Transactions of the SMiRT 27
Yokohama, Japan, March 3-8, 2024
Division XII

DEEP LEARNING BASED FRACTURE SURFACE SEGMENTATION FOR MICROSTRUCTURE FEATURE ANALYSIS IN DUCTILE CAST IRON

Johannes Rosenberger[1], Johannes Tlatlik[2], Sebastian Münstermann[3]

[1] M.Sc., Doctoral Researcher, Institute of Metal Forming, RWTH Aachen University, Aachen, Germany (johannes.rosenberger@iwm-extern.fraunhofer.de)
[2] Dr.-Ing., Research Scientist, Fraunhofer Institute for Mechanics of Materials, Freiburg, Germany
[3] Professor, Institute of Metal Forming, RWTH Aachen University, Aachen, Germany

ABSTRACT

Transport casks for spent nuclear fuel assemblies in Germany are largely made of ductile cast iron with nodular graphite. The microstructure of these materials affects their crack resistance and reliability, but assessing it manually based on micrographs is time-consuming and subject to variance. To address this issue, we developed a deep learning network that can extract local information from SEM images of fracture surfaces, using the open-source library Detectron2. Our study provides a framework for using deep learning models to analyze nodular cast iron fracture surfaces and extract microstructure parameters. For this purpose, the graphite particles in the fracture surface are detected, segmented and their properties are then statistically analyzed. The results are then compared to parameters obtained through micrograph preparation and manual analysis using techniques like thresholding.

INTRODUCTION

Prior to storage into a final long-term disposal facility, nuclear waste is stored in a dry state in special cast iron transport and storage containers. The storage time in these containers is supposed to be more than 40 years. Throughout this period, structural integrity of the containers must be guaranteed, especially considering transportation accidents and other impacts scenarios. According to current standards, the structural integrity is assessed in a very conservative deterministic manner, in which the most unfavorable material properties, loading conditions and defects are postulated and combined with each other. While this methodology is conservative and well established, it provides several disadvantages, i.e. no statistical variation of material properties is considered. The extracted microstructure parameters such as the graphite area fraction f_A, the particle size (diameter) d_G, particle distance λ, the shape factor f and or the number of particles per area N_A determine the material properties and can for example, be used to calculate material properties such as the fracture toughness (Pusch et al. (2016)). However, these parameters are derived from micrographs, which represent a 2D-simplification of a 3D-structure in a random cross section, only rarely revealing the true diameter of a particle. Also, while they do allow a general representation of the microstructure, local variation regarding the microstructure that was directly involved in fracture for this specific test cannot be obtained by this method. Besides the fact that an individual micrograph is an additional analysis, the micrograph preparation (grinding, etc.) will not allow the extraction of the desired information. In contrast, many particles are often completely exposed on the fracture surface, which can provide a more accurate particle quantification in 3D-space.

Therefore, we investigated the possibility of quantifying local microstructure parameters of fracture surfaces of SEM-images (scanning electron microscope) with the aid of artificial intelligence in order to compare the results with the results of conventional micrographs. An AI-assisted fracture surface analysis creates remedy with reference to the disadvantages discussed above, provided a solid model can be trained and verified. We implemented a Mask-RCNN instance segmentation model for fracture surface assessment.

Using deep-learning models for tasks in materials science is an emerging focus. For materials used in reactor technology, the successful evaluation and measurement of fracture mechanism assessment regarding macro images has been demonstrated earlier by Rosenberger et al. (2023). Due to its earlier success in detecting martensite-austenite islands in bainitic steel, Ackermann et al. (2022), we chose to use the open-source library Detectron2 by Wu et al. (2019) for the task of nodular cast iron fracture surface analysis. The fracture surface contains local information about the actual microstructure at the failure location rather than representative information from a micrograph analysis. Additionally, rather than performing manual micrograph analysis, the fracture surface segmentation is automated and the model generalizes well. The objective of the present contribution is to validate a deep-learning-based microstructure analysis of fracture surfaces for a future probabilistic safety assessment procedure that allows to determine a statistically quantifiable failure probability for cast iron. To validate our approach and highlight possible advantages of this type of analysis, the results are compared to manually determined microstructure parameters from micrographs.

METHODOLOGY

Material

The investigated material is nodular cast iron, whereby the material properties in the container vary considerably due to the manufacturing process. The material in the container wall is comparable to GJS-350-22, i.e., a ductile cast iron material. Similar cast iron materials with spheroidal graphite are used for containers for the storage and transport of nuclear waste. The specimens were tested as part of the research project "ProCast – Probabilistic safety assessment of cast iron containers." Hereafter referred to as ProCast (2023). All considered specimens in this study are Single Edge Notched Bending SE(B) 25x25 specimens with a precrack.

Data Acquisition, Pre-Processing and Augmentation

For imaging two different microscopes were used. The fracture surfaces have been captured with SEM, whereas the micrographs have been captured with a light microscope. For training 3 magnification levels are considered. An image at 100x magnification and resolution 1280x896 pixels can also be considered as a slice of an image at 50x magnification (if the image has a resolution of 2560x1792 pixels). Additionally, if an image at 100x is sliced into 4 patches of shape 640x448, each patch equals a magnification of 200x. At 100x magnification the used SEM has a measurement error of 2.6 %. The light microscope used, has a measurement error of 1.4 % at 25x magnification. All micrographs were captured at a resolution of 2560x1786 pixels. For labelling the open-source tool Labelme by Wada (2022) was used. Graphite particles were labelled as "graphite", the surrounding matrix was labelled as background, no matter the actual fracture mechanism. In total 24 images with a magnification of 100x were manually annotated. 12 fracture surface images resemble container wall material (denoted 1.7) and 12 images come from bottom material (2.7). In Figure 1 an overview over the different occurring microstructures and fracture mechanisms is given. In addition to the varying graphite particle parameters, the material differs greatly in perlite content. All specimens were tested at $5x10^3$ MPa√m/s (40 mm/s) and a temperature of -40 °C. The observed fracture mechanisms vary from cleavage (left) to mixed or mostly ductile (middle and right). Through patching the dataset could then be expanded to 120 images. The dataset was split in training, validation and test data with a split of 72-24-24 images (60-20-20 %). To improve the model's generalizability, augmentations were applied during training. In addition to horizontal and vertical flips, random brightness, contrast and saturation augmentations were applied.

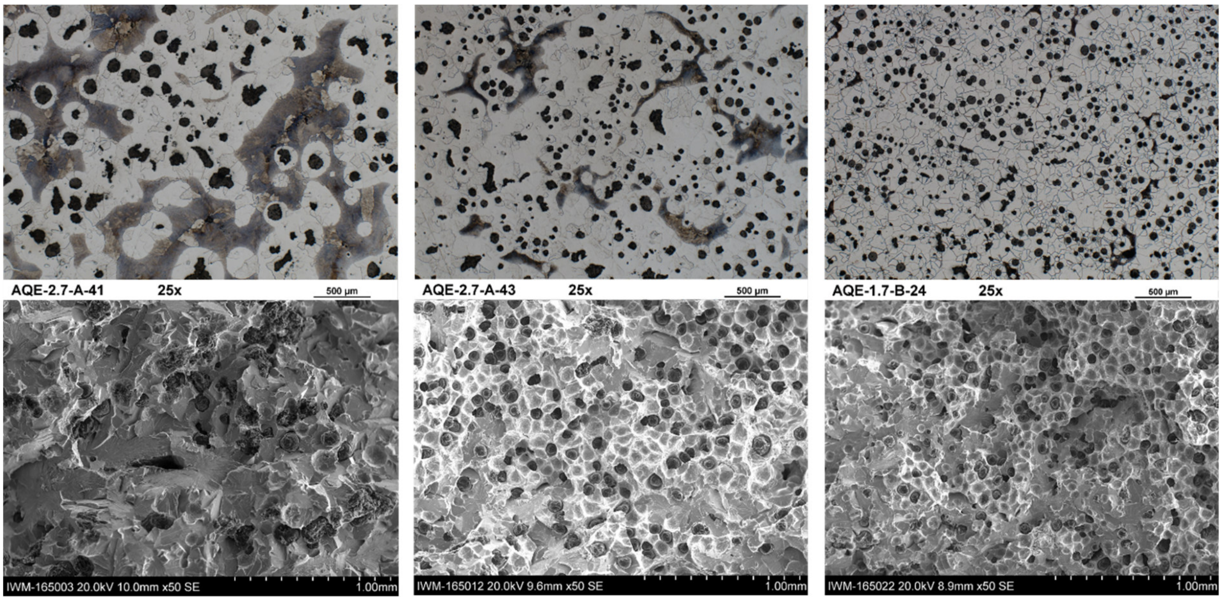

Figure 1. Exemplary overview of the various present microstructures and fracture mechanisms (all specimens tested at $5x10^3$ MPa√m/s and T = -40 °C).

Instance Segmentation Model

Detectron2 is an open-source computer vision library developed by Facebook AI Research (FAIR). It is the second iteration of the Detectron framework and is designed to facilitate the development and implementation of object detection algorithms. Detectron2 is built on top of the PyTorch deep learning framework and provides a flexible and modular structure for building, training, and deploying computer vision models. We used a pretrained Mask R-CNN (Region-based Convolutional Neural Network) model (He et al. (2017)) with a ResNet-50 backbone and a Feature Pyramid Network (FPN) as introduced by Lin et al. (2017) for feature extraction. The schematic architecture of the model is given in Figure 2. The model was trained for 369 epochs with the lowest evaluation loss achieved after 269 epochs.

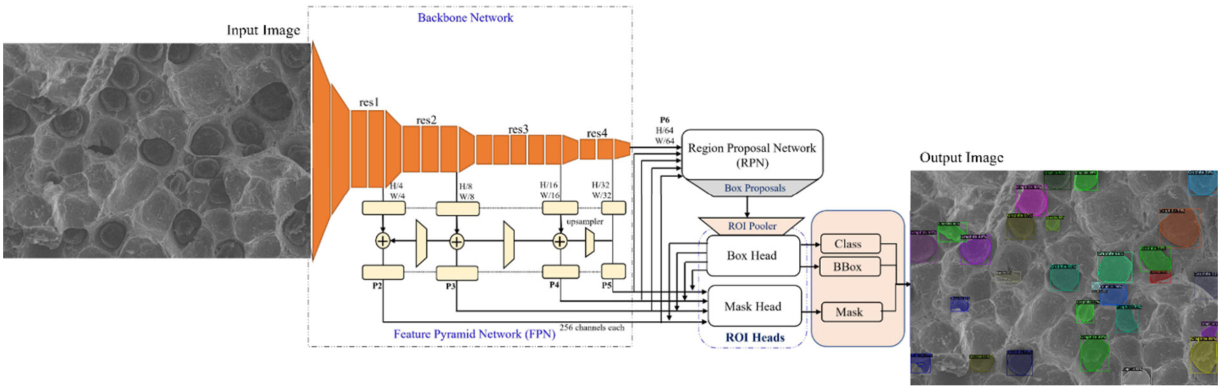

Figure 2. Schematic architecture of Detectron2 (redrawn figure based on Wu et al. (2019)).

Determination of Microstructure Parameters

Each micrograph was manually pre-processed in the image processing tool ImageJ by Schneider et al. (2012) using thresholding and watershed algorithms. Afterwards the particles could be measured within the software. The fracture surface images were segmented by the trained Mask-RCNN model and the resulting prediction masks were then automatically analyzed utilizing the Python package scikit-image. Both measurement methods have the same underlying parameter definitions and equations. The nodule diameter of the graphite particle d_G can be calculated as the mean of its major axis length a and its minor axis length b. To measure the mean particle distance, most commonly, the nearest neighbour distance λ_{nn} is used. The graphite content f_A is calculated as a percentage of the measured area, and N_A is the mean number of particles per measured area. According to Hütter et al. (2015) the nodularity of the graphite particles can be described as:

$$f = \frac{4\pi A}{P^2} \tag{1}$$

Where f is the shape factor (circularity) of the graphite particle, A is the area of a graphite particle and P is the perimeter. The pixel size can be used to transform the values to µm values. For the SEM images, each pixel equals 0.984 µm, whereas for the light microscope each pixel equals 1.359 µm.

RESULTS

Hyperparameter Tuning

Before the actual training process a hyperparameter optimization was carried out using the Python library Ray Tune by Liaw et al. (2018). We used an Asynchronous Successive Halving Algorithm (ASHA) to parallelize the search for optimal hyperparameters by running multiple trials concurrently, meaning the least promising trials are terminated early. The training has been optimized for the following hyperparameters: base learning rate (base lr), batch size, number of Regions of Interest (ROIs). Additionally, enabling an additional cropping augmentation was considered. The tuning showed that the base lr can be set rather high at 10^{-2}. The batch size and the ROIs did not seem to play the most important role in training success and the best results was achieved for a batch size of 8 and 512 ROIs. Disabling the cropping augmentation proved to be more feasible than enabling it.

Inference

Three randomly chosen examples for the results achieved with the trained model on the test data are shown in Figure 3. The images in the left column are the input images, as they are captured in the SEM. The middle column shows the manually labelled ground truth masks and on the right the prediction masks are displayed. From a qualitative point of view the model performance is very satisfying. The model shows good generalization regarding different lighting, contrast, as well as ductile and brittle fracture. Also, different particle sizes do not seem to influence the performance. For inference, two metrics are used. Precision describes the proportion of positive (graphite) identifications that were correct:

$$Precision = \frac{TP}{TP+FP} \tag{2}$$

Where TP are the true positive (actual graphite) pixels and FP are the false positive pixels (background predicted as graphite). The Recall metric describes the proportion of graphite pixels correctly identified:

$$Recall = \frac{TP}{TP+FN} \tag{3}$$

Where FN denotes the false negative pixels (graphite predicted as background). Additionally, the metrics are subdivided by the object size. Small objects (denoted by s) have an area smaller than 32^2 pixels, whereas large objects (l) are larger 96^2 pixels and medium objects (m) lie in between those two (32^2 < area < 96^2). In our test data (1807 particles total), 22.9 % of the objects are small, 71.8 % are medium sized and 5.3 % are considered large.

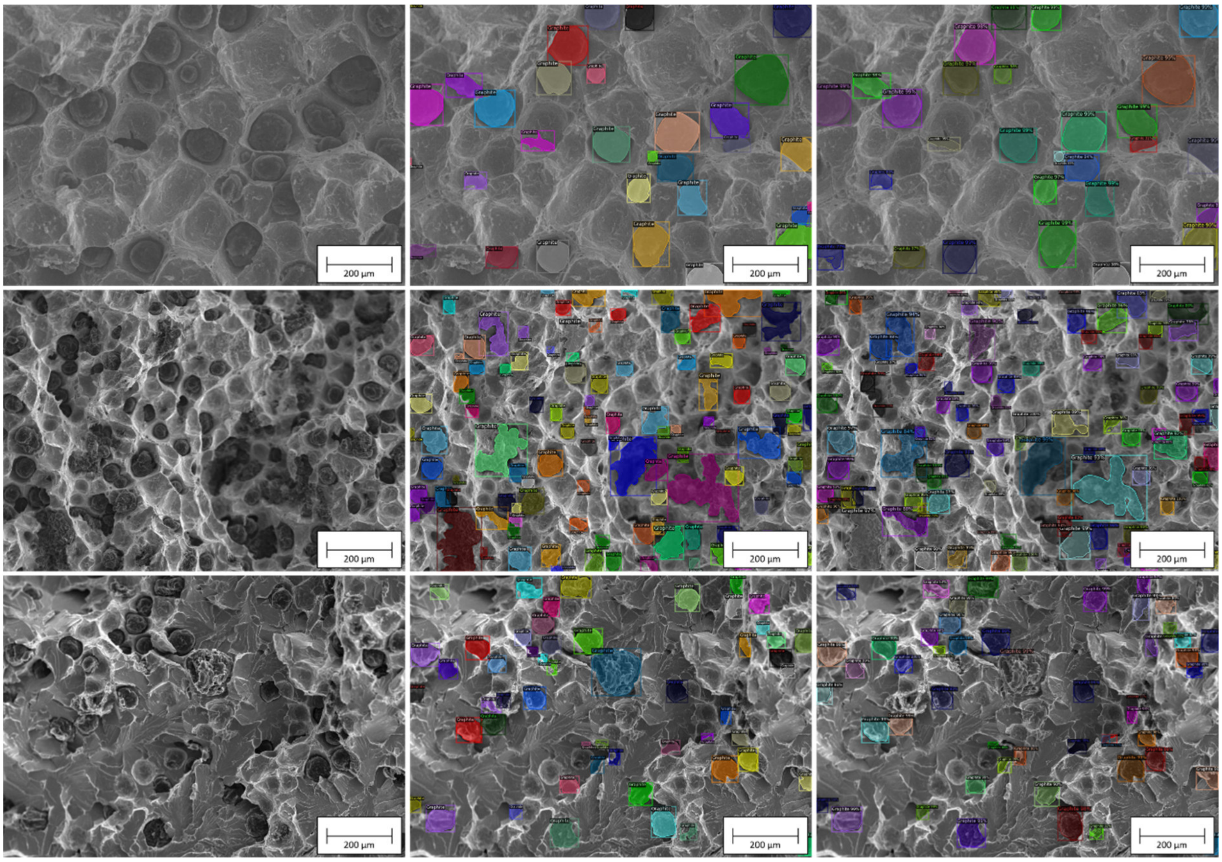

Figure 3. Prediction quality for randomly selected test images (left), ground truth (middle) and prediction (right) masks.

Table 1 shows that an average precision of around 24% could be achieved for the bounding box and the segmentation task. The recall is rated slightly higher at around 26 %. The bounding boxes define the spatial extent of objects within an image, while the segmentation masks are pixel-wise annotations that define the outline or boundary of objects within an image. The bounding box results show us how good the model is at detecting the object, while the segmentation represents the ability to predict the correct shape and class. All metrics are determined for a total of 10 intersection over union (IoU) thresholds (0.5 to 0.95), so that additionally to the actual segmentation performance we also get an insight on whether the model over- or underpredicts the graphite particles. The small AP and AR values can be explained by the fact that the ground truth bounding box that has the highest agreement with the predicted bounding box is used for comparison. On the one hand, if, for example, a particle is not detected as a whole but two particles are detected, both will still be compared with the original but much larger bounding box (Figure 4, example A). The number of false negatives increases accordingly and the recall value decreases. On the other hand, the prediction of a particle not present in the ground truth has an effect on the number of false positive identifications. Accordingly, the precision value will decrease.

Table 1: Evaluation results in terms of average precision (AP) and average recall (AR) for small (s), medium (m) and large (l) objects.

Evaluation	AP	AP_s	AP_m	AP_l	AR	AR_s	AR_m	AR_l
Bounding Box	24.0	6.5	67.5	58.5	25.9	6.7	72.4	73.0
Segmentation	24.6	6.1	69.3	63.5	26.3	6.7	74.1	71.5

Apparently the medium and largely sized graphite particles are by far better detected, than the small particles. This can be explained with single pixel misclassifications having a higher impact in small objects, as well as missing or overpredicting entire particles is more likely for small objects. Since graphite particles with a measured diameter of less than 10 µm are not considered in the following microstructure analysis, no further optimization regarding this metric was performed. For all graphite particle sizes, the recall values are higher than the precision. This suggests that the model is sensitive to capturing positive instances (minimizing false negatives i.e., the misclassification of graphite as background) but may be less precise in distinguishing between true positives and false positives. In other words, the model might tend to produce more graphite predictions, and some of them might be incorrect. It's often a trade-off between precision and recall, and the optimal balance depends on the specific requirements of the application. In our case a higher recall is favorable, since missing out on whole graphite nodules would influence the microstructure parameters (f_A, λ, N_A) more than the misclassification of a few pixels slightly influencing d_G and f.

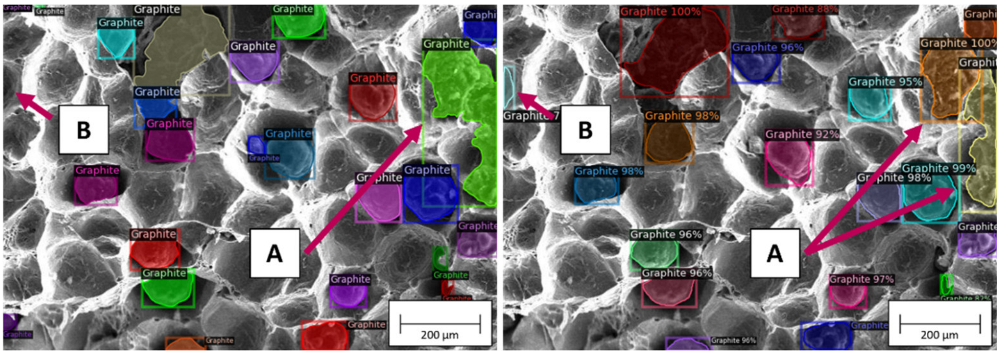

Figure 4. Examples for model predictions (right) decreasing AR (A) and AP (B), when compared to ground truth (left).

The results presented above show that the instance segmentation of graphite particles is possible and a good generalization can be achieved. Medium and large size particles are detected with good precision, while the small graphite particles are not always detected well. The subsequent microstructure analysis results are discussed in the following paragraph.

Microstructure Parameters

For the analysis of the microstructure parameters 10 specimens were investigated. The research project ProCast (2023) showed that the microstructure varies significantly within the container. Therefore, 5 specimens are taken from the container wall (denoted as 1.7) and 5 from the bottom (2.7). For the analysis of the microstructure parameters on the fracture surfaces, 16 images at a 100x magnification are considered for each specimen. To validate the results, the parameters found are compared to manually determined values from actual micrographs of the same specimens. For the micrographs the number of light microscope images per specimen varies from 3 to 5 at 25x magnification. Graphite particles that share pixels with the image border are not considered in both the SEM and the light microscope images. As mentioned, detected graphite particles with a measured diameter of less than 10 µm are not considered in the following analysis since they can be considered artifacts.

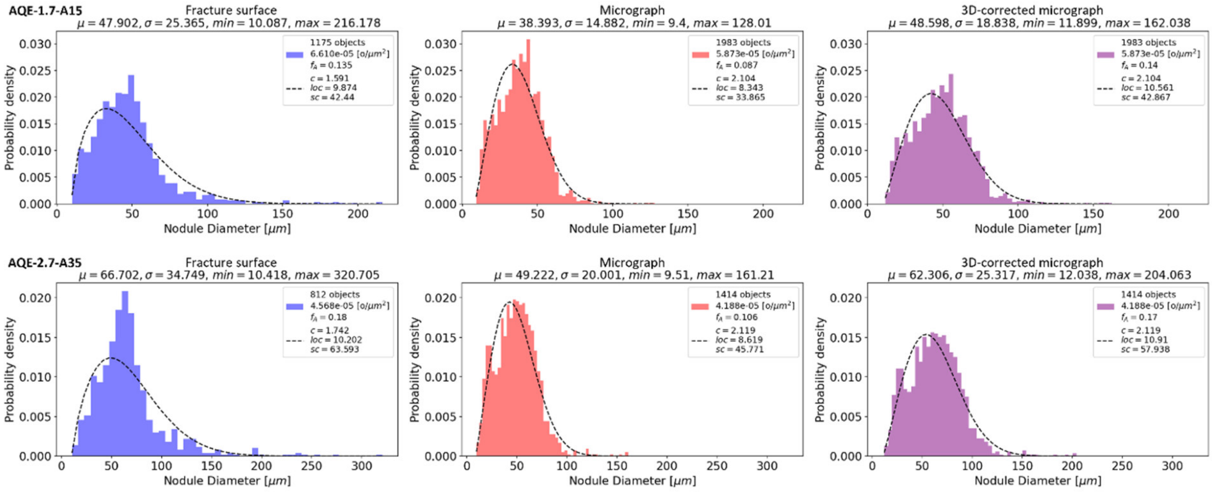

Figure 5. Exemplary probability (Weibull) distributions for the nodule diameter obtained from the fracture surface (left), micrograph (middle) and 3D-correction (right).

Figure 5 shows exemplary results for the nodule diameter measurements for a wall and a bottom specimen. For both specimens the fracture surface results obtained with the Mask-RCNN model (left) differ from the results obtained in the micrographs (middle). In particular, the fracture surface analysis produces wider distributions which are shifted towards higher diameter sizes. Moreover, the method detects significantly more particles of very high diameter (100 µm). By applying a 3D-correction to the measured particles sizes in the 2D micrographs, according to Williams et al. (2008), the histograms shift to the right and the standard deviation is increased. This is in closer agreement with the fracture surface results. However, the mentioned particles of very high diameter are still not present in the results. The correction itself is a rather simple multiplication of the measured diameters with a factor of 0.79 as the mean measured particle diameter is 79 % of the actual particle diameter. Additionally, to obtain a corrected area fraction f_A, the correction is also applied to the measured particle area A by multiplying it with 0.79^2. In Table 2 the mean values for the measured microstructure parameters are given for the container wall and bottom. The values measured in fracture surfaces are in close agreement to the values obtained by the 3D-correction of the 2D micrographs. The graphite area fraction f_A values range from 14.9 to 16 %. For all three methods, the circularity f of the graphite particles are in close agreement. The number of measured particles per area N_A is slightly increased for the fracture surface. For the wall material, the mean nodule diameter d_G measured in the fracture surface is larger than in the micrograph. Compared to the micrographs, the fracture surface segmentation considers bigger particles. In Figure 5 the biggest measured nodule diameters are 216 and 320 µm, while for the 3D-corrected macrograph its only 162 and 204 µm. The standard deviation shows that the graphite particle size is more uniform in the fracture surface. On the fracture surface the particles are rather uncovered (visible as a whole) than being cut in plane. In microstructures with less nodularity (bottom material), this effect is not observable. The mean nearest neighbour distance measured in the fracture surfaces is higher than in the micrographs. In general, the microstructure parameter determination via fracture surfaces utilizing instance segmentation works well.

Table 2: Mean values (and standard deviation) of microstructure parameters measured by different methods: 1.7 wall material, 2.7 bottom material.

Method	Type [-]	f_A [%]	N_A [mm^{-2}]	d_G [µm]	λ_{nn} [µm]	f [-]	$\lambda_{nn} * f$ [µm]
Micrograph	1.7	9.3 (0.4)	75 (15.0)	35 (3.8)	65 (6.9)	0.804 (0.005)	52.6
	2.7	10.0 (0.6)	27 (9.9)	62 (11.8)	103 (14.5)	0.734 (0.050)	70.7
3D-corrected Micrograph	1.7	14.9 (0.5)	75 (15.0)	44 (4.6)	65 (6.9)	0.811 (0.007)	53.0
	2.7	16.0 (1.0)	27 (9.9)	79 (14.1)	103 (14.5)	0.730 (0.045)	74.3
Fracture Surface	1.7	16.0 (2.0)	79 (12.3)	48 (2.5)	71 (4.6)	0.793 (0.013)	55.9
	2.7	16.2 (2.6)	34 (11.1)	75 (12.7)	113 (22.5)	0.731 (0.046)	81.3

To further validate the quality of the model, we can compare the nearest neighbour distance to a mean particle distance calculated according to Hütter et al. (2013):

$$\lambda = (f_0/\alpha_V)^{1/6}/\sqrt{N_A} \tag{4}$$

Where the graphite volume fraction f_0 is equal to the is the graphite content as a percentage of the measured area f_A, α_V is a correction factor equal to 1 for Rayleigh distributions. If the particles are regularly arranged the ratio between the mean distance according to Equation (4) and the nearest neighbour distance λ_{nn}/λ should be equal to 1.

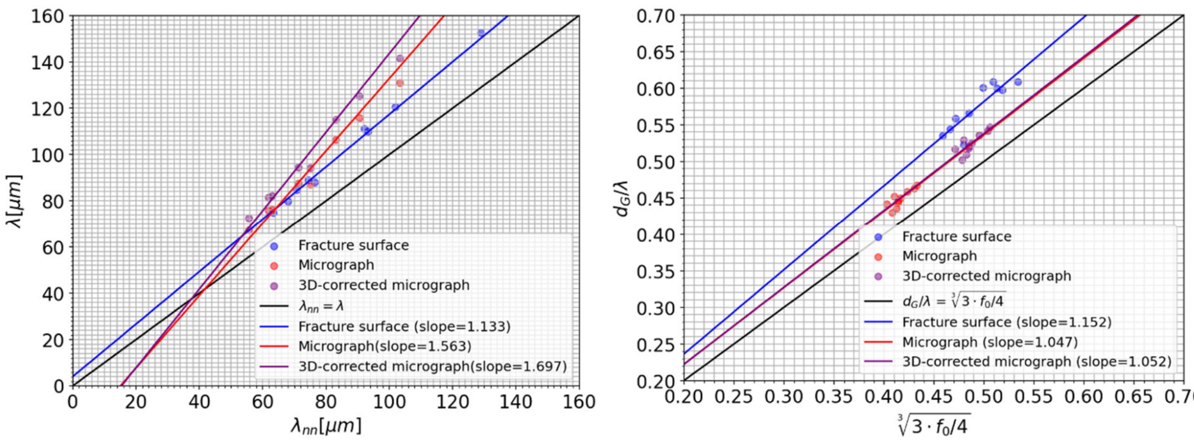

Figure 6. Ratio of mean distance according to Equation (4) and the nearest neighbour distance (left). Ratio of graphite particle size and distance vs. graphite volume fraction (right).

Figure 6 (left) shows this ratio is met better for the parameters measured in the SEM fracture surface image and that the difference is increasing with a growing nearest neighbour distance. Upon that, we investigated whether a fixed graphite volume fraction of $f_0 \approx 0.07 \dots 0.15$ scales to the graphite particle distance and the graphite particle size as Hütter et al. (2015) suggested:

$$\frac{d_G}{\lambda} = \sqrt[3]{\frac{3f_0}{4}} \approx 0.37 \dots 0.48 \tag{5}$$

Figure 6 (right) suggests, that this criterion is only met for the uncorrected micrograph data, where d_G/λ is slightly higher than the volume fraction term. As the plot shows, the 3D-correction shifts the values to a higher range up to 0.55 without having the same slope and offset. For the fracture surface data, the intercept for d_G/λ is increased and the mismatch is increasing with higher volume fractions.

DISCUSSION

The results of the microstructure parameter determination on fracture surfaces in ductile cast iron are promising. Nonetheless, various influences need to be investigated further. The fracture surfaces only contain half of the graphite particles. However, since it is assumed that these are random influences, the statistical analysis of the particles is not biased. The results show a slightly increased number of graphite particles per area N_A although we must consider that 50 % of the graphite particles are missing (located on the counterpart). This could be explainable since the crack does not propagate in plane and therefore the actual surface area of the fracture surface is bigger than the plane micrograph and more particles are uncovered, which would also correlate with the fracture mechanism. Moreover, it is very likely that cleavage generally takes place with less nodule interaction, since ferrite provides less resistance here, creating fewer visible nodules on the fracture surface. Figure 7 also shows that for cleavage (bottom right) fewer particles are uncovered as for ductile fracture (bottom left). Additionally, the fracture surface analysis reveals big and irregularly shaped particles as a whole and the instance segmentation can clearly differ them from particle clusters. This is not always possible in the micrograph analysis. Now that these particles are not randomly and also fully detected in a micrograph, the method could potentially be used to identify such critical local microstructure parameters more reliably.

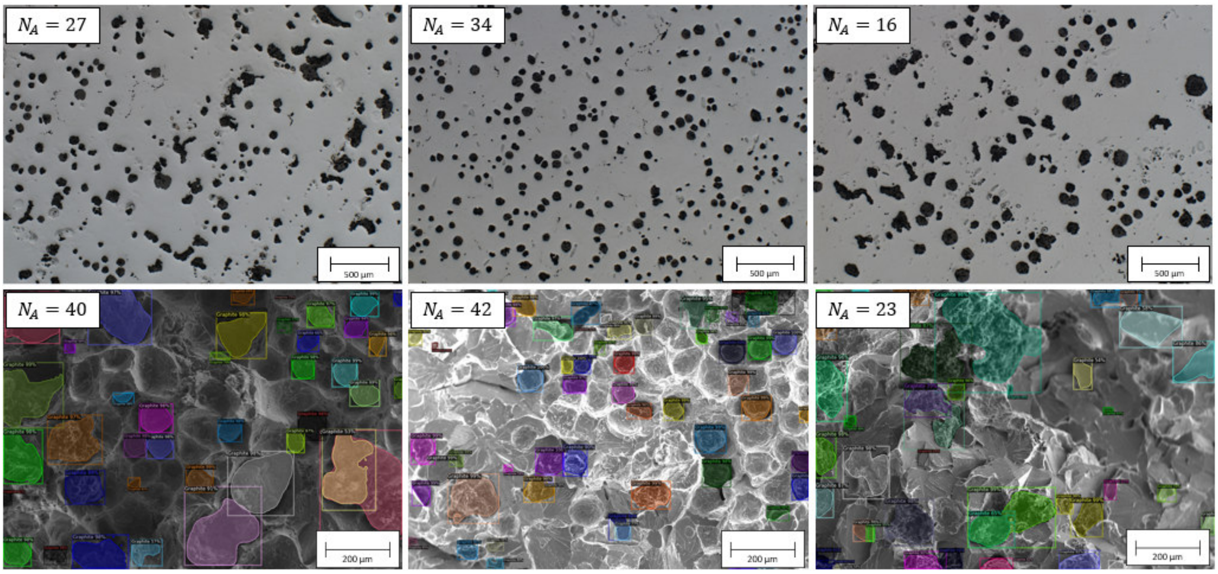

Figure 7. Differences in N_A in different fracture mechanisms compared to micrographs (all specimens are bottom material and tested at 5x10^3 MPa√m/s, left at T = 20 °C, middle and right at T = -40 °C).

CONCLUSIONS

In conclusion, this study demonstrates the successful measurement of graphite particle parameters without the need for manual micrograph preparation and analysis but from analysis of fracture surfaces through instance segmentation. The results indicate that the segmentation of graphite particles, particularly medium and large-sized ones, can be achieved with high precision using the Mask-RCNN model. Despite challenges in detecting small particles, the overall performance of microstructure analysis utilizing Detectron2 proves effective. The agreement between mean values and standard deviations of parameters measured in micrographs and those determined from fracture surfaces is notable. Moving forward, it will be interesting to see how the enhanced visibility of entire particles on fracture surfaces, as opposed to partial cuts in plane in micrographs, present a distinct advantage, especially for probabilistic safety assessments. The obtained data will be used in simulation and can be compared to results achieved with micrograph data, further contributing to the field of cast iron container safety assessment.

AKNOWLEDGEMENT

The research project "KEK – Automated analysis of fracture surfaces using artificial neural networks (KNN) for nuclear-relevant safety components" is funded by the German Federal Ministry for the Environment, Nature Conservation, Nuclear Safety and Consumer Protection (Project No. 1501621) on basis of a decision by the German Bundestag.

REFERENCES

Ackermann M., Iren D., Wesselmecking S., Shetty D., Krupp U. (2022). "Automated segmentation of martensite-austenite islands in bainitic steel," *Materials Characterization*, Elsevier, UK, 191 112091.

He K., Gkioxari G., Dollár P., Girshick R. (2017). "Mask R-CNN," *2017 IEEE International Conference on Computer Vision (ICCV)*, Venice, Italy.

Hütter G., Zybell L., Mühlich U., Kuna M. (2013). "Consistent simulation of ductile crack propagation with discrete 3D voids," *Computational Materials Science*, Elsevier, UK, 80 61–70.

Hütter G., Zybell L., Kuna M. (2015). „Micromechanisms of fracture in nodular cast iron: From experimental findings towards modeling strategies – A review," *Engineering Fracture Mechanics*, Elsevier, UK, 144 118-41.

Liaw R., Liang E., Nishihara R., Moritz P., Gonzalez J.E., Stoica I. (2018). "Tune: A Research Platform for Distributed Model Selection and Training," *ICML 2018 AutoML Workshop*, Stockholm, Sweden.

Lin T.-Y., Dollár P., Girshick R., He K., Hariharan B., Belongie S. (2017). "Feature Pyramid Networks for Object Detection," *2017 IEEE International Conference on Computer Vision (ICCV)*, Venice, Italy.

ProCast (2023). "Probabilistic safety assessment of cast iron containers," *Research project funded by the Federal Ministry for Environment, Nature Conservation, Nuclear Safety and Consumer Protection (Project No. 1501576).*

Pusch G., Henkel S., Biermann H., Hübner P., Ludwig A. (2016) "Determination of Fracture Mechanics Parameters for Cast Iron Materials Under Static, Dynamic and Cyclic Loading," *Recent Trends in Fracture and Damage Mechanics*, Springer International Publishing, Switzerland, 159-96.

Rosenberger J., Tlatlik J., Münstermann S. (2023) "Deep learning based initial crack size measurements utilizing macroscale fracture surface segmentation," *Engineering Fracture Mechanics*, Elsevier, UK, 293 109686.

Schneider, C. A., Rasband, W. S., & Eliceiri, K. W. (2012). "NIH Image to ImageJ: 25 years of image analysis," *Nature Methods*, 9(7) 671-675.

Wada (2023). "Labelme: Image Polygonal Annotation with Python," *GitHub Repository*, Available at https://github.com/wkentaro/labelme.

Williams A.M., Garner C.P., Binner J.G. (2008). "Measuring Pore Diameter Distribution of Gelcast Ceramic Foams from Two-Dimensional Cross Sections," *Journal of the American Ceramic Society*, John Wiley & Sons, USA, 91(9) 3113-3116.

Wu, Y., Kirillov, A., Massa, F., Lo, W.-Y., & Girshick, R. (2019). "Detectron2," *GitHub Repository*, Available at https://github.com/facebookresearch/detectron2.