that all figures, tables and their legends are included. Also check the accuracy of special characters, equations, and electronic supplementary material if applicable. If necessary refer to the *Edited manuscript*.
- The publication of inaccurate data such as dosages and units can have serious consequences. Please take particular care that all such details are correct.
- Please **do not** make changes that involve only matters of style. We have generally introduced forms that follow the journal's style.
  Substantial changes in content, e.g., new results, corrected values, title and authorship are not allowed without the approval of the responsible editor. In such a case, please contact the Editorial Office and return his/her consent together with the proof.
- If we do not receive your corrections **within 48 hours**, we will send you a reminder.
- Your article will be published **Online First** approximately one week after receipt of your corrected proofs. This is the **official first publication** citable with the DOI. **Further changes are, therefore, not possible.**
- The **printed version** will follow in a forthcoming issue.

Please note

After online publication, subscribers (personal/institutional) to this journal will have access to the complete article via the DOI using the URL: http://dx.doi.org/[DOI].
If you would like to know when your article has been published online, take advantage of our free alert service. For registration and further information go to: http://www.link.springer.com.

Due to the electronic nature of the procedure, the manuscript and the original figures will only be returned to you on special request. When you return your corrections, please inform us if you would like to have these documents returned.

Metadata of the article that will be visualized in OnlineFirst

ArticleTitle	Machine-Learning-based Algorithms for Automated Image Segmentation Techniques of Transmission X-ray Microscopy (TXM)	
Article Sub-Title		
Article CopyRight	The Minerals, Metals & Materials Society (This will be the copyright line in the final PDF)	
Journal Name	JOM	
Corresponding Author	Family Name	**Chawla**
	Particle	
	Given Name	**Nikhilesh**
	Suffix	
	Division	School of Materials Engineering
	Organization	Purdue University
	Address	West Lafayette, IN, 47907, USA
	Phone	
	Fax	
	Email	nikc@purdue.edu
	URL	
	ORCID	
Author	Family Name	**Torbati-Sarraf**
	Particle	
	Given Name	**Hamidreza**
	Suffix	
	Division	School of Materials Engineering
	Organization	Purdue University
	Address	West Lafayette, IN, 47907, USA
	Phone	
	Fax	
	Email	
	URL	
	ORCID	
Author	Family Name	**Niverty**
	Particle	
	Given Name	**Sridhar**
	Suffix	
	Division	School of Materials Engineering
	Organization	Purdue University
	Address	West Lafayette, IN, 47907, USA
	Phone	
	Fax	
	Email	
	URL	

	ORCID	
Author	Family Name	**Singh**
	Particle	
	Given Name	**Rajhans**
	Suffix	
	Division	School of Arts Media and Engineering
	Organization	Arizona State University
	Address	Tempe, AZ, 85281, USA
	Phone	
	Fax	
	Email	
	URL	
	ORCID	
Author	Family Name	**Barboza**
	Particle	
	Given Name	**Daniel**
	Suffix	
	Division	School of Arts Media and Engineering
	Organization	Arizona State University
	Address	Tempe, AZ, 85281, USA
	Phone	
	Fax	
	Email	
	URL	
	ORCID	
Author	Family Name	**Andrade**
	Particle	**De**
	Given Name	**Vincent**
	Suffix	
	Division	Advanced Photon Source
	Organization	Argonne National Laboratory
	Address	Argonne, IL, 60439, USA
	Phone	
	Fax	
	Email	
	URL	
	ORCID	
Author	Family Name	**Turaga**
	Particle	
	Given Name	**Pavan**
	Suffix	
	Division	School of Arts Media and Engineering
	Organization	Arizona State University
	Address	Tempe, AZ, 85281, USA
	Phone	

Fax
Email
URL
ORCID

Schedule	Received	15 January 2021
	Revised	
	Accepted	22 April 2021

Abstract

Four state-of-the-art Deep Learning-based Convolutional Neural Networks (DCNN) were applied to automate the semantic segmentation of a 3D Transmission x-ray Microscopy (TXM) nanotomography image data. The standard U-Net architecture as baseline along with UNet++, PSPNet, and DeepLab v3+ networks were trained to segment the microstructural features of an AA7075 micropillar. A workflow was established to evaluate and compare the DCNN prediction dataset with the manually segmented features using the Intersection of Union (IoU) scores, time of training, confusion matrix, and visual assessment. Comparing all model segmentation accuracy metrics, it was found that using pre-trained models as a backbone along with appropriate training encoder–decoder architecture of the Unet++ can robustly handle large volumes of x-ray radiographic images in a reasonable amount of time. This opens a new window for handling accurate and efficient image segmentation of in situ time-dependent 4D x-ray microscopy experimental datasets.

Footnote Information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11837-021-04706-x.

JOM
https://doi.org/10.1007/s11837-021-04706-x

MICROSTRUCTURE CHARACTERIZATION: DESCRIPTORS, DATA-INTENSIVE TECHNIQUES, AND UNCERTAINTY QUANTIFICATION

Machine-Learning-based Algorithms for Automated Image Segmentation Techniques of Transmission X-ray Microscopy (TXM)

HAMIDREZA TORBATI-SARRAF,[1] SRIDHAR NIVERTY,[1] RAJHANS SINGH,[2] DANIEL BARBOZA,[2] VINCENT DE ANDRADE,[3] PAVAN TURAGA,[2] and NIKHILESH CHAWLA[1,4]

1.—School of Materials Engineering, Purdue University, West Lafayette, IN 47907, USA. 2.—School of Arts Media and Engineering, Arizona State University, Tempe, AZ 85281, USA. 3.—Advanced Photon Source, Argonne National Laboratory, Argonne, IL 60439, USA. 4.—e-mail: nikc@purdue.edu

Four state-of-the-art Deep Learning-based Convolutional Neural Networks (DCNN) were applied to automate the semantic segmentation of a 3D Transmission x-ray Microscopy (TXM) nanotomography image data. The standard U-Net architecture as baseline along with UNet++, PSPNet, and DeepLab v3+ networks were trained to segment the microstructural features of an AA7075 micropillar. A workflow was established to evaluate and compare the DCNN prediction dataset with the manually segmented features using the Intersection of Union (IoU) scores, time of training, confusion matrix, and visual assessment. Comparing all model segmentation accuracy metrics, it was found that using pre-trained models as a backbone along with appropriate training encoder–decoder architecture of the Unet++ can robustly handle large volumes of x-ray radiographic images in a reasonable amount of time. This opens a new window for handling accurate and efficient image segmentation of in situ time-dependent 4D x-ray microscopy experimental datasets.

INTRODUCTION

X-ray microtomography has become a very important characterization technique for understanding materials behavior. Depending on the particular modality, this technique can enable image resolutions ranging from tens of micrometers to the nanometer scale. More importantly, the non-destructive nature of this technique allows us to conduct in situ and/or time-resolved (4D) investigations where the evolution of microstructure[1–6] or propagation of a defect[7–19] can be captured as a function of time. In this study, the x-ray tomography technique of interest is transmission x-ray microscopy (TXM) which can be used to conduct in situ 4D experiments at high spatial resolutions ($\sim$ 20 nm)[20–27] using the 32-ID beamline at the Advanced Photon Source (APS).

(Received January 15, 2021; accepted April 22, 2021)

One of the challenges with 4D x-ray microtomography experiments is the large amounts of data that are generated (often in the TB range). In particular, 3D rendering and statistical quantification of microstructurally evolving features (for example, crack growth, corrosion propagation, and phase transformation) require image segmentation. Thus, the subsequent image processing and feature classification is often the rate-limiting step for tomographic data analysis. Pixels with a wide distribution of grayscale values in the reconstructed images need to be segmented based on the features' density and homogeneity. These features often cannot be segmented using simple histogram thresholding or edge-based filtering. The presence of beam-hardening, scattered or other ionizing x-ray -generated "zinger" artifacts, ring artefacts, edge blurring due to motion artefacts, and phases with similar attenuation make the segmentation process more complicated.[28,29] This renders the

Journal : 11837_JOM Dispatch : 4-5-2021 Pages : 12
Article No.: 4706 ☐ LE ☐ TYPESET ☑ CP ☑ DISK

requirement of human hands-on intervention for careful manual segmentation, making the process extremely cumbersome and time-intensive. Manual image segmentation is also subjective, depending on the experience and visual acuity of the person doing the analysis.[30] Thus, there is a need for a new set of tools to move toward automating image segmentation of large 4D x-ray tomography datasets.

With the success of machine-learning algorithms in general image analysis, numerous promising segmentation and classification approaches can be applied for image segmentation for materials science applications. Traditional machine-learning image-processing methods, such as k-means clustering and thresholding, have been employed for various segmentation tasks.[31,32] Contemporary approaches, such as Gabor filtering, rely on neural networks to solve a given end-task, such as classification or segmentation, by learning the model parameters from training datasets[33]. Nevertheless, these methods still require minimal human intervention and are known as semi-automatic segmentation approaches.

In recent developments, deep-learning (DL) has pushed the boundaries in improving the robustness of segmentation methods with minimal multi-pass post-segmentation human intervention. Deep convolutional neural networks (DCNNs), have achieved excellent performance for various image-processing tasks. Krizhevsky et al.[34] developed the first large-scale application of DCNNs on difficult natural image classification problems, and, since then, different deep architectures have been modified in different domains to improve the time and accuracy of automated image processing. Although DL-based image segmentation techniques have been extensively implemented in medical x-ray radiographic images, only a handful of studies have applied this approach for x-ray microscopy images in materials science and engineering.[5,23,35–41]

In this work, we study the performance of four state-of-the-art deep learning architectures for automatic image segmentation of a TXM dataset. We chose U-Net, UNet++, PSPNet, and DeepLab v3+ DCNN architectures, as the implementation of these networks are readily accessible and have shown outstanding segmentation results in multiple domains, such as intelligent transportation, geo sensing, and medical imaging.[42–46] The architecture of all the adopted networks contains an encoder and a decoder sub-network. First, the encoder extracts the features from a given image by contracting the image into different depths (resolutions) using different down-sampling convolutions and operational layers. Then, the decoder takes the feature map from different depths of the encoder, predicts the class of the pixel, recovers spatial information, and reconstructs the image using up-sampling convolutions.[47] However, each model uses different strategies, operational layers, and convolutional arithmetic to extract and predict the features. The U-Net was chosen as the baseline, which comprises a symmetric encoder–decoder architecture and extracts features by applying consecutive convolutions into a fixed depth. UNet++ applies similar strategies to U-Net, but its architecture has been redesigned to operate at the optimal depth.[43,48] In contrast to U-Net and UNet++, PSPNet and DeepLab v3+ simultaneously apply multi-scale convolutional modules to convert the image into different depths, and the decoder fuses all the features at different scales to the prediction output. Applying this strategy can potentially accelerate the speed of image processing by extracting global information in the image more efficiently.[46,49–51]

Although the application of DL tools for x-ray microscopy-based imaging is still in its early stages, here, we provide a unified framework for the analysis of x-ray tomography datasets using DCNN. To obtain an optimal segmentation output, optimized hyperparameters, methods, and backbones were identified for each architecture (a general description of DCNN parameters and terms are provided in the electronic supplementary). A supervised DL approach using stacks of 2D TXM grayscale slices (the raw data) and the corresponding manually segmented RGB images (ground-truth) were used as a training dataset. Then the adopted networks were trained, and the extracted models were used to predict the rest of the dataset images. A comparison of the ground-truth versus the predicted datasets was quantified and showed excellent agreement. In addition, the computational efficiency, based on the time taken to process an image, was also found to be favorable.

EXPERIMENTAL METHODS AND PROCEDURE

In this section, we describe the experimental details for acquiring the XCT dataset, the image segmentation, hardware details, and the analysis used for comparing the accuracy of the predicted segmented images.

X-ray Synchrotron Tomography

This study was conducted on 2D x-ray slices obtained from a TXM scan performed on 7075 aluminum alloy (AA7075). The principal constituent particles of AA7075 are the Al-Cu-Fe inclusions and the Cu-Mg-Zn precipitates.[52,53] The size, geometry, and distribution of these phases play a significant role in determining the mechanical behavior and corrosion performance of this alloy. [13,54,55] Rods of the AA7075 were overaged using the following heat-treatment protocol: solution treatment at 510°C for 2 h → water-quenching → overageing at 107°C for 6 h and at 163°C for 40 h, followed by further coarsening of the precipitates at 300°C for 86 h. This resulted in a significant coarsening of the precipitate particles. Al-Cu-Fe inclusions, with the composition Al_7Cu_2Fe, are intermetallic particles present

Journal : 11837_JOM | Dispatch : 4-5-2021 | Pages : 12
Article No.: 4706 | ☐ LE | ☐ TYPESET | ☑ CP | ☑ DISK

in the form of stringers along the rolling direction. A micropillar of overaged AA7075, approximately 15 μm in diameter and 30 μm in height, was milled using a Ga^{+} focused ion beam. A TXM scan was conducted on the pillar at the 32-ID-C beamline at the APS at Argonne National Laboratory. The combination of a condenser lens and custom-made Fresnel zone plates yielded a voxel size of 18 nm. Details about the beamline can be found in previous studies[26,27]. A monochromatic x-ray beam energy of 9.75 keV (just above the zinc absorption edge) was used for the scan to obtain the maximum contrast between the Cu-Mg-Zn precipitate particles, the Al_7Cu_2Fe inclusions, and the matrix. A total of 1200 projections were captured over an angular range of 0–180° with an exposure time of 1 s/projection. A filtered back-projection algorithm was used to reconstruct the projections using the software TomoPy[27]. The reconstructed dataset was converted into a 2D slice stack of 730 images, having an image size of 768 × 768 pixels with a high bit-depth of 32 bits.

Manual Segmentation Recipe

Due to different attenuating properties originating from different atomic weight densities of the particles, these phases can be distinguished based on their gray value range. The presence of composite phases and the formation of near-field phase contrast fringes around the periphery of the particles (shown in Fig. 1), makes the manual segmentation complicated. To this end, a non-local mean filter was employed to reduce noise, and an unsharp mask filter was used to sharpen the relevant microstructural features. Avizo 9 (Bethesda, MD, USA) was then used to perform manual image segmentation to generate the ground-truth.

Deep Convolutional Neural Network (DCNN) Architectures

U-Net

The U-Net is a symmetric U-shaped encoder–decoder network originally developed for medical image processing[56]. Its general architecture shown in Fig. 2(a). First, in the contraction path (encoder), the image features are extracted using consecutive 3 × 3 convolutions followed by 2 × 2 rectified linear unit (ReLU) activation and 2 × 2 max-pooling operations. Then, in the expansion path (decoder), the dense output of the encoder is progressively expanded. In each step of the expansion path, the spatial information of up-convolution is concatenated with the corresponding feature maps from the contraction pathway, followed by 3 × 3 convolutions and ReLU layers. However, this design has a major limitation. Depending on the feature sizes and the numbers of labeled classes for training, the optimal depth of an encoder–decoder network can vary for different segmentation tasks. Hence, this network cannot be used for multi-scale feature segmentation as it is unnecessarily restricted to fuse feature maps into the fixed depth[43].

UNet++

The UNet++ is a redesigned U-Net architecture which extends the U-Net's abilities for achieving multi-scale and more accurate semantic segmentation.[43,48] As compared to the U-Net (Fig. 2(b)), the UNet++ consists of varying depths, and the decoders are densely connected at the same resolution of encoders via skip connections. It bridges the feature maps from different depths of the contraction path to the expansion path before merging them. This architectural modification not only improves the overall segmentation accuracy but also enhances the learning and prediction time by enabling the network pruning itself.[43] Furthermore, using deep supervision enables the model to operate in different modes by averaging all segmentation branches. In the current study, we also used pre-trained ResNet 152 as a backbone to enhance the training speed.

PSPNet

The pyramid scene parsing network (PSPNet) uses the spatial pyramid pooling module with different-region-based contexts to achieve superior segmentation performance.[43,48,51,57] As shown in Fig. 2(c), the PSPNet architecture takes the feature map from the last convolutional layer as an input image and fuses the features under four different pyramid scales. The pyramid levels form pooled representations of the feature map. The low-dimension feature maps are then up-sampled to the input image size and concatenated with the original input image.[51,58] Using multi-scale pyramid pooling, convolution aids the network to extract global features in the image more efficiently. A graphical interpretation of spatial pyramid pooling can be found in supplementary Fig. S-2. In this study, a pre-trained ResNet 152 backbone was used to extract the feature map as an input to the pyramid pooling module.

DeepLab v3+

This network applies atrous convolutions and atrous spatial pyramid pooling (ASPP) approach to extract the feature in its encoder sub-network. Atrous convolution are also called "dilated convolutions". A graphical interpretation of ASPP is depicted in supplementary Fig. S-3. Using this module, DeepLab v3+ is able to extract global and multi-scale features of the image simultaneously, which results in a faster computational process compared to conventional convolutions used in U-Net base architectures.[49,50] As shown in Fig. 2(d), the extracted features from atrous convolutions with different sampling rates and strides are fused

Journal : 11837_JOM | Dispatch : 4-5-2021 | Pages : 12
Article No.: 4706 | ☐ LE ☑ CP | ☐ TYPESET ☑ DISK

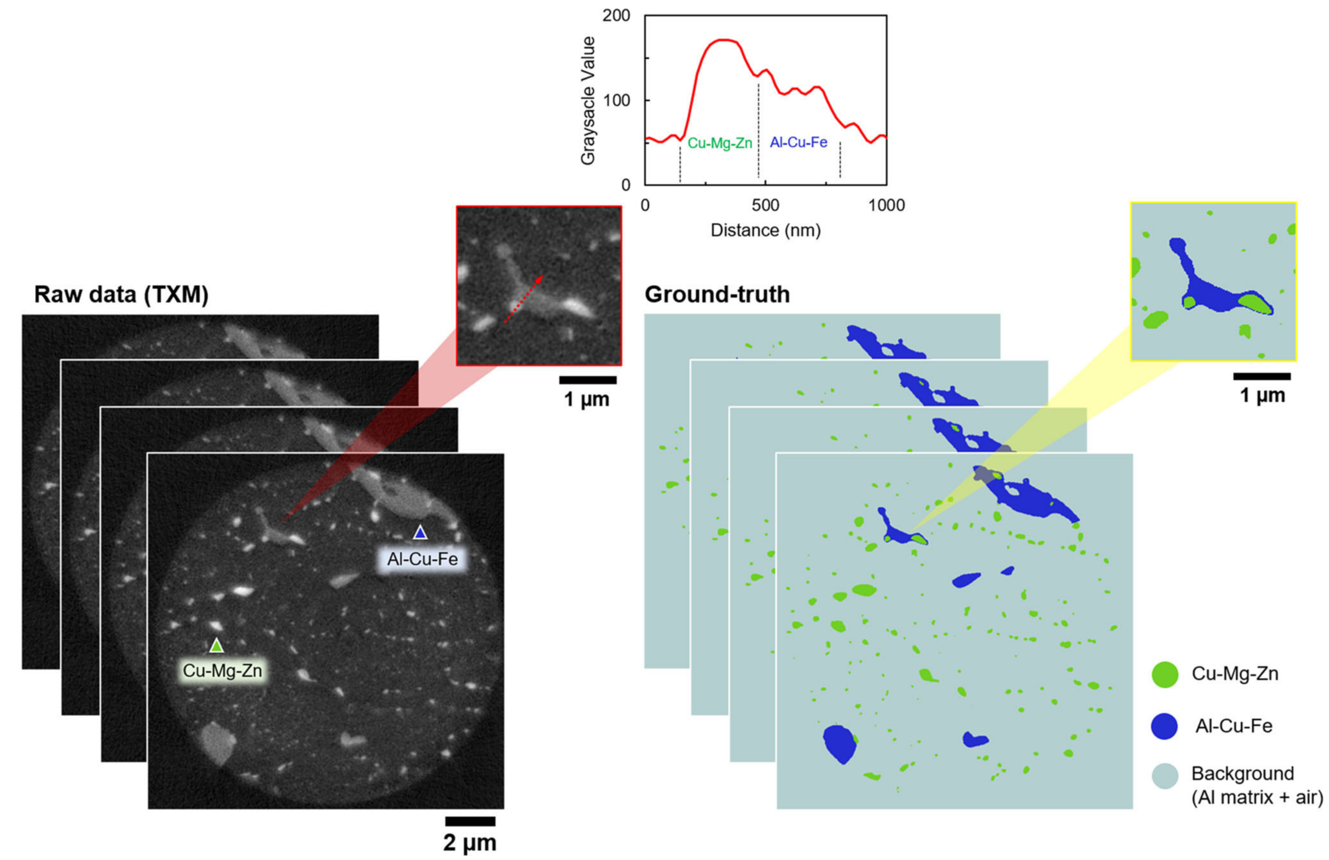

Fig. 1. Typical raw and ground-truth (manually segmented) images from a 15-μm-diameter micropillar of an overaged AA 7075 dataset obtained by the TXM technique.

to generate the encoder part of the network. Then, the decoder uses the fused map from the ASPP and the low-level features received from the preliminary extracted feature by DCNN layers as input and generates the output predicted output[49]. In this study, this technique was paired with an Xception 71 model as a backbone.

Training, Testing, and Evaluation of CNN Algorithms

After normalizing of the entire batch of TXM dataset (a description of batch normalization can be found in the supplementary file: S-1), the consecutive image stack (here, each slice is a pair of raw and ground-truth images (shown in Fig. 1 as an example) was randomly split into training and testing batches based on their slice numbers. 550 images (75% of the whole data) were used for training and 180 images (25% of the whole data) were used for testing. During training, the ground-truth dataset was used as the target output to minimize the loss error. For testing, the ground-truth images were used to calculate the accuracy of the prediction. To validate and monitor training progress after each epoch, 20 images were randomly separated from the training batch for the validation loss checkpoint. All models were trained and tested with the use of GPU (NVIDIA Quadro RTX 6000) computational resources.

We use the intersection over union (IoU) as a metric to quantify the accuracy of the segmentation models[36]. Given the ground-truth mask and predicted image, the IoU for a class m can be computed using Eq. 1:

$$\mathrm{IoU_m} = \frac{t_{\mathrm{pm}}}{t_{\mathrm{pm}} + f_{\mathrm{pm}} + f_{\mathrm{nm}}} \tag{1}$$

where t_{pm} and f_{pm} are the numbers of true and false positives, respectively, and f_{nm} refers to the number of false-negative pixels in the predicted image. This metric measures the ratio of the area of overlap between the predicted feature and the ground-truth, divided by the area of union between the predicted feature and the ground-truth, as shown in Fig. 3. The mean of IoU (mIoU) scores for every class present in the predicted image were used to compare different models.

RESULTS AND DISCUSSION

The baseline U-Net architecture used for the initial experimentation was implemented in a deep-learning toolbox for x-ray imaging obtained from the "Xlearn" Github repository (github.com/tomography/xlearn). A detailed description of the Xlearn network is provided in[40]. To our knowledge, for the first time, Xlearn, as a CNN segmentation tool, has

Journal : 11837_JOM Dispatch : 4-5-2021 Pages : 12
Article No.: 4706 ☐ LE ☑ CP ☐ TYPESET ☑ DISK

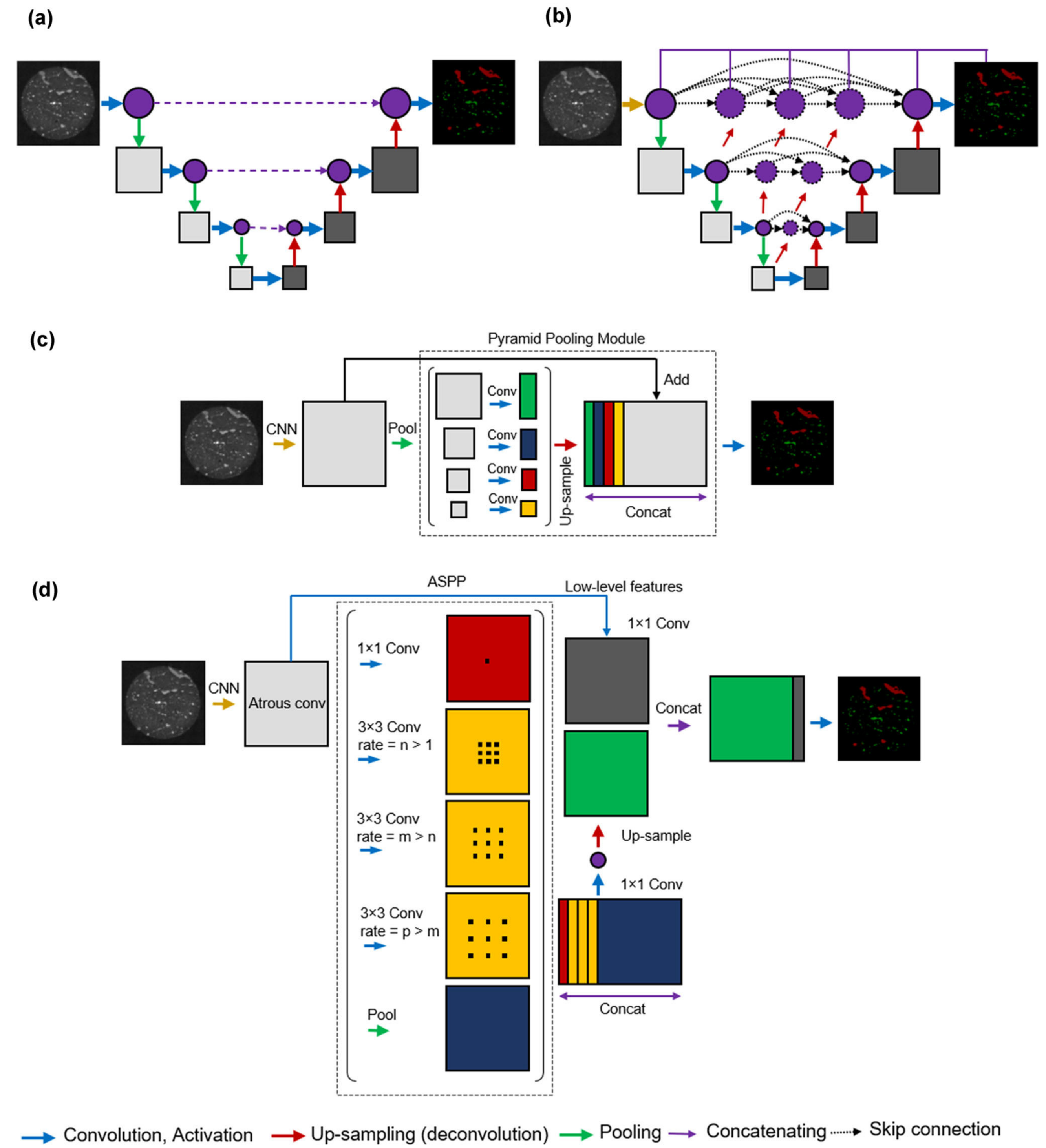

Fig. 2. The architecture of the four networks used for this study: (a) U-Net, (b) UNet++, (c) PSPNet, and (d) DeepLab v3+.

been implemented on a TXM dataset to segment 2D images and visualize the microstructural features in an Al-4%Cu alloy micropillar. This alloy mainly consists of plate-like and needle-like Al_2Cu precipitates in an α-Al matrix and the shape, size, and distribution of each phase is required to understand the mechanical behavior of this material. [59]However, marginal differences in grayscale values between the existing precipitates in this alloy made it almost impossible to conduct manual segmentation on the entire TXM dataset. This entailed the application of an efficient automated technique to fully segment the whole dataset. To this end, a sub-volume (only 1/32 of the whole scan) of the TXM slices were segmented manually within 36 h. The Xlearn algorithm was trained to emulate

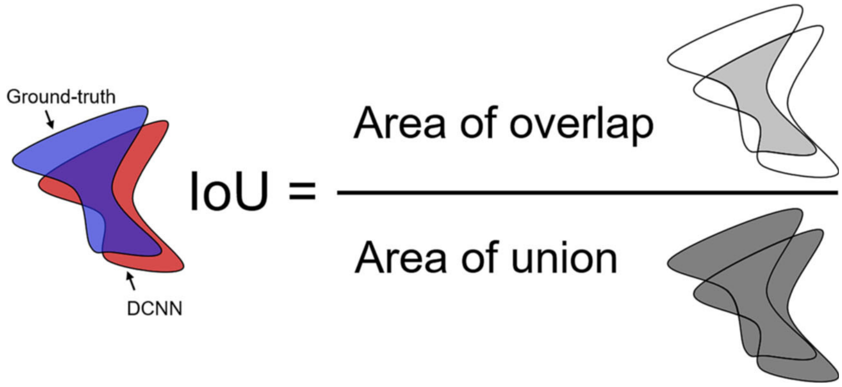

Fig. 3. Graphical interpretation of the intersection over union (IoU) metric used to compare the accuracy of the predicted feature (DCNN) with the manual segmented feature (ground-truth).

Table I. U-Net results on the testing dataset for binary (2-class) and 3-class segmentation

Segmentation	Loss function	Epoch	mIoU (%)
2-class (binary)	Means square error)	10	95.9
3-class	Categorial cross-entropy	10	91.7

segmenting of the precipitates of the alloy using the 2D images provided from the manually segmented sub-volume. Using a CPU, this process only took 2 h. Then, the entire dataset was segmented automatically using the trained Xlearn model within just 20 h using a CPU. The quantitative volume comparison of the 3D-rendered data from the manually and Xlearn segmented dataset revealed appreciable accuracy of the DCNN approach for such TXM segmentation[23]. The time for automated segmentation is also a function of the hardware computation capability.

In our current study, we executed the Xlearn and other DCNN algorithms training and testing process on a GPU cluster and compared the computational time reduction. The Xlearn was trained on all 550 training images with a patch size of 768 × 768 pixels, using the Adam optimizer with the MSE loss (binary segmentation) as default loss function. However, we changed the loss function to categorial cross-entropy be able to classify and segment the 3 classes. For the binary (2-class) segmentation, both the Cu-Mg-Zn precipitates and Al-Cu-Fe inclusions were classified as a single class with the background being another class; while for 3-class segmentation, each precipitate, inclusions, and background were classified as separate classes. As shown in Table I, the U-Net gave better mIoU scores (mIoU overall all classes) in the case of binary segmentation compared with 3-class segmentation. However, to segment Al-Cu-Fe and Cu-Mg-Zn particles separately, multi-class segmentation was employed for all four architectures under consideration.

The UNet++ and PSPNet were implemented on a nested U-Net architecture from the GitHub repository (github.com/MrGiovanni/UNetPlusPlus). The DeepLab v3+ architecture with the exception 71 as backbone were referred to existing open-source TensorFlow Model Garden implementations (github.com/tensorflow/models). The training and testing dataset were converted to Tensorflow's '.tfrecord' format. Used this way, the models can become trained faster, as less memory is consumed during processing, while the data can be read quickly from memory. Also, the atrous rates were set to 3, 6, and 9 for training and testing.

In our code implementation, based on the initial epochs and monitoring the loss convergence and accuracy progression of the validation batch, the best sets of hyperparameters were chosen before initiating the main training process. These parameters included learning-rate, dropout, batch size, loss function, and optimizer (the detailed description of each term is provided in supplementary file: S-1). A fixed batch size was used for all the models. Various loss functions provided in[60,61] were tested, and the summation of the categorial cross-entropy and dice loss (1 − dice similarity coefficient) as the loss function gave the best training and testing

Journal : **11837_JOM** Dispatch : **4-5-2021** Pages : **12**
Article No.: **4706**
☐ LE ☑ CP ☐ TYPESET ☑ DISK

Table II. Performance analysis of the implemented architectures

Model	Batch size	Epoch	Loss function	Training time (h)	Prediction time per image (s)	mIoU score (%)
U-Net	1	10	Categorical cross-entropy + dice loss	12	1.9	92.2
UNet++	1	10	Categorical cross-entropy + dice loss	8	1.1	95.1
PSPNet	1	10	Categorical cross-entropy + dice loss	5	0.9	88.2
DeepLab v3+	1	10	Categorical cross-entropy + dice loss	4	0.8	89.1

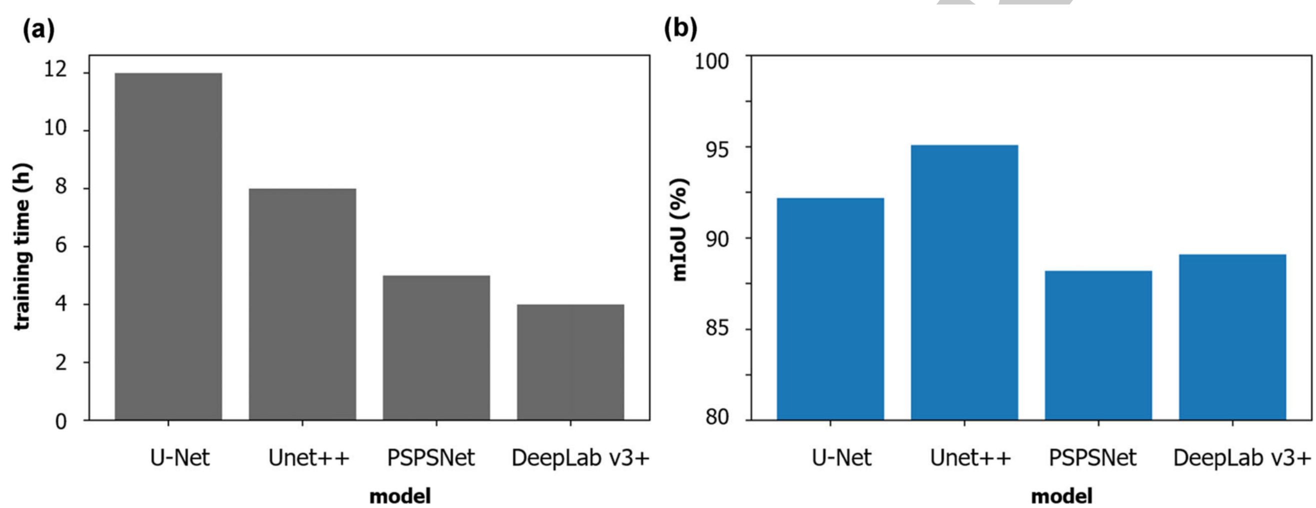

Fig. 4. Performance of the implemented architectures: (**a**) time of the training, (**b**) accuracy.

results for all the algorithms. In all the training processes, the learning rate was set to 10^{-4} and the dropout was 0.5.

Numbers of epochs, training time, prediction time, and mIoU scores obtained from all the models are listed in Table II, and the preferences are plotted in Fig. 4. Typical predicted segmentations by trained models are presented in Fig. 5. The U-Net implementation took the longest amount of time owing to the process of fine-tuning and presumably down-sampling the input image to a fixed depth. Note that a small increment was observed in the U-Net mIoU score by applying a summation of the categorial cross-entropy with dice loss (Table II) instead of merely categorial cross-entropy loss (Table I) as the loss function. This improvement is attributed to the dice loss function, as it not only evaluates the number of pixels correctly labelled but also penalizes instances of incorrect segmentation (false-positive and false-negative) and determines the accuracy of the segmentation boundaries[42].

The UNet++ training and testing process was faster than U-Net as it used the pre-trained model as backbone. Also, the architecture of UNet++ takes advantage of skip connections to operate at an optimal depth. Considering all of these strategies, UNet++ achieved significantly higher mIoU performance compared to all the other architectures. PSPNet was able to achieve favorable mIoU scores in a training time of just 17 h. The faster performance of PSPNet compared to UNet++ can be attributed to the application of pyramid pooling convolutions. However, comparing all the training times, the DeepLab v3+ outperformed the other architectures, but it has a slightly lower segmentation accuracy than the best one. As DeepLab v3+ takes advantage of atrous pooling convolutions with different rates, the kernel can move faster across the input feature map and extract global information more efficiently than the other encoder sub-networks used in other models (an interpretation of atrous pooling operation can be found in Figure S-3 in supplementary file).

In this study, the background class occupies the major portion of the x-ray micrographs compared to other constituent particles. Hence, the high values of IoU score might be coming from background labels. To highlight the accuracy of segmentation in

Journal : 11837_JOM | Dispatch : 4-5-2021 | Pages : 12
Article No.: 4706 | ☐ LE ☑ CP | ☐ TYPESET ☑ DISK

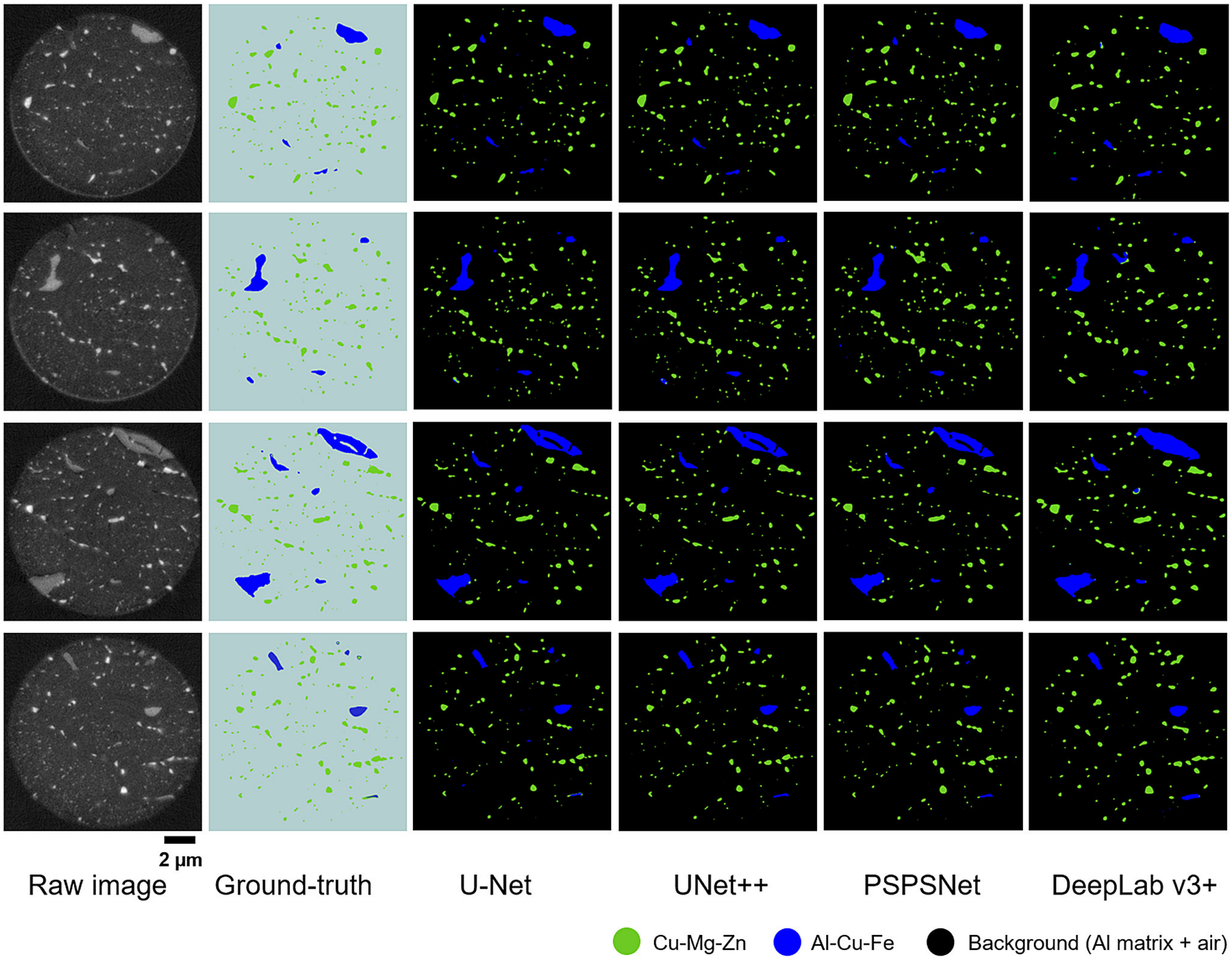

Fig. 5. Comparison of raw, ground-truth, and segmented output of images predicted by different DL architectures.

precipitates and inclusions, we calculated the confusion matrix to evaluate the pixel-level classification accuracy for each label (The description and graphical interpretation of the confusion matrix can be found in the supplementary document: S-4, 5). Figure 6 shows the confusion matrixes extracted from pixel-by-pixel comparison of ground-truth and predicted images used for testing the networks. A detailed look at the numbers clearly reveals that the background has the highest true-positive values in all the architectures, compared to other labels which led to high mIoU scores. The significant errors in all the architectures can be attributed to the misclassification of Cu-Mg-Zn and Al-Cu-Fe labels to background. These errors are more pronounced for PSPNet and DeepLab v3+, as they use larger kernels for convolution operation which can potentially dilate the boundaries or small features' gray value into the background in the maxpooling layer. However, similar to what was observed in the mIoU score in Table II, the Unet++ followed by U-Net outperformed the other models in term of true-positive values. In contrast, a great portion of the Cu-Mg-Zn precipitates and Al-Cu-Fe inclusions has been segmented as background by the PSPNet model.

In the following, intuitive examples of segmentation and classification are presented. Figure 7 shows an example of the composite particle segmentation predicted by all models. Surrounding and embedding a particle into another phase makes the x-ray image segmentation more difficult, as the phase boundaries are barely distinguishable. As indicated in the outlined boxes in Fig. 5, the U-Net and UNet++ outperformed other networks to segment Al-Cu-Fe rims around the Cu-Mg-Zn precipitate, presumably due to the appropriate depth of the convolutions. However, it appears that employing pyramid pooling and atrous pooling convolutions by PSPNet and DeepLab v3+, respectively, dilates the boundaries of the classes and overlooks the details of the feature's periphery. This inaccurate segmentation shows the limitation of the dilated pooling strategy used by PSPNet and DeepLab v3+, where the kernel matrix size is larger than the area of the particle.

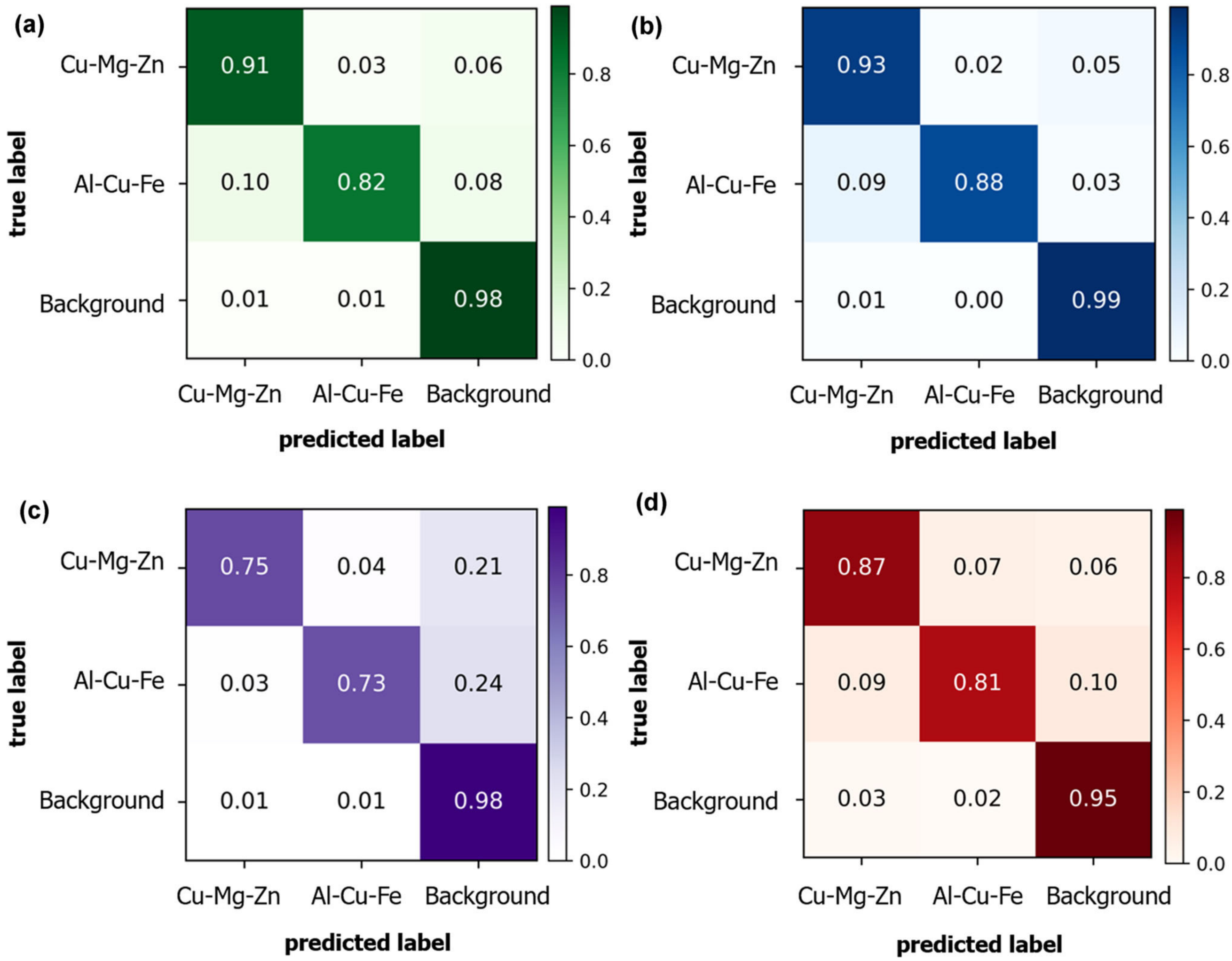

Fig. 6. Normalized confusion matrixes of segmentation results from (a) U-Net, (b) UNet++, (c) PSPNet, and (d) DeepLab v3+.

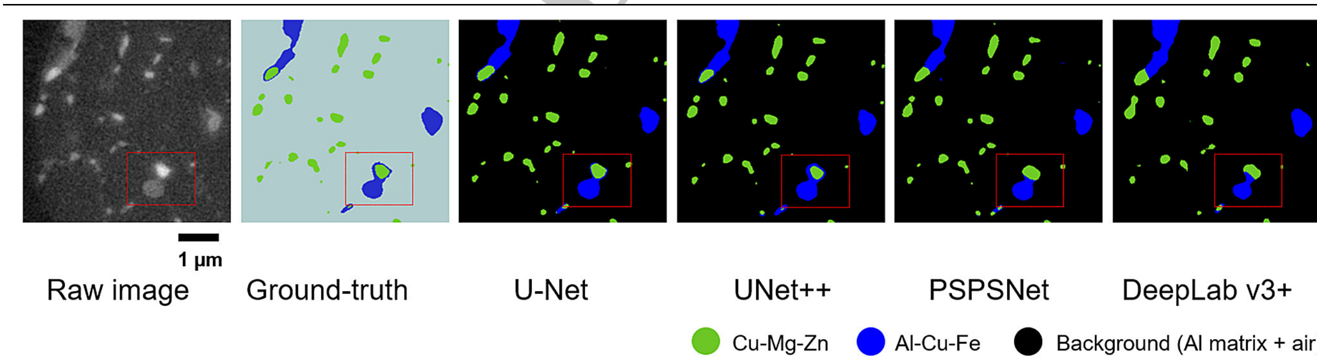

Fig. 7. Example of composite particle segmentation outputs by different DL architectures.

Figure 8 shows some examples of incorrect segmentation/classification outputs. All the architectures misclassified the features where the classes (background and particles) seemed to share similar gray values. Formation of near-field phase contrast fringes around the periphery of the particles or sample edges can potentially form brighter spots in some regions OF THE x-ray micrographs. For a few output images (10% of the Prediction dataset), U-Net and Unet++ classified some regions of the micropillar edges (background class) as Al-Cu-Fe inclusions. In addition, in numerous cases, Al-Cu-Fe inclusions were partially classified as Cu-Mg-Zn precipitates.

Journal : **11837_JOM** Dispatch : **4-5-2021** Pages : **12**
Article No.: **4706** ☐ LE ☑ CP ☐ TYPESET ☑ DISK

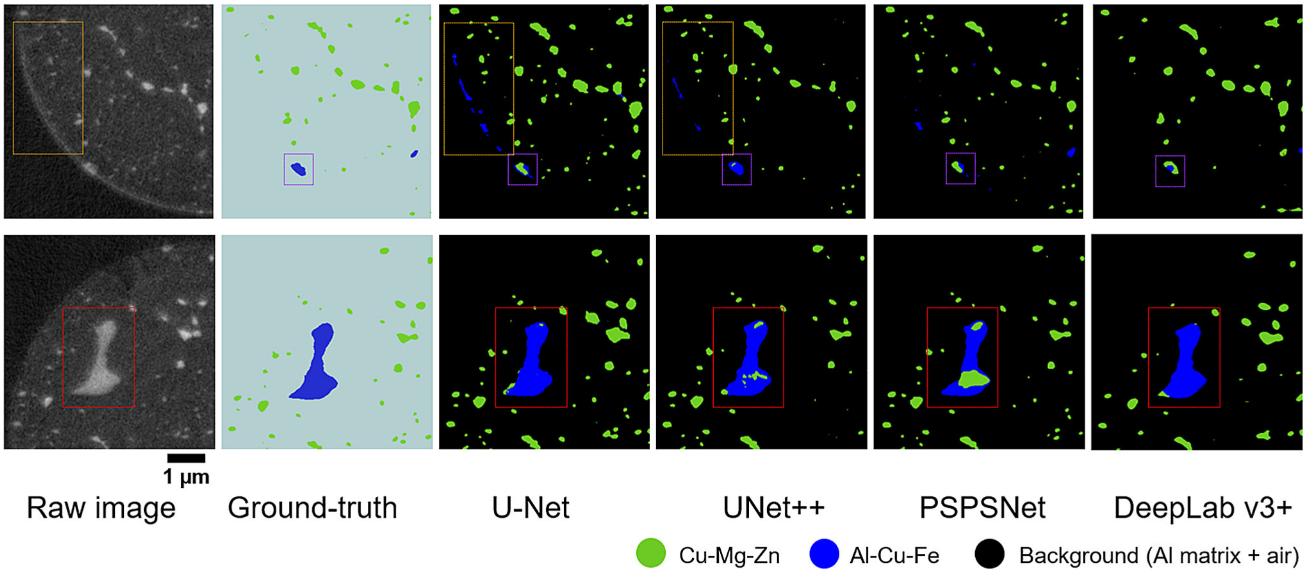

Fig. 8. Examples of incorrect misclassified feature outputs by different DL architectures.

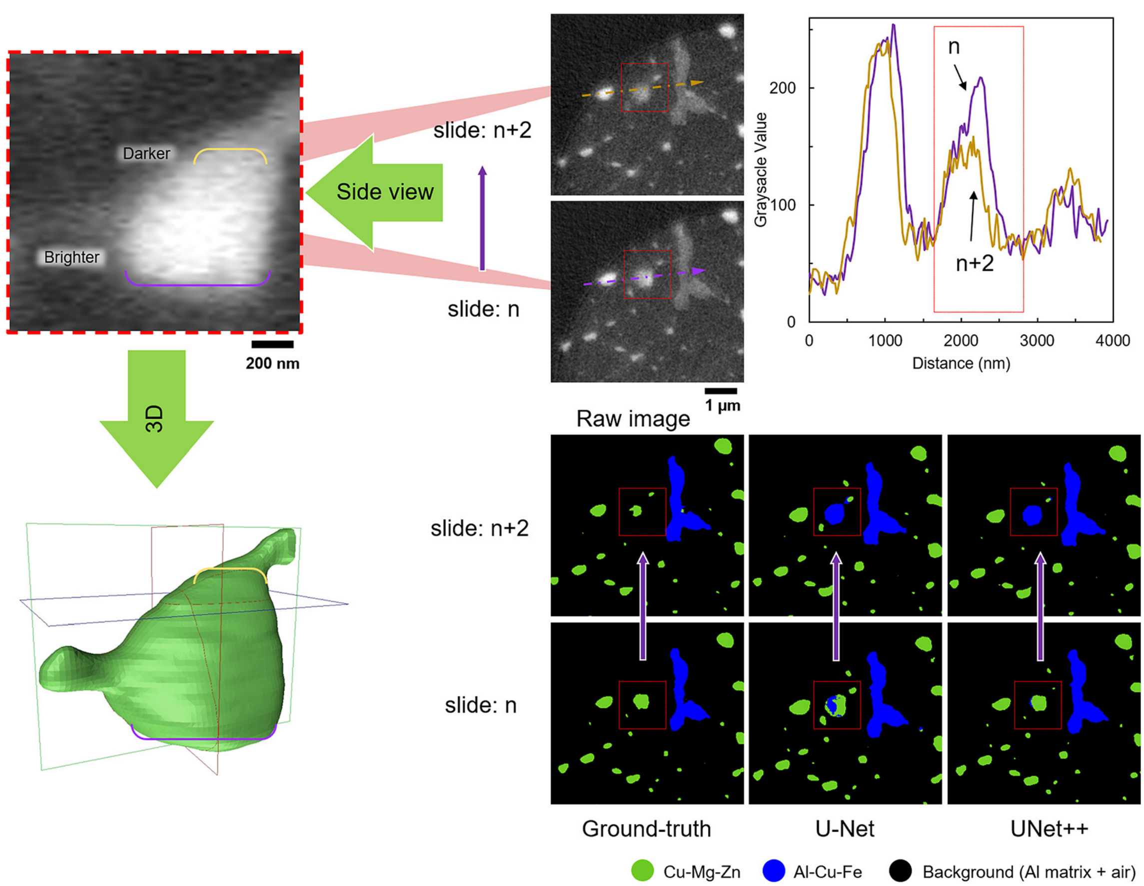

Fig. 9. Examples of inconsistent segmentation outputs predicted by U-Net and UNet++ architectures.

Inconsistent classifications were also observed for a few cases predicted by U-Net and UNet++. Figure 9 shows an example misclassification output on a particular precipitate in different slices. Both models were able to outline the Cu-Mg-Zn precipitate as a feature, but this particle was misclassified as Al-Cu-Fe inclusions in the neighboring slice. As shown in the grayscale value plot and 3D-rendered volume of the outlined Cu-Mg-Zn precipitate, this misclassification is primarily due to

Journal : 11837_JOM | Dispatch : 4-5-2021 | Pages : 12
Article No.: 4706 | ☐ LE | ☐ TYPESET
☑ CP | ☑ DISK

gray value changes within the thickness of this particle. The thinner region of the particle appears to have decreased attenuation compared with center of the particle. This leads the Cu-Mg-Zn precipitate to appear darker and look like an Al-Cu-Fe inclusion in its thinner cross-section in the nearby slice.

CONCLUSION

This study focused on the effectiveness of utilizing four start-of-the art deep learning architectures to perform automated segmentation on a complex nanotomography dataset obtained by TXM. A workflow was introduced to train, apply, and compare the models. All four architectures were successfully implemented and shown to perform well on an x-ray tomography (XRT) dataset.

The U-Net as the baseline and most common model in x-ray microscopy imageprocessing methods showed a modest performance and the slowest training time as compared with other models. By redesigning U-Net and applying skip connections as in UNet++, a significantly improved performance was achieved.

Furthermore, it was shown that backbones from open source libraries such as imagenet could also be used for XRT image-processing tasks. Promising performances with superior training times were achieved by application of pyramid and ASPP convolutions. However, there is room for further improvement in the configurations and implementations of the PSPNet and DeepLab v3+ models using other libraries, backbones, and shape-aware loss functions.

In addition to metrics such as mIoU, extracting the confusion matrix and visual assessment of the output help to interpret the strength of different CNN architectures for multi-class semantic segmentation, especially when the sizes of the labels (pixel proportion of the classes) are not balanced. This approach guides a practitioner to select an optimized architecture and parameters for automated segmentation. It has to be noted that the images used in this study was not pre-processed. In future, it would be ideal to apply various 2D and 3D filters, and also to implement data augmentation techniques to reduce misclassification, inconsistency, and incorrect segmentations.

ACKNOWLEDGEMENTS

H.T., S.N., and N.C. are grateful for financial support from the Office of Naval Research (ONR) under Contract No. N00014-10-1-0350 (Dr. W. Mullins, Program Manager). We acknowledge the use of resources at Beamline 32-ID-C of the Advanced Photon Source, a U.S. Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under Contract No. DE-AC02-06CH11357.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

SUPPLEMENTARY INFORMATION

The online version contains supplementary material available at https://doi.org/10.1007/s11837-021-04706-x.

REFERENCES

1. Q. Zhang, S. Niverty, A.S.S. Singaravelu, J.J. Williams, E. Guo, T. Jing, and N. Chawla, *Mater. Charact.* 148, 52. (2019).
2. J.M. Yu, N. Wanderka, A. Rack, R. Daudin, E. Boller, H. Markötter, A. Manzoni, F. Vogel, T. Arlt, I. Manke, and J. Banhart, *Acta Mater.* 129, 194. (2017).
3. Q. Krol, and H. Löwe, *Acta Mater.* 151, 478. (2018).
4. N. Limodin, L. Salvo, E. Boller, M. Suéry, M. Felberbaum, S. Gailliègue, and K. Madi, *Acta Mater.* 57, 2300. (2009).
5. C.S. Kaira, V. De Andrade, S. Singh, C. Kantzos, A. Kirubanandham, F. De Carlo, and N. Chawla, *Adv. Mater.* 29, 1703482. (2017).
6. E. Boulard, C. Denoual, A. Dewaele, A. King, Y. Le Godec, and N. Guignot, *Acta Mater.* 192, 30. (2020).
7. S. Niverty, C. Kale, K.N. Solanki, and N. Chawla, *Corros. Sci.* 185, 109429. (2021).
8. M.B. Kelly, S. Niverty, and N. Chawla, *J. Alloys Compds.* 818, 152918. (2020).
9. A.S.S. Singaravelu, J.J. Williams, H.D. Goyal, S. Niverty, S.S. Singh, T.J. Stannard, X. Xiao, and N. Chawla, *Metall. Mater. Trans. A* 51, 28. (2020).
10. M.B. Kelly, S. Niverty, and N. Chawla, *Acta Mater.* 189, 118. (2020).
11. V. Mazars, O. Caty, G. Couégnat, A. Bouterf, S. Roux, S. Denneulin, J. Pailhès, and G.L. Vignoles, *Acta Mater.* 140, 130. (2017).
12. S. Niverty, (2020).
13. S.S. Singh, T.J. Stannard, X. Xiao, and N. Chawla, *JOM* 69, 1404. (2017).
14. J. Samei, C. Pelligra, M. Amirmaleki, and D.S. Wilkinson, *Mater. Lett.* 269, 127664. (2020).
15. A. Isaac, F. Sket, W. Reimers, B. Camin, G. Sauthoff, and A.R. Pyzalla, *Mater. Sci. Eng. A* 478, 108. (2008).
16. T. Lacondemine, J. Réthoré, É. Maire, F. Célarié, P. Houizot, C. Roux-Langlois, C.M. Schlepütz, and T. Rouxel, *Acta Mater.* 179, 424. (2019).
17. H.A. Bale, A. Haboub, A.A. MacDowell, J.R. Nasiatka, D.Y. Parkinson, B.N. Cox, D.B. Marshall, and R.O. Ritchie, *Nature Mater.* 12, 40. (2013).
18. A.S.S. Singaravelu, J.J. Williams, J. Ruppert, M. Henderson, C. Holmes, and N. Chawla, *J. Mater. Sci.* (2020).
19. B.M. Patterson, L. Kuettner, T. Shear, K. Henderson, M.J. Herman, A. Ionita, N. Chawla, J. Williams, T. Sun, K. Fezzaa, X. Xiao, and C. Welch, *J. Mater. Sci.* 55, 11353. (2020).
20. C.S. Kaira, C.R. Mayer, V. De Andrade, F. De Carlo, and N. Chawla, *Microsc. Microanal.* 22, 808. (2016).
21. X. Yang, D. Gürsoy, C. Phatak, V. De Andrade, E.B. Gulsoy, and F. De Carlo, *Microsc. Microanal.* 22, 240. (2016).
22. C.S. Kaira, V. De Andrade, S.S. Singh, C. Kantzos, F. De Carlo, and N. Chawla, *Microsc. Microanal.* 23, 2220. (2017).
23. C. Shashank Kaira, X. Yang, V. De Andrade, F. De Carlo, W. Scullin, D. Gursoy, and N. Chawla, *Mater. Charact.* 142, 203. (2018).

Journal : 11837_JOM Dispatch : 4-5-2021 Pages : 12
Article No.: 4706 ☐ LE ☑ CP ☐ TYPESET ☑ DISK

24. Y. Wang, J. Gao, Y. Ren, V. De Andrade, and A.J. Shahani, *JOM* 72, 2965. (2020).
25. L.J. Ausderau, H.J. Gonzalez Malabet, J.R. Buckley, V. De Andrade, Y. Liu, and G.J. Nelson, *JOM* 69, 1478. (2017).
26. V. De Andrade, A. Deriy, M.J. Wojcik, D. Gürsoy, D. Shu, K. Fezzaa and F. De Carlo, SPIE Newsroom, (2016).
27. D. Gürsoy, F. De Carlo, X. Xiao, and C. Jacobsen, *J. Synchrotron Radiat.* 21, 1188. (2014).
28. J.J. Williams, Z. Flom, A.A. Amell, N. Chawla, X. Xiao, and F. De Carlo, *Acta Mater.* 58, 6194. (2010).
29. J.C.E. Mertens, J.J. Williams, and N. Chawla, *Nucl. Instrum. Methods Phys. Res. A* 800, 82. (2015).
30. C. Gobert, A. Kudzal, J. Sietins, C. Mock, J. Sun, and B. McWilliams, *Add. Manuf.* 36, 101460. (2020).
31. A. Kumar, and G.K.H. Pang, *IEEE Trans. Syst. Man Cybernet. B* 32, 553. (2002).
32. P.I. Guntoro, G. Tiu, Y. Ghorbani, C. Lund, and J. Rosenkranz, *Miner. Eng.* 142, 105882. (2019).
33. C. Bishop, *Pattern Recognition and Machine Learning* (Springer, New York, 2006).
34. T.F. Gonzalez, *Handbook of Approximation Algorithms and Metaheuristics 1* (Taylor & Francis, London, 2007).
35. B. Ma, X. Ban, H. Huang, Y. Chen, W. Liu, and Y. Zhi, *Symmetry* 10, 107. (2018).
36. T. Stan, Z.T. Thompson, and P.W. Voorhees, *Mater. Charact.* 160, 110119. (2020).
37. A. Tekawade, B.A. Sforzo, K.E. Matusik, A.L. Kastengren and C.F. Powell, in *Developments in X-Ray Tomography XII*, ed. B. Müller and G. Wang (SPIE, 2019), p. 67.
38. S. Evsevleev, S. Paciornik, and G. Bruno, *Adv. Eng. Mater.* 22, 1. (2020).
39. D. Chen, D. Guo, S. Liu, and F. Liu, *Symmetry* 12, 639. (2020).
40. X. Yang, F. De Carlo, C. Phatak, and D. Gürsoy, *J. Synchrotron Radiat.* 24, 469. (2017).
41. X. Yang, V. De Andrade, W. Scullin, E.L. Dyer, N. Kasthuri, F. De Carlo, and D. Gürsoy, *Sci. Rep.* 8, 2575. (2018).
42. I. Rizwan, I. Haque, and J. Neubert, *Inform. Med. Unlock.* 18, 100297. (2020).
43. Z. Zhou, R. Siddiquee, N. Tajbakhsh, and J. Liang, 1 (n.d.).
44. W. Zhang, X. He, W. Li, Z. Zhang, Y. Luo, L. Su, and P. Wang, *Image Vis. Comput.* 93, 103824. (2020).
45. X. Liu, Z. Deng, and Y. Yang, *Artif. Intell. Rev.* 52, 1089. (2019).
46. Q. Liu, A.B. Salberg and R. Jenssen, International Geoscience and Remote Sensing Symposium (IGARSS) 2018-July, 6943 (2018).
47. I. Goodfellow, Y. Benjio, and A. Courville, *Deep Learning* (MIT, Cambridge, 2016).
48. Z. Zhou, M. M. Rahman Siddiquee, N. Tajbakhsh, and J. Liang, in (2018), pp. 3–11.
49. L. C. Chen, Y. Zhu, G. Papandreou, F. Schroff and H. Adam, Proceedings of the European Conference on Computer Vision (ECCV) 801 (2018).
50. L.C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A.L. Yuille, *IEEE Trans. Pattern Anal. Mach. Intell.* 40, 834. (2018).
51. H. Zhao, J. Shi, X. Qi, X. Wang and J. Jia, Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017 2017-Jan, 6230 (2017).
52. S.S. Singh, C. Schwartzstein, J.J. Williams, X. Xiao, F. De Carlo, and N. Chawla, *J. Alloys Compds.* 602, 163. (2014).
53. M. Gao, C.R. Feng, and R.P. Wei, *Metall. Mater. Trans. A* 29, 1145. (1998).
54. S. Dey, M.K. Gunjan, and I. Chattoraj, *Corros. Sci.* 50, 2895. (2008).
55. W. Tian, S. Li, B. Wang, J. Liu, and M. Yu, *Corros. Sci.* 113, 1. (2016).
56. O. Ronneberger, P. Fischer, and T. Brox, in (2015), pp. 234–241.
57. G. Shi, S. Guan and X. Yang, (2011).
58. K. Jamart, Z. Xiong, G. D. Maso Talou, M. K. Stiles and J. Zhao, *Front. Cardiovasc. Med.*, 7, (2020).
59. C.S. Kaira, C. Kantzos, J.J. Williams, V. De Andrade, F. De Carlo, and N. Chawla, *Acta Mater.* 144, 419. (2018).
60. J. Bertels, T. Eelbode, M. Berman, D. Vandermeulen, F. Maes, R. Bisschops,and M. B. Blaschko, in (2019), pp. 92–100.
61. S. Jadon, 2020 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology, CIBCB 2020 (2020).

Journal : **11837_JOM** Dispatch : **4-5-2021** Pages : **12**
Article No.: **4706** ☐ LE ☑ CP ☐ TYPESET ☑ DISK

Journal : **11837**
Article : **4706**

Author Query Form

Please ensure you fill out your response to the queries raised below and return this form along with your corrections

Dear Author

During the process of typesetting your article, the following queries have arisen. Please check your typeset proof carefully against the queries listed below and mark the necessary changes either directly on the proof/online grid or in the 'Author's response' area provided below

Query	Details Required	Author's Response
AQ1	Author names: Please confirm that the author names are presented accurately and in the correct sequence (given names/initials, family name).	
AQ2	Journal instruction requires a city and country for affiliations. Please verify if the provided city and country are correct and amend if necessary.	