Journal of the European Ceramic Society 46 (2026) 118431

Contents lists available at ScienceDirect

Journal of the European Ceramic Society

journal homepage: www.elsevier.com/locate/jeurceramsoc

Grain size measurement of calcium phosphate ceramic using deep learning

Stefan Jakobs *, Bastien Le Gars Santoni [1], Moritz Liesegang, Marc Bohner

RMS Foundation, Robert Mathys-Strasse 1, 2544 Bettlach, Switzerland

ARTICLE INFO

Keywords:
Grain size
Segmentation
U-Net
Binary cross-entropy
Beta-tricalcium phosphate

ABSTRACT

Grain size determination is crucial for correlating microstructure with material properties in ceramics, but manual measurement is time-consuming and operator dependent. This study presents a neural network (U-Net) for automated grain boundary segmentation in scanning electron microscopy (SEM) images of β-tricalcium phosphate ceramics. Two loss functions, binary cross-entropy (BCE) and binary focal cross-entropy (BFCE), were evaluated. High segmentation performance was achieved across 70 validation images of varying quality, with intersection-over-union (IoU) scores of 0.953 (BCE) and 0.943 (BFCE). No image enhancement methods were necessary. The automated approach achieved higher precision than manual measurements (relative standard deviations $< 1.5\%$). Linear regression confirmed strong correlation between automated and manual measurements. Although no ground truth is available, systematic differences can be addressed through dataset-specific calibration. The workflow provides a robust and reproducible framework for automated grain size analysis, remaining effective in the presence of intra-grain topography, grain fractures, and blurred grain boundaries.

1. Introduction

Grain size measurement is a fundamental technique in the characterization of materials, especially in the field of ceramic materials, where the microstructure is closely related to mechanical properties [1–4]. Understanding and quantifying the grain structure provides insights into the behaviour and performance of the material under various conditions. Conventionally, this process is carried out manually through techniques such as line-intercept methods, which follow standards like ASTM E112 [5] or ISO 13383–1:2012 [6].

Nevertheless, manual procedures are labor-intensive and time-consuming, as they require careful and consistent tracing of grain boundaries across numerous micrographs. This raises the question of whether such evaluations can be automated. In this context, image segmentation offers a computational approach to partitioning micrographs into meaningful regions, enabling more efficient identification of grain boundaries. Classical segmentation algorithms (thresholding, random walker segmentation, watershed algorithms or other rule-based methods) have been demonstrated to be effective for grain boundary extraction in micrographs of polycrystalline materials, particularly when image contrast and edge definition are sufficient [7–11]. These methods often fail in the presence of pronounced intra-grain topography or under suboptimal imaging conditions, where low contrast and blurred grain boundaries hinder edge detection. Manual segmentation demands significant effort, introduces variability between users, and is inherently restricted in the number of grains that can be effectively analyzed within a given timeframe. The limitations of this approach are particularly apparent in materials with more than one phase or complex microstructures, where the evaluation is very time-consuming. This limits the applicability of manual grain size determination and makes the process vulnerable to bias introduced by the operator.

In order to address these challenges, the integration of machine learning techniques into materials science has gained traction in recent years [12–19]. Especially neural networks (NN) have demonstrated significant potential in the segmentation of microstructural images, such as those obtained via scanning electron microscopy (SEM) [13,15]. These deep learning methods have the advantage of processing large datasets within a fraction of the time required for manual analysis [19], while also enabling the extraction of additional microstructural descriptors, such as grain orientation or information about defects that provide a more comprehensive characterization of the material [13,15,16]. Also, the potential to extract new, easily accessible metrics (for example, aspect ratios of grains) is of interest for various applications, such as grain shape analysis in sintering studies or the calculation of phase-specific grain metrics in multiphase systems for automated quality control in industrial processes. Machine learning-based segmentation

* Corresponding author.
E-mail address: stefan.jakobs@rms-foundation.ch (S. Jakobs).
[1] Current address: Stryker, Dr. Homer Stryker Str. 1, 2545 Selzach, Switzerland

https://doi.org/10.1016/j.jeurceramsoc.2026.118431
Received 13 January 2026; Received in revised form 15 April 2026; Accepted 16 April 2026
Available online 17 April 2026

methods offer an additional technical benefit in that they can be applied to complex images [13,20]. This includes images affected by low contrast, inconsistent brightness, or non-optimal sample preparation. This suggests that the algorithm may be capable of identifying local and global patterns that are not perceptible to the human eye [21]. In manual grain size evaluation, image imperfections such as grain fractures, blurred features, or other artifacts introduced during sample preparation can lead to inaccuracies [8,9]. NN can be trained to distinguish between true grain boundaries and artefacts, thus offering enhanced robustness in challenging scenarios.

This work focuses on leveraging machine learning techniques to address several key challenges in grain boundary segmentation. The U-Net, a widely recognised framework for image segmentation [22], is employed here to analyse SEM images of β-tricalcium phosphate (β-TCP). β-TCP is a biocompatible and resorbable calcium phosphate ceramic commonly used as a bone graft substitute [23]. Its grain size strongly influences both mechanical stability [24] and biological performance [25], making reliable characterization an important factor for biomedical applications. The training process is done using two loss functions: binary cross-entropy (BCE), a widely utilised function in segmentation tasks, and binary focal cross-entropy (BFCE), a function specifically designed to address class imbalance issues and to focus on difficult-to-classify pixels during the training phase [26]. The objective of this study is to evaluate the efficacy of a machine learning approach in achieving robust and reproducible grain size analysis, with the aim of providing an alternative to manual procedures. The working hypothesis of this study is that a U-Net can achieve a precision comparable to manual grain size determination in a ceramic material, while substantially reducing user dependence and evaluation time.

2. Materials and methods

2.1. Sample preparation and SEM imaging

β-Tricalcium phosphate (β-TCP; $Ca_3(PO_4)_2$) samples were fabricated by slip casting as described in [27]. After sintering at 1100 °C for 3 h, dense cylindrical specimens were obtained. One end of each cylinder was ground using silicon carbide (SiC) abrasive papers and subsequently polished with diamond suspensions down to 0.25 µm. Final polishing was performed using a 0.2 µm silica suspension. To remove polishing residues, the specimens were cleaned in an ultrasonic bath: two cycles of 5 min in acetone, followed by one 5-minute cycle in ethanol. For SEM imaging, the cleaned samples were mounted on aluminium stubs using carbon adhesive tape and carbon coated. Imaging was performed on the polished side of the β-TCP cylinder with a EVO MA 25 scanning electron microscope in backscattered electron (BSE) mode at an acceleration voltage of 7 kV.

2.2. Model input data

The dataset comprised 120 SEM images of β-tricalcium phosphate (β-TCP), used without scale bars. In contrast to Fig. 1, most images simultaneously exhibited multiple challenging features, such as blurred regions or fractured grains, reflecting typical variations in SEM images. A total of 50 images were randomly selected for training and testing. These 50 images were automatically split into 90% training and 10% test data using the Scikit-learn Python library [28]. The five test images served solely as an internal test set during the training process. The remaining 70 images were reserved for the final validation and evaluation of the model performance. The ground-truth annotations were created manually by a single experienced operator, who used a touchscreen tablet to draw red lines (3 pixels wide) on each image. Annotation time per image ranged from 45 to 75 min, depending on the magnification, with lower magnifications requiring more time than

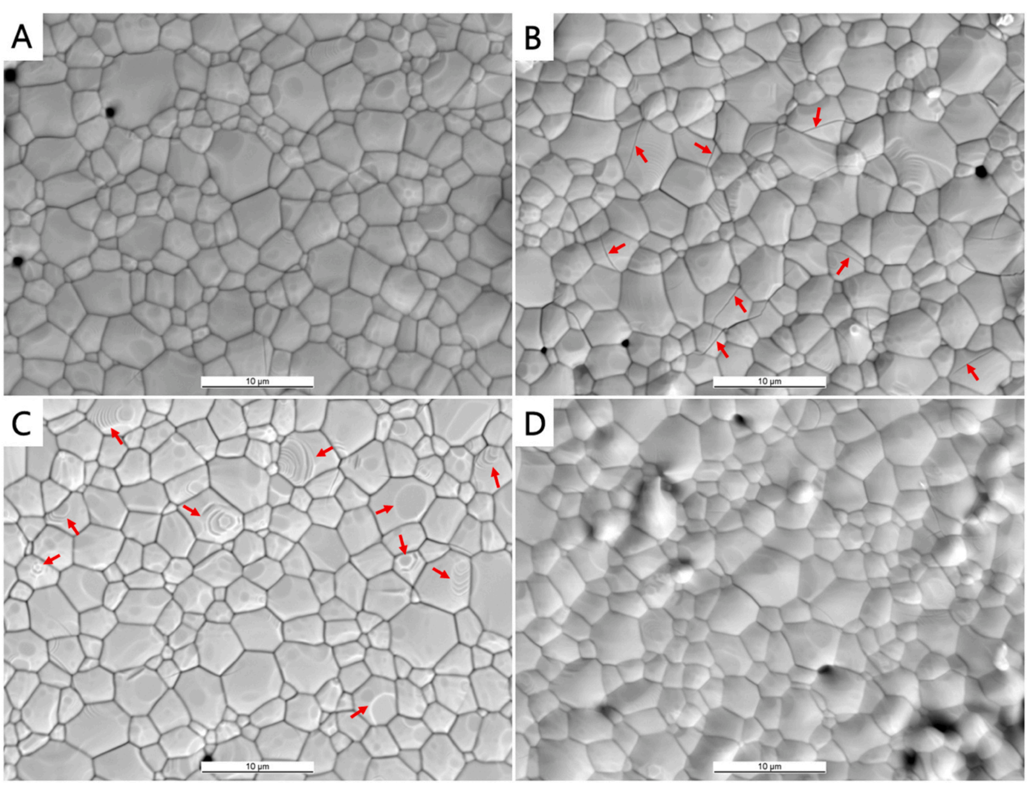

Fig. 1. Representative BSE images of Beta-tricalcium phosphate ceramic with varying image and sample preparation quality. A: Good image. B: Image with grain fractures (red arrows). C: Image with visible topography in the grains (red arrows). D: Blurred image. All images shown are from the test set.

higher magnifications. These images with red lines were converted into binary images, enabling the network to learn the relationship between the raw input and the segmentation target (see Fig. 2). The SEM images were acquired at magnifications ranging from 2500 × to 5000 × and have a resolution of 2048 × 1536 pixels. Each image was processed in its raw state, without the application of any filters or image enhancements, such as contrast or brightness adjustments. Each image was divided into 12 non-overlapping patches of 512 × 512 pixels to enable efficient training. Data augmentation consisted of rotations by 90°, 180°, and 270°, as well as horizontal mirroring (left–right flip) to increase the dataset eightfold, resulting in 96 patches per original image.

2.3. Neural network architecture

The U-Net [22], widely recognized for its effectiveness in image segmentation tasks, was employed in this study. It consists of a symmetric encoder–decoder structure with skip connections linking corresponding layers. These connections preserve spatial information during downsampling and enhance segmentation accuracy, especially with limited training data. The implemented U-Net (Fig. 3) contained five layers and approximately 1.9 million trainable parameters. All layers used ReLU activation, He-normal initialization, and "same" padding. Dropout was applied to reduce overfitting, with rates ranging from 0.1 (first layers) to 0.3 (deep layers). Training was conducted using a batch size of 16, the Adam optimizer used the default learning rate (0.001), and training was capped at 40 epochs. The dataset exhibited a notable class imbalance: on average, only 12.7% (standard deviation (SD): 3.8%) of all pixels across the 120 images correspond to grain boundaries, while the remaining 87.3% represent background. To address this, two loss functions were compared: BCE and BFCE. BCE corresponds to $\gamma = 0$. BFCE applies a gamma-weighting (here used with $\gamma = 4$) to better handle the class imbalance and to improve robustness in difficult to classify regions [26]. Both loss functions were implemented and tested under identical network configurations. Model outputs were computed using sigmoid activation, yielding per-pixel probability maps. Pixels with a value ≥ 0.80 were classified as grain boundary (black); all others were classified as background (white). This threshold was determined empirically by applying it to four images (Fig. 1), using them as representatives for the whole dataset. The model was implemented in Keras with a TensorFlow (version 2.9.1) backend [29,30].

2.4. Grain size measurement in binary images

Following the segmentation process, the grain size of the binary images was analysed using established Python-based image processing routines (Fig. 2). Grains were separated using the Watershed algorithm [31], assigning a unique ID to each grain, while grains intersecting the image boundaries were excluded from further evaluation (Fig. 5). Segmentation results were visualized using Napari [32].

Where necessary, the influence of different grain boundary thicknesses on the calculated grain areas was minimized by introducing a post-processing step prior to the watershed-based separation. In this approach, the binarized grain boundary masks were reduced to one-pixel-wide centerlines by morphological skeletonization using the scikit-image library [18,33]. The interiors of the grains were then re-labeled by expanding the connected white regions until they met the skeleton lines, ensuring a uniform boundary thickness of one pixel across the entire image.

2.5. Performance measures

The quality of segmentation was primarily assessed using the intersection-over-union (IoU) metric [34], which quantifies the overlap between predicted and ground truth segmentations. Additionally, the

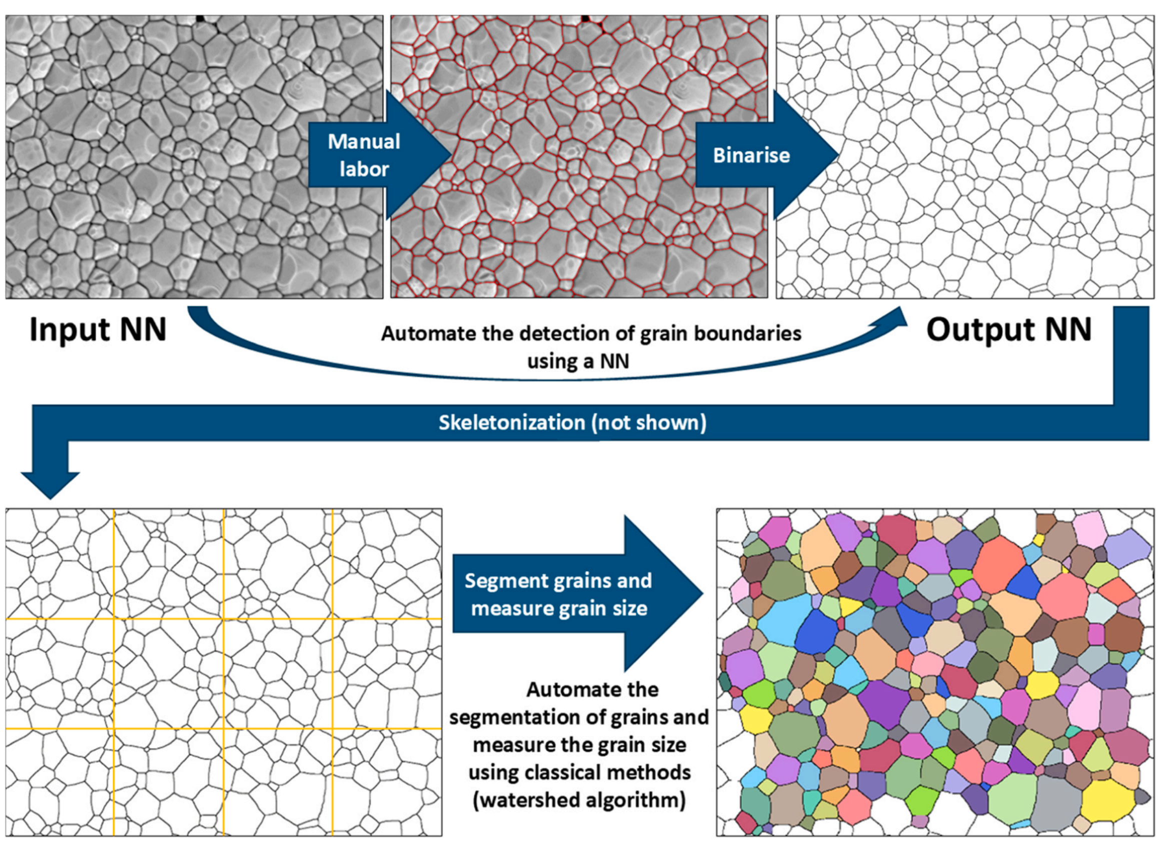

Fig. 2. Workflow during training (top row) and workflow during application with trained neural network (bottom row). The image on the bottom left shows how a complete image is divided into patches.

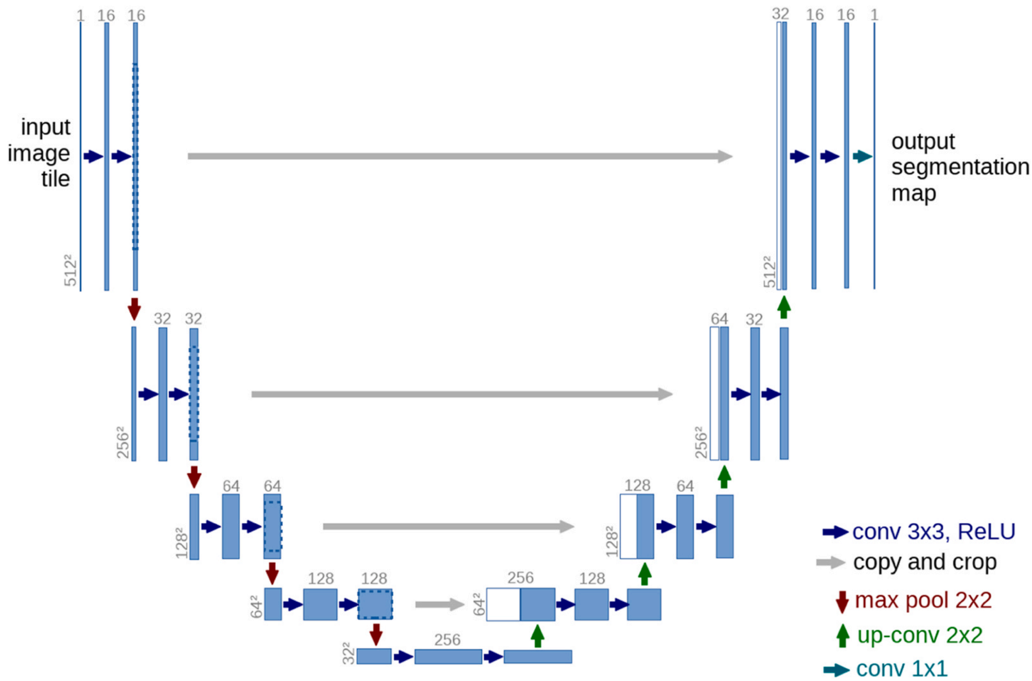

Fig. 3. Image of the network architecture used. Figure taken and adapted to the self-implemented network from [22].

Dice score [35] was utilised to obtain a more comprehensive evaluation of the model's performance. To validate grain size predictions, the average equivalent circle diameter from the segmented images was compared to manual measurements according to the procedure of Heyn in section 13 of ASTM E112 [5] (regarding the difference between equivalent circle diameter and the manual line-intercept length, see

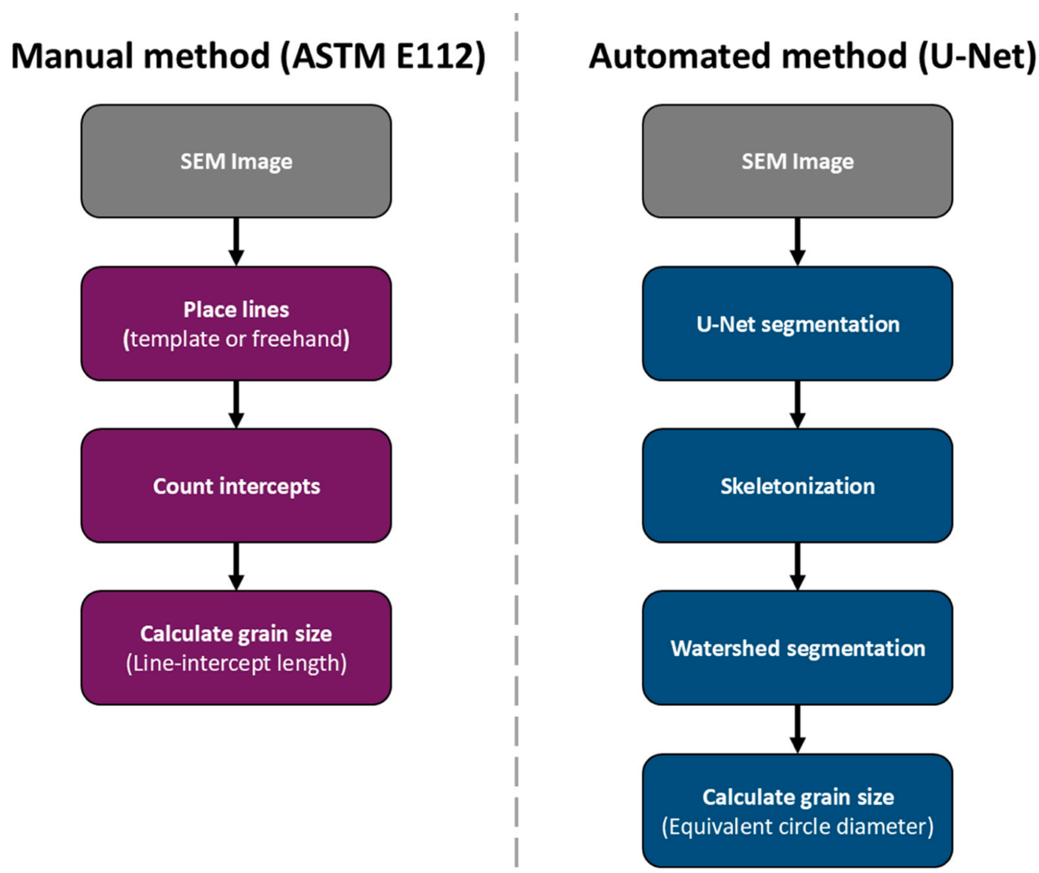

Fig. 4. Comparison of the manual grain size determination workflow according to ASTM E112 (left) and the automated workflow based on U-Net segmentation proposed in this study (right).

Section 4.2). Fig. 4 provides a comparative overview of the automated segmentation workflow and the manual line-intercept method. To evaluate precision, manual grain size analysis was performed by three independent operators using two variants of the line-intercept method: (i) a template with pre-positioned lines and (ii) without a template, allowing free placement of the lines. The total length of all test lines placed on the micrograph is defined as L, and the number of grain boundary intersections is recorded as N_I. The mean line-intercept length $\bar{l}$ is then calculated according to $\bar{l} = \frac{L}{N_I}$, following the procedure described by Heyn in ASTM E112. For each image, at least 100 intercepts were evaluated (Fig. 6). Precision was reported as the relative standard deviation (RSD) across evaluators.

Because the automated method is operator-independent, precision was assessed by evaluating the algorithm's response to image variations. For this purpose, the 70 images that were not used for training were manipulated, and three image properties, brightness, contrast, and sharpness, were systematically varied. Each property was adjusted to match its 5th and 95th percentile values, based on the respective distribution across the full dataset (120 images). This resulted in six modified images per original input. A detailed description of the image manipulation procedures for contrast, brightness, and sharpness, along with representative examples of the modified images and their corresponding segmentation results, is provided in the Supplementary Material (Figures S1 to S12). ChatGPT [36] was used to support programming tasks described in the supplement and to assist with language refinement during manuscript preparation. All outputs generated by the tool were carefully reviewed and verified by the authors. Regression analysis was performed using Origin Software [37].

3. Results

3.1. Binary cross-entropy as loss function

The neural network that was trained with BCE achieved a mean intersection-over-union (IoU) score of 0.953 and a Dice score of 0.979 across the 70 validation images. Fig. 5 presents representative raw

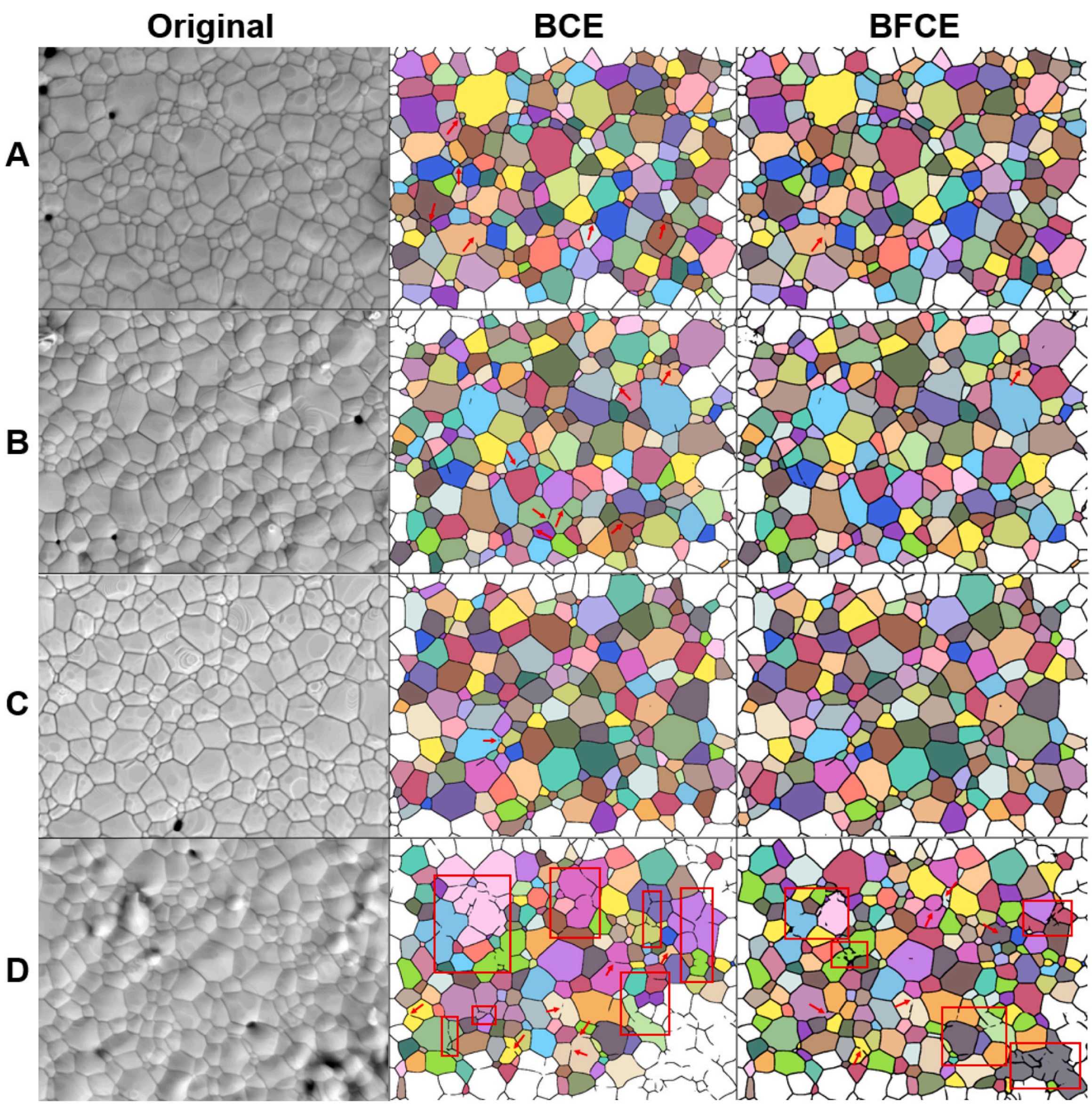

Fig. 5. Segmentation results of the neural networks applied on the Images of Fig. 1 (Original, left) trained with binary cross-entropy (BCE, middle) and binary focal cross-entropy (BFCE, right). All images represent the *raw* segmentation outputs prior to the skeletonization-based boundary normalization described in Section 2.4. The BFCE model tends to produce thicker grain boundary lines compared to BCE, resulting in visibly narrower enclosed grain areas. Marked regions (red arrows and boxes) highlight local differences in segmentation continuity and boundary definition.

segmentation outputs (prior to skeletonization; see Section 2.4) for the four image types. Across the images, several observations regarding segmentation accuracy can be made. Image A in Fig. 5 (BCE column) yields a segmentation that enables further quantitative grain size evaluation. However, local inconsistencies exist at the patch boundaries. These areas pose challenges for the pixel classification algorithm. Image B in Fig. 5 (BCE column) also shows an segmentation result suitable for quantitative analysis, despite many areas where grain fractures are present. In some instances, parts of the grain boundaries are misclassified, although this does not affect the final watershed segmentation outcome. Image C in Fig. 5 (BCE column) shows that intra-grain topography does not interfere with the segmentation process. Image D in Fig. 4 (BCE column) illustrates the effect of blurred imaging conditions, where grain boundaries appear incompletely closed, resulting in discontinuities in the segmentation.

Based on these examples, the following statistical analysis evaluates the network performance on a subset of 24 original images selected from the independent 70-image test set. These 24 images comprised six

Table 1

Summary of grain size evaluations for a set of 4 × 6 SEM images using four methods: manual segmentation in accordance with ASTM E112 with template, manual segmentation without template, and automated segmentation using a neural network with binary cross-entropy (BCE) and binary focal cross-entropy (BFCE) as loss functions. Each image number (1–24) refers to one original SEM image. For each image, the mean grain size ± SD is followed by the corresponding RSD. The RSD values shown at the category level (e.g., 'Good', 'Grain fractures') represent the average RSD across all images within that category. For the machine learning methods, six modified variants of every image were generated (brightness decreased/increased, contrast decreased/increased, sharpness decreased/increased), and the reported mean values ± SD were calculated from the six corresponding segmentation results. The RSD therefore represents the variability of the predicted grain size within these six modified variants of the same image. For each manual method, three independent evaluators annotated every image. The values from the machine learning methods correspond to the average equivalent circle diameter calculated by Scikit-learn [33]. The values of the line-intercept method correspond to the line-intercept length. Depending on grain shape and microstructural anisotropy, the line-intercept method can calculate grain sizes that are between 10% and 70% smaller than the equivalent circle diameter (in this dataset probably around 13%, see Section 4.2). The lower grain size values observed for the blurred images are attributed to a slightly higher Ca/P molar ratio of the corresponding β-TCP batch, which is known to reduce grain growth in calcium phosphate ceramics [25]. Unitless numbers are in μm. ± values represent standard deviation (SD); percentage values indicate relative standard deviation (RSD).

		Line-intercept method (ASTM E112)						Machine learning method					
Image	Type	Average grain size with template			Average grain size without template			Average grain size, BCE, γ = 0			Average grain size, BFCE, γ = 4		
1	Good	1.99 ± 0.06	3.08%	3.51%	2.52 ± 0.22	8.65%	5.79%	2.52 ± 0.04	1.50%	0.62%	2.40 ± 0.01	0.60%	0.65%
2		2.47 ± 0.10	3.97%		2.50 ± 0.14	5.51%		2.42 ± 0.01	0.48%		2.39 ± 0.01	0.12%	
3		1.41 ± 0.06	4.03%		1.45 ± 0.15	10.60%		1.53 ± 0.01	0.43%		1.47 ± 0.01	0.81%	
4		2.01 ± 0.04	2.20%		2.29 ± 0.10	4.19%		2.36 ± 0.02	0.91%		2.22 ± 0.03	1.43%	
5		1.60 ± 0.07	4.25%		1.72 ± 0.09	4.98%		1.77 ± 0.01	0.17%		1.70 ± 0.01	0.24%	
6		1.23 ± 0.04	3.51%		1.54 ± 0.01	0.79%		1.47 ± 0.01	0.26%		1.40 ± 0.01	0.71%	
7	Grain fractures	2.15 ± 0.03	1.24%	2.46%	2.61 ± 0.22	8.31%	6.97%	2.72 ± 0.02	0.59%	1.29%	2.63 ± 0.02	0.77%	1.48%
8		2.24 ± 0.03	1.30%		2.61 ± 0.07	2.56%		2.53 ± 0.01	0.48%		2.45 ± 0.04	1.49%	
9		2.22 ± 0.08	3.56%		2.49 ± 0.12	4.93%		2.48 ± 0.02	0.72%		2.41 ± 0.02	0.84%	
10		1.97 ± 0.05	2.31%		2.32 ± 0.21	9.13%		2.56 ± 0.02	0.65%		2.34 ± 0.03	1.08%	
11		2.69 ± 0.14	5.26%		2.79 ± 0.21	7.53%		2.66 ± 0.04	1.36%		2.31 ± 0.09	3.69%	
12		2.50 ± 0.03	1.10%		2.88 ± 0.27	9.39%		2.64 ± 0.10	3.91%		2.43 ± 0.02	0.99%	
13	Intra-grain topography	2.44 ± 0.01	0.54%	1.95%	2.70 ± 0.14	5.30%	4.61%	2.68 ± 0.01	0.23%	0.25%	2.70 ± 0.01	0.24%	0.52%
14		2.48 ± 0.05	1.89%		2.65 ± 0.17	6.45%		2.70 ± 0.01	0.25%		2.66 ± 0.01	0.52%	
15		2.41 ± 0.06	2.29%		2.52 ± 0.13	5.23%		2.58 ± 0.01	0.20%		2.57 ± 0.01	0.40%	
16		2.51 ± 0.06	2.49%		2.80 ± 0.12	4.11%		2.84 ± 0.01	0.17%		2.82 ± 0.02	0.65%	
17		2.05 ± 0.05	2.36%		2.39 ± 0.03	1.23%		2.39 ± 0.01	0.17%		2.36 ± 0.01	0.59%	
18		2.38 ± 0.05	2.11%		2.49 ± 0.13	5.35%		2.56 ± 0.01	0.47%		2.42 ± 0.02	0.73%	
19	Blurred	1.35 ± 0.04	3.17%	3.71%	1.60 ± 0.05	3.11%	8.52%	1.66 ± 0.03	1.66%	2.66%	1.57 ± 0.02	1.57%	1.77%
20		1.03 ± 0.02	1.95%		1.25 ± 0.12	9.44%		1.35 ± 0.04	2.96%		1.17 ± 0.03	2.81%	
21		1.06 ± 0.02	2.03%		1.21 ± 0.07	5.90%		1.32 ± 0.05	4.02%		1.12 ± 0.02	1.65%	
22		1.05 ± 0.02	2.35%		1.14 ± 0.13	11.66%		1.29 ± 0.03	1.98%		1.13 ± 0.02	1.92%	
23		1.21 ± 0.08	6.35%		1.34 ± 0.11	8.22%		1.44 ± 0.06	4.09%		1.24 ± 0.02	1.28%	
24		1.18 ± 0.08	6.42%		1.48 ± 0.19	12.79%		1.39 ± 0.02	1.26%		1.24 ± 0.02	1.38%	
		Mean RSD:	2.91%		Mean RSD:	6.47%		Mean RSD:	1.28%		Mean RSD:	1.11%	

representative images for each of the four image categories (A–D) shown in Fig. 1. For the robustness analysis presented in Table 1, each of these 24 images was additionally modified by systematically varying brightness, contrast, and sharpness as described in Section 2.5 (resulting in six modified images for every singe image). The quantitative results (Table 1) were obtained after applying the skeletonization-based boundary normalization described in Section 2.4, ensuring that all grain size values are independent of segmentation line thickness. The grain size values derived from the neural network correspond to the equivalent circle diameter, whereas the manual measurements are based on the mean line-intercept length. These two metrics are not directly comparable. The results of these 24 test images, evaluated by multiple users and methods, show a RSD of 1.28% for the U-Net trained with BCE. Manual calculation using the line-intercept method with a template yields a RSD of 2.91%. Without the use of a template, the RSD increases to 6.47%.

To assess the precision and systematic behavior of the algorithm, a linear regression analysis was performed (Fig. 7). All 70 images that had not been used during training were evaluated using the neural network (with post-processing) and compared to the corresponding grain size measurements obtained manually by a single operator using the line-intercept method without a template, which was the only manual procedure applied across all images. This allowed for a point-by-point comparison over a broad data range. The regression analysis yielded a coefficient of determination (R^2) of 0.941.

3.2. *Binary focal cross-entropy as loss function*

Fig. 5 illustrates the segmentation performance of the neural network trained with BFCE ($\gamma = 4$), using the same representative images A to D as in the BCE evaluation for direct comparison. The images represent the unprocessed network outputs before applying the skeletonization-based boundary normalization. When applied to all 70 validation images, BFCE yielded an IoU score of 0.943 and a slightly improved Dice score of 0.980, compared to BCE. Across the four example images, several observations regarding segmentation accuracy can be made. In Image A and B in Fig. 5 (BFCE column), the BFCE model corrects some of the local inconsistencies previously observed with BCE (see markings in Fig. 5), resulting in a more coherent segmentation. For Image C in Fig. 5 (BFCE column), no substantial differences compared to BCE are observed. In contrast, Image D in Fig. 5 (BFCE column), despite the low image quality, shows clear improvements in segmentation with BFCE, particularly in areas where the lack of sharpness previously hindered segmentation. The statistical evaluation of the 24 test images (Table 1) was performed on the skeletonized segmentations. The resulting RSD of 1.11% for the BFCE model is similarly low to that obtained with BCE (1.28%).

Again, a linear regression analysis was performed to evaluate the systematic behavior of the algorithm. As in the previous case, all 70 validation images were segmented and post-processed using the skeletonization-based normalization before comparison to the corresponding manual grain-size values obtained via the line-intercept method without a template. This one-to-one comparison yielded a coefficient of determination (R^2) of 0.960. As with the BCE model, it must be emphasized that the regression analysis does not provide any information about accuracy, as no true values are available for comparison.

4. Discussion

4.1. *Advantages of the neural network approach*

The primary objective of this study was to evaluate whether automated image segmentation using a convolutional neural network can complement or, depending on the application, substitute manual grain size determination in ceramic materials. Manual procedures, such as the line-intercept method defined in ISO 13383–1:2012 [6], are time-consuming and operator-dependent, with reported deviations of up to ±10% for homogeneous, single-phase ceramics when 100 grains are measured [6]. This variability between users and the limited number of grains typically analyzed (illustrated in Fig. 6) highlight the need for a more efficient and reproducible approach.

In this context, a U-Net was applied to segment SEM images of a calcium phosphate ceramic and evaluated the influence of two different loss functions, BCE and BFCE, on segmentation performance under varying image quality conditions. This approach was used to assess the suitability of deep learning-based segmentation for grain size analysis, particularly in cases where segmentation is challenged by grain fractures, intra-grain topography, or blurred grain boundaries (Fig. 1 and Fig. 5). One important advantage of the neural network-based approach lies in its ability to consistently segment all fully depicted grains in an image. While manual grain size determination evaluates only a subset of grains within an image, the neural network analyses all complete grains present in the image. Since no ground truth is available for grain size measurements, accuracy could not be directly assessed. This limitation is rooted in the fact that true grain size distributions are inherently three-dimensional, while standard methods rely on 2D projections. Several studies emphasize that a reliable true grain size can only be determined through full 3D reconstruction of the microstructure, typically involving techniques such as serial sectioning or X-ray tomography, which are impractical for routine analysis [38]. Instead, precision was quantified via the RSD, which remained below 1.5%. This value quantifies the algorithm's sensitivity to variations in image brightness, contrast, and sharpness for a given image, rather than the variability of grain size across different images or microstructures or inter-operator variability, which is not applicable to an automated method. Accordingly, the precision assessment presented here is closer to an evaluation of the algorithm's robustness against variations in imaging conditions. Despite this distinction, the consistently low RSD values (below 1.5%) indicate a better reproducibility for the neural network compared to manual procedures (RSD = 2.9% with template; 6.5% without). These findings therefore suggest that a convolutional neural network can achieve such segmentation that the deduced average grain size is comparable to manual grain size determination while substantially reducing user dependence and evaluation time.

Although the training was scheduled for up to 40 epochs, convergence was typically reached after approximately 20 epochs. Training with 45 images for 20 epochs on a CPU (AMD Ryzen 9 5950X, 16 cores) required roughly 15.5 h, whereas inference took about ten seconds per image. On a GPU (Graphics Processing Unit) with CUDA (Compute Unified Device Architecture [39]) acceleration, the training time would likely be reduced substantially, potentially enabling near real-time evaluation. It should also be noted that the overall efficiency gain of the automated approach also depends on practical factors such as the imaging setup and data management workflow, including how easily scale information can be extracted from the image management software, particularly when images are acquired at varying magnifications.

Previous studies [15,21] suggest that useful segmentation performance can be obtained with relatively small training datasets. For instance, DeCost et al. [13] reported robust segmentation results using only 24 annotated micrographs. However, the ultrahigh carbon steel microstructures analyzed in that study contain clearly distinguishable carbide features along grain boundaries, which may simplify the segmentation task. Consequently, the required training dataset size may vary depending on the specific microstructure and image characteristics.

In the present study, the combination of dropout regularization and data augmentation, together with the consistently high segmentation performance on 70 unseen images of varying quality, suggests that the model generalizes well without overfitting. Overall, the convolutional neural network offers a clear improvement over conventional manual methods by combining precision, reproducibility, and efficiency. It eliminates user-dependent variability, consistently segments all fully depicted grains across large image sets, and provides a scalable

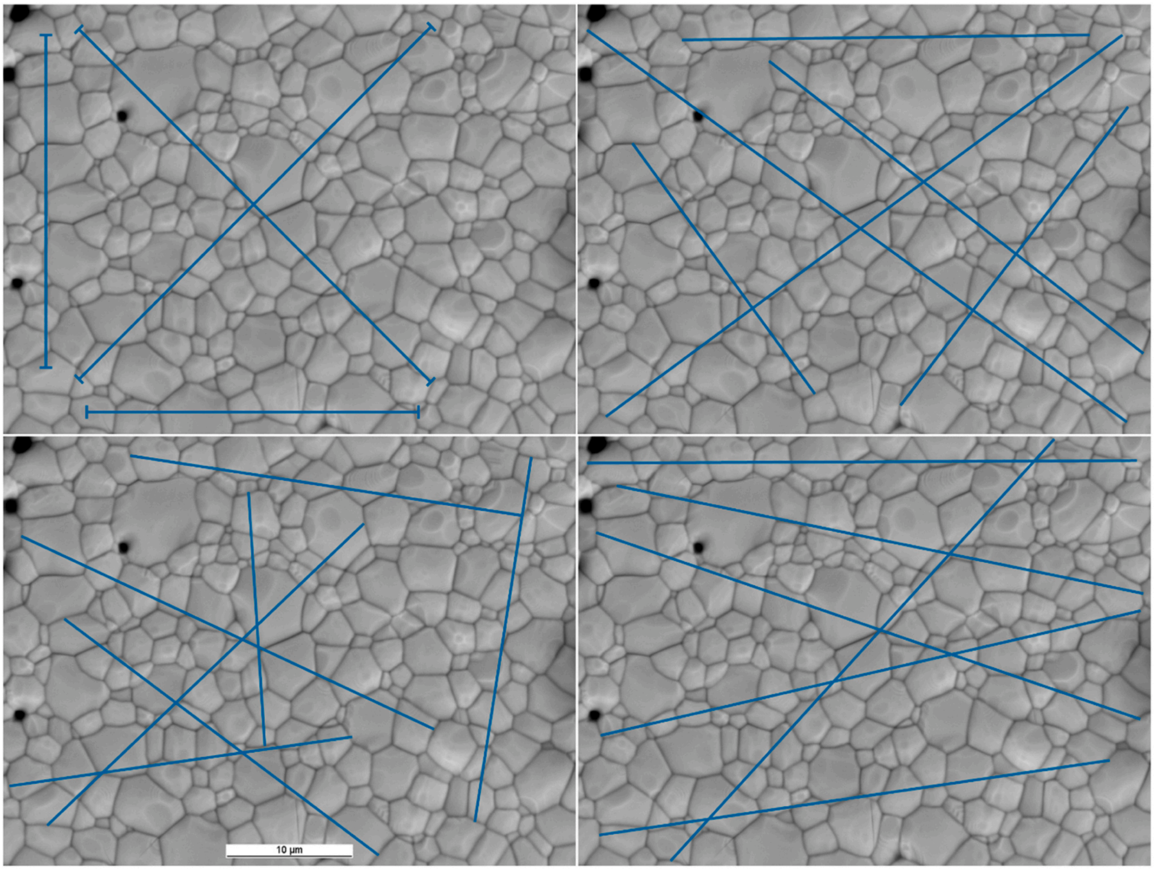

Fig. 6. Comparison of the manual line cutting method according to ASTM E112 (procedure according to Heyn, section 13 of the standard). The top left shows the line placement using a template. The other three images were created by three different people and show how different the random line placement can be (The goal for line placement was defined as at least 100 grain intersections in each image).

framework for future applications in automated microstructure analysis.

4.2. Evaluation of loss functions and performance metrics

The lower IoU scores obtained with BFCE, despite the often visually more accurate segmentations, required the inclusion of an additional performance metric, the Dice score. While both metrics are commonly used in segmentation tasks, they emphasize different aspects of performance. IoU penalizes false positives and false negatives more uniformly, whereas the Dice score places greater weight on true positives, making it more sensitive to the accurate segmentation of narrow or underrepresented structures like grain boundaries [40,41]. This is particularly relevant for the BFCE loss, which applies a weighting factor to less frequent pixel classes. Several studies have shown that focal loss can outperform standard BCE in handling class imbalance and improving segmentation of small or low-contrast structures, especially in medical and structural imaging tasks where such imbalances are common [42,43]. As a result, segmentation performance can improve in image regions with blurred features or low contrast. However, this can come at the expense of performance in simpler or well-defined areas, potentially explaining the slightly lower IoU scores observed with BFCE. Although the absolute difference between the Dice scores of BCE and BFCE models was small (0.979 vs. 0.980), the nonlinear behavior of the metric implies that such improvements near the upper limit may still indicate relevant segmentation gains. Such marginal numerical differences illustrate the limitations of using a single metric for assessing segmentation quality. Therefore, visual inspection remains an important supplement to quantitative measures, especially for evaluating segmentation performance in challenging or heterogeneous image conditions.

A linear regression analysis revealed coefficients of determination of 0.941 (BCE) and 0.960 (BFCE), indicating that the predicted grain sizes follow a consistent and approximately linear trend relative to manual measurements. Both loss functions have their advantages: While the BFCE loss improves segmentation robustness under poor image quality, the BCE loss is simpler to apply because it does not require tuning of the γ parameter. Testing both losses allows selecting the optimal configuration for different datasets. Without considering the difference between the equivalent circle diameter and the line-intercept length, the regression slopes for both loss functions were close to one (Fig. 7), indicating that the predicted and manually measured grain sizes were nearly proportional. Even when accounting for the inherent metric difference (equivalent circle diameter /line-intercept length), the accuracy could be further improved by applying a simple linear correction factor, for instance, based on manually determined reference values available for a specific material.

Systematic deviations between neural-network-based grain size values and reference measurements could be compensated by applying a linear calibration factor, thereby combining the high precision of the automated segmentation with an accuracy consistent with established reference methods. However, such corrections raise the fundamental question of what reference the adjustment should be based on. Even manually derived values, although supported by standardized procedures such as ISO 13383–1 and ASTM E112, do not represent a ground truth, but rather a well-established approximation. Any correction strategy must therefore balance improved alignment with these reference values against the risk of introducing new biases or limiting the model's generalizability.

In addition to the IoU and Dice scores discussed above, further quantitative metrics were calculated to assess classification consistency at the individual pixel level. Specifically, overall accuracy and precision

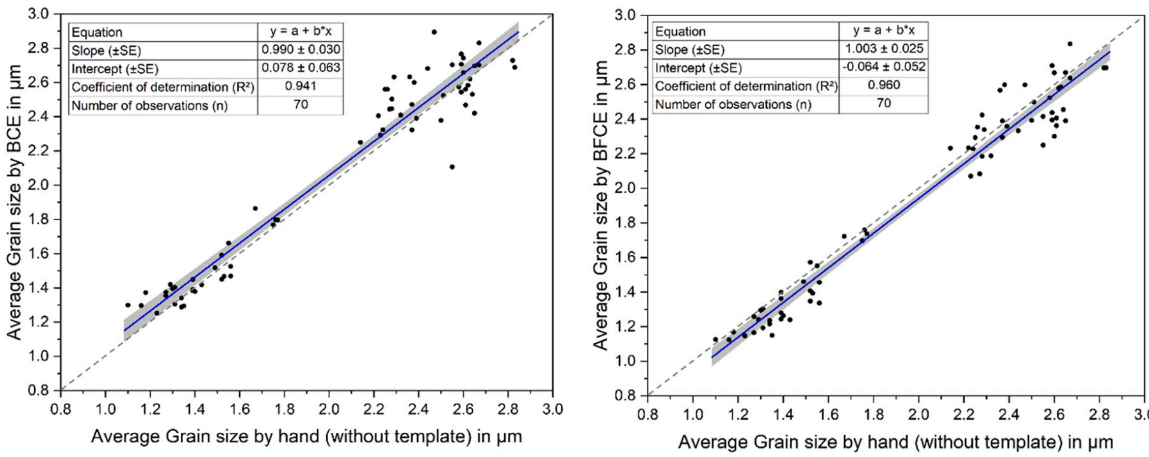

Fig. 7. Linear regression of automated (neural network) versus manual (ASTM E112) grain size measurements for 70 validation images not used during training, shown for both loss functions (BCE and BFCE).

were determined to complement the metrics by quantifying the ratio of correctly and incorrectly classified pixels. For this analysis, the binary images from the test subset (10% of the 50-image dataset) were used as the reference ("ground truth"). Both accuracy and precision were computed according to the definitions in [44] and using the implementation provided in the scikit-learn library [28]. For BCE, a precision of 96.1% and an accuracy of 97.7% were obtained. BFCE achieved a precision of 97.6% and an accuracy of 96.1%. These values are in a comparable range to those reported in the literature, such as Azimi et al. [45], who reported maximum precision values of 99.1% and maximum accuracy values of 95.2% for steel microstructural classification.

4.3. Limitations and segmentation challenges

Despite the promising results, certain limitations were identified. One issue concerns segmentation inconsistency at the boundaries of the 512×512 patches used during training (Fig. 2, bottom left and Fig. 5). These boundary artifacts may be reduced by overlapping patches or applying post-processing techniques to smooth transitions. The patch size of 512×512 pixels was chosen as a practical compromise between computational efficiency and spatial context, and is closely aligned with the originally proposed U-Net input size of 572×572 pixels [22]. It also fits well with the 2048×1536 pixel resolution of the dataset images, allowing for a convenient division into 12 non-overlapping patches. Tests with smaller patches (e.g. 256×256 pixels) did not improve segmentation outcomes and reduced the amount of contextual information available to the network. Another challenge was encountered in images containing very blurred grain boundaries (Fig. 1). In such cases, the network occasionally merged several grains into one larger connected region. In images containing more than 100 grains, these errors had only a limited influence on the overall average grain size. This effect can cause segmentations that appear visually inaccurate (e.g. Fig. 5D) but yield quantitative results close to the manually determined values. Furthermore, the network showed consistent segmentation in images exhibiting pronounced intra-grain topography, indicating that internal textural features within grains do not substantially affect boundary detection.

An additional observation concerned the behavior of the BFCE loss function. This model tended to produce more coherent and continuous segmentation maps in difficult image conditions but often generated thicker grain boundaries compared to BCE (Fig. 5). After post-processing, the results showed a stronger correlation with manual measurements then BCE (Fig. 7). Attempts were also made to optimize the pixel classification threshold applied to the sigmoid activation output (i.e. before post-processing). However, due to the variable SEM image quality, a uniform threshold did not yield consistent improvements across all cases. Similarly, varying the γ parameter of the BFCE loss function was explored. While the present study used $\gamma = 4$, additional tests with the commonly used $\gamma = 2$ [26] resulted in intermediate behavior between BCE ($\gamma = 0$) and full focal loss ($\gamma = 4$).

Manual segmentations were annotated with lines of 3-pixel thickness. Using thinner lines might reduce the width of grain boundaries, but would considerably increase the effort required for annotations, making it more difficult to connect boundary lines precisely at grain intersections. Given the image resolution of 2048×1536 pixels, the selected line thickness was considered appropriate for balancing precision and practical feasibility. Another point of consideration was the potential benefit of applying image preprocessing to improve edge definition in blurred images before segmentation. While such enhancements (e.g., sharpening) may improve segmentation quality, initial tests showed that manually processing the images significantly increased preparation time. This undermines one of the main strengths of the proposed method, its efficiency, so no image enhancement was applied in this study.

It also should be noted that the equivalent circle diameter used in the automated evaluation is not directly comparable to the mean line-intercept length obtained from manual measurements. The literature reports varying relationships between the equivalent circle diameter and the mean line-intercept length, depending on grain shape and microstructural anisotropy. Mingard et al. [46] found differences of approximately 10% within a round-robin comparison of manual grain size measurements, whereas Roebuck et al. [47] reported deviations of up to 70% for tungsten carbide–cobalt hard metals. Coutinho et al. [48] investigated different conversion factors between these metrics and emphasized that selection of the most suitable method depends on grain topography: the line-intercept method is more representative for elongated or oriented microstructures (e.g., after rolling of a material), while the equivalent circle diameter is appropriate for equiaxed grains. Taking these findings into account, a correction of approximately 13% represents a reasonable estimate for the present dataset and is close to the empirically determined value of $\sqrt{\frac{4}{\pi}}$ [47]. If such factors were considered, the slopes of the regression curves (Fig. 7) would change from 0.990 to 0.861 (BCE) and from 1.003 to 0.872 (BFCE).

Beyond the systematic differences arising from the choice of grain size metric (equivalent circle diameter vs. line-intercept length) and the inherent limitations of 2D cross-sections as approximations of 3D microstructures, additional sources of systematic bias should be considered. In this study, all ground truth annotations were produced by a single operator. Consequently, the neural network inherently learns the annotation style and subjective decisions of that individual, such as the precise placement of grain boundary lines in ambiguous regions. This single-annotator bias could be mitigated in future work by incorporating annotations from multiple independent operators or by using consensus-based labeling strategies to create more representative training data. Furthermore, the empirically determined pixel classification threshold of 0.80 applied to the sigmoid output introduces another potential source of systematic bias. A higher threshold produces thinner grain boundaries, while a lower threshold results in thicker boundaries, both of which systematically affect the measured grain areas. Although the skeletonization-based post-processing step (Section 2.4) reduces this effect by normalizing all boundaries to a uniform width of one pixel, it does not fully eliminate the influence of the threshold, as pixels near ambiguous grain boundaries may still be systematically included or excluded depending on the chosen value.

4.4. *Transferability to other materials*

Future studies should investigate the generalizability of the presented method to other ceramic systems, particularly those with multiphase compositions or more complex microstructures. Functions available in the scikit-image library [33], such as the calculation of grain-level descriptors (e.g. equivalent circle diameter, orientation, axis lengths), can be used to identify material- or phase-specific patterns in segmentation results. Widely used multiphase ceramics such as zirconia-toughened alumina (ZTA) may serve as suitable case studies for evaluating the adaptability of this segmentation workflow to a broader range of materials and application scenarios.

4.5. *Implications*

The presented segmentation approach offers promising opportunities for standardizing grain size analysis workflows in materials research and industrial quality control. By supporting consistent evaluation across a wide range of image qualities, machine-learning-based methods can help reduce user variability and improve data throughput. Moreover, the automated extraction of grain-level features can facilitate microstructural characterization and can give more insights into the study of ceramics. The increasing automation and digital transformation of processes highlight the potential of integrating neural network-based segmentation tools to complement manual expertise with scalable and reproducible evaluation.

5. Conclusions

This study demonstrates that a U-Net enables precise and reproducible grain boundary segmentation in SEM images of β-tricalcium phosphate ceramics. Both loss functions, BCE and BFCE, yielded consistent grain size results with RSD below 1.5%, surpassing manual evaluation in precision. Linear regression analyses confirmed a strong proportional relationship between automated and manual measurements. These results confirm the hypothesis that a U-Net based on a convolutional neural network can produce results, in term of average grain size, comparable to those determined manually, even with improved precision.

At the same time, this minimizes user dependency and the time required per image. While the automated approach exhibits high precision, its absolute accuracy depends on the reference metric (equivalent circle diameter or line-intercept length). Systematic differences between equivalent circle diameter and line-intercept length may be compensated through dataset-specific calibration against known reference data. Such correction could align absolute values with established standards without compromising reproducibility. The presented workflow provides a robust framework for automated grain size analysis in ceramics, maintaining consistent performance across images with varying quality and microstructural features such as intra-grain topography, grain fractures, and blurred grain boundaries. Future efforts should focus on refining calibration strategies, extending the approach to multiphase microstructures, and evaluating the generalizability of the neural network to other, previously unseen ceramic materials.

CRediT authorship contribution statement

Stefan Jakobs: Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Bastien Le Gars Santoni:** Writing – review & editing, Validation, Resources. **Moritz Liesegang:** Writing – review & editing, Validation. **Marc Bohner:** Writing – review & editing, Validation, Supervision, Formal analysis.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used ChatGPT (version 4o) in order to support programming tasks described in Section 2.5 and to assist with language refinement during manuscript preparation. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This project was internally funded by RMS Foundation. The authors thank Pascal Michel for his support during sample preparation, Oliver Loeffel for assistance with SEM imaging, and Markus Schaudinn for his contributions to the initial application of the neural network to SEM images. The authors wish to thank two anonymous reviewers for their detailed and thoughtful comments that improved the quality of the manuscript, and Jérôme Chevalier for editorial handling.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.jeurceramsoc.2026.118431.

References

[1] T. Watanabe, S. Tsurekawa, The control of brittleness and development of desirable mechanical properties in polycrystalline systems by grain boundary engineering, Acta Mater. 47 (1999) 4171–4185.

[2] J. Wang, M. Rainforth, R. Stevens, The grain size dependence of the mechanical properties in TZP ceramics. Zirconia'88 Adv. Zirconia Sci. Technol, Springer, 1989, pp. 337–366.

[3] D. Stoyan, H.-D. Schnabel, Description of relations between spatial variability of microstructure and mechanical strength of alumina ceramics, Ceram. Int. 16 (1990) 11–18.

[4] E. Hall, The deformation and ageing of mild steel: III discussion of results, Proc. Phys. Soc. Sect. B 64 (1951) 747.

[5] ASTM E112-24 - Standard Test Methods for Determining Average Grain Size, West Conshohocken, PA (USA), 2024. ⟨https://doi.org/10.1520/E0112-24⟩.

[6] International Organization for Standardization, ISO 13383–1:2016 – Ceramic materials – Microstructural characterization – Part 1: Determination of grain size and size distribution, Geneva, Switzerland, 2016..

[7] A. Campbell, P. Murray, E. Yakushina, S. Marshall, W. Ion, New methods for automatic quantification of microstructural features using digital image processing, Mater. Des. 141 (2018) 395–406.

[8] H. Peregrina-Barreto, I. Terol-Villalobos, J. Rangel-Magdaleno, A. Herrera-Navarro, L. Morales-Hernández, F. Manríquez-Guerrero, Automatic grain size determination in microstructures using image processing, Measurement 46 (2013) 249–258.

[9] S. Banerjee, P.C. Chakraborti, S.K. Saha, An automated methodology for grain segmentation and grain size measurement from optical micrographs, Measurement 140 (2019) 142–150.

[10] B. Lu, M. Cui, Y. Wang, others, Automated grain boundary detection using the level set method, Comput. Geosci. 35 (2009) 267–275.
[11] M. Ananyev, D. Medvedev, A. Gavrilyuk, S. Mitri, A. Demin, V. Malkov, P. Tsiakaras, Cu and Gd co-doped BaCeO3 proton conductors: experimental vs SEM image algorithmic-segmentation results, Electrochim. Acta 125 (2014) 371–379.
[12] B.L. DeCost, H. Jain, A.D. Rollett, E.A. Holm, Computer vision and machine learning for autonomous characterization of AM powder feedstocks, JOM 69 (2017) 456–465, https://doi.org/10.1007/s11837-016-2226-1.
[13] B.L. DeCost, B. Lei, T. Francis, E.A. Holm, High throughput quantitative metallography for complex microstructures using deep learning: a case study in ultrahigh carbon steel, Microsc. Micro 25 (2019) 21–29, https://doi.org/10.1017/S1431927618015635.
[14] J. Na, J. Lee, S.-H. Kang, S.-J. Kim, S. Lee, Label-free grain segmentation for optical microscopy images via unsupervised image-to-image translation, Mater. Charact. 206 (2023) 113410, https://doi.org/10.1016/j.matchar.2023.113410.
[15] K. Pazdernik, N.L. LaHaye, C.M. Artman, Y. Zhu, Microstructural classification of unirradiated LiAlO2 pellets by deep learning methods, Comput. Mater. Sci. 181 (2020) 109728, https://doi.org/10.1016/j.commatsci.2020.109728.
[16] B.-I. Bachmann, M. Müller, D. Britz, A.R. Durmaz, M. Ackermann, O. Shchyglo, T. Staudt, F. Mücklich, Efficient reconstruction of prior austenite grains in steel from etched light optical micrographs using deep learning and annotations from correlative microscopy, Front. Mater. 9 (2022) 1033505, https://doi.org/10.3389/fmats.2022.1033505.
[17] P. Shi, M. Duan, L. Yang, W. Feng, L. Ding, L. Jiang, An improved U-Net image segmentation method and its application for metallic grain size statistics, Materials 15 (2022) 4417, https://doi.org/10.3390/ma15134417.
[18] X. Li, Y. Yuan, L. Chen, S. Guan, S. Feng, S. McMains, A novel approach to metallographic grain size measurement using deep learning and computational geometry techniques, Comput. -Aided Des. Appl. 23 (2026).
[19] A. Bordas, J. Zhang, J.C. Nino, Application of deep learning workflow for autonomous grain size analysis, Molecules 27 (2022) 4826.
[20] A. Chowdhury, E. Kautz, B. Yener, D. Lewis, Image driven machine learning methods for microstructure recognition, Comput. Mater. Sci. 123 (2016) 176–187, https://doi.org/10.1016/j.commatsci.2016.05.034.
[21] R. Poplin, A.V. Varadarajan, K. Blumer, Y. Liu, M.V. McConnell, G.S. Corrado, L. Peng, D.R. Webster, Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning, Nat. Biomed. Eng. 2 (2018) 158–164.
[22] O. Ronneberger, P. Fischer, T. Brox, U-Net: Convolutional Networks for Biomedical Image Segmentation, (2015). ⟨http://arxiv.org/abs/1505.04597⟩ (accessed September 30, 2024).
[23] M. Bohner, B.L.G. Santoni, N. Döbelin, β-tricalcium phosphate for bone substitution: synthesis and properties, Acta Biomater. 113 (2020) 23–41, https://doi.org/10.1016/j.actbio.2020.06.022.
[24] F.H. Perera, F.J. Martínez-Vázquez, P. Miranda, A.L. Ortiz, A. Pajares, Clarifying the effect of sintering conditions on the microstructure and mechanical properties of β-tricalcium phosphate, Ceram. Int. 36 (2010) 1929–1935, https://doi.org/10.1016/j.ceramint.2010.03.015.
[25] B. Le Gars Santoni, L. Niggli, S. Dolder, O. Loeffel, G.A. Sblendorio, R. Heuberger, Y. Maazouz, C. Stähli, N. Döbelin, P. Bowen, W. Hofstetter, M. Bohner, Effect of minor amounts of β-calcium pyrophosphate and hydroxyapatite on the physico-chemical properties and osteoclastic resorption of β-tricalcium phosphate cylinders, Bioact. Mater. 10 (2022) 222–235, https://doi.org/10.1016/j.bioactmat.2021.09.003.
[26] T.-Y. Lin, P. Goyal, R. Girshick, K. He, P. Dollár, Focal Loss for Dense Object Detection, (2018). ⟨http://arxiv.org/abs/1708.02002⟩ (accessed September 30, 2024).
[27] M. Gallo, B. Le Gars Santoni, T. Douillard, F. Zhang, L. Gremillard, S. Dolder, W. Hofstetter, S. Meille, M. Bohner, J. Chevalier, S. Tadier, Effect of grain orientation and magnesium doping on β-tricalcium phosphate resorption behavior, Acta Biomater. 89 (2019) 391–402, https://doi.org/10.1016/j.actbio.2019.02.045.
[28] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, et al., Scikit-learn, Mach. Learn. Python J. Mach. Learn. Res 12 (2011) 2825–2830.
[29] F. Chollet, et al., Keras, ⟨https://github.com/fchollet/keras-resources⟩, 2015. ⟨https://keras.io⟩.
[30] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, others, TensorFlow: a system for Large-Scale machine learning, 12th USENIX Symp. Oper. Syst. Des. Implement. OSDI 16 (2016) 265–283.
[31] S. Beucher, The watershed transformation applied to image segmentation, Scanning Microsc. (1992) 299–314.
[32] C.-L. Chiu, N. Clack, the napari community, napari: a python multi-dimensional image viewer platform for the research community, Microsc. Micro 28 (2022) 1576–1577, https://doi.org/10.1017/S1431927622006328.
[33] S. Van der Walt, J.L. Schönberger, J. Nunez-Iglesias, F. Boulogne, J.D. Warner, N. Yager, E. Gouillart, T. Yu, scikit-image: image processing in Python, PeerJ 2 (2014) e453.
[34] M.A. Rahman, Y. Wang, Optimizing intersection-over-union in deep neural networks for image segmentation, in: G. Bebis, R. Boyle, B. Parvin, D. Koracin, F. Porikli, S. Skaff, A. Entezari, J. Min, D. Iwai, A. Sadagic, C. Scheidegger, T. Isenberg (Eds.), Adv. Vis. Comput, Springer International Publishing, Cham, 2016, pp. 234–244, https://doi.org/10.1007/978-3-319-50835-1_22.
[35] P. Bajcsy, A. Cardone, J. Chalfoun, M. Halter, D. Juba, M. Kociolek, M. Majurski, A. Peskin, C. Simon, M. Simon, A. Vandecreme, M. Brady, Survey statistics of automated segmentations applied to optical imaging of mammalian cells, BMC Bioinforma. 16 (2015) 330, https://doi.org/10.1186/s12859-015-0762-2.
[36] OpenAI, ChatGPT, OpenAI, San Francisco, CA, USA, 2024. chat.openai.com..
[37] OriginLab Corporation, OriginPro, Version 2022, OriginLab Corporation, Northampton, MA, USA, 2022. ⟨https://www.originlab.com⟩.
[38] A. Berger, M. Herwegh, J.-O. Schwarz, B. Putlitz, Quantitative analysis of crystal/grain sizes and their distributions in 2D and 3D, J. Struct. Geol. 33 (2011) 1751–1763.
[39] CUDA C++ Programming Guide, (n.d.). ⟨https://docs.nvidia.com/cuda/cuda-c-programming-guide/⟩ (accessed October 2, 2024).
[40] A.K. Kakumani, L.P. Sree, B.V. Kumar, S.K. Rao, M. Garrepally, M. Chandrakanth, Segmentation of cell nuclei in microscopy images using modified ResUNet, in: 2022 IEEE 3rd Glob. Conf. Adv. Technol. GCAT, IEEE, 2022, pp. 1–6.
[41] I.-A. Taciuc, M. Dumitru, A. Marinescu, C. Serboiu, G. Musat, M. Gherghe, A. Costache, D. Vrinceanu, Enhancing malignant lymph node detection in ultrasound imaging: a comparison between the artificial intelligence accuracy, dice similarity coefficient and intersection over union, J. Mind Med. Sci. 12 (2025) 29.
[42] M.S. Neyestanak, H. Jahani, M. Khodarahmi, J. Zahiri, M. Hosseini, A. Fatoorchi, M.S. Yekaninejad, A quantitative comparison between focal loss and binary cross-entropy loss in brain tumor auto-segmentation using U-Net, J. Biostat. Epidemiol. 11 (2025) 15–35.
[43] M. Yeung, E. Sala, C.-B. Schönlieb, L. Rundo, A mixed focal loss function for handling class imbalanced medical image segmentation, ArXiv Prepr. ArXiv210204525 (2021).
[44] D.M. Powers, Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation, ArXiv Prepr. ArXiv201016061 (2020).
[45] S.M. Azimi, D. Britz, M. Engstler, M. Fritz, F. Mücklich, Advanced steel microstructural classification by deep learning methods, Sci. Rep. 8 (2018) 2128, https://doi.org/10.1038/s41598-018-20037-5.
[46] K. Mingard, P. Quested, M. Peck, Determination of grain size by EBSD-Report on a round robin measurement of equiaxed titanium, NPL Rep. MAT 56 (2012).
[47] B. Roebuck, C. Phatak, I. Birks-Agnew, A comparison of the linear intercept and equivalent circle methods for grain size measurement in WC/Co hardmetals, NPL Rep. MATCA 149 (2004).
[48] Y. Coutinho, S. Rooney, E. Payton, Analysis of EBSD grain size measurements using microstructure simulations and a customizable pattern matching library for grain perimeter estimation, Metall. Mater. Trans. A 48 (2017) 2375–2395.