 DOI: 10.3217/978-3-85125-869-1-07

A study on robust feature representations for grain density estimates in austenitic steel

Filip Ilic[1], Marc Masana[1,2], Lea Bogensperger[1], Harald Ganster[3] and Thomas Pock[1]

***Abstract*— Modern material sciences and manufacturing techniques allow us to create alloys that help shape our way of living; from jet turbines that withstand extreme stresses to railroad tracks that retain their intended shape. It is therefore an important aspect of quality control to estimate the microstructural properties of steel during and after the manufacturing process, as these microstructures determine the mechanical properties of steel. This estimation has for a long time been a labor intensive and non-trivial task which requires years of expertise.**

We show that modern deep neural networks can be used to estimate the grain density of austenitic steel, while also applying a visualization technique adapted to our task to allow for the visual inspection of why certain decisions were made. We compare classification and regression models for this specific task, and show that the learned feature representations are vastly different, which might have implications for other tasks that can be solved via discretization into a classification problem or treating it as an estimation of a continuous variable.

I. INTRODUCTION

Not all steel is created equally. Other than the ratio of carbon and other metals that are used in the alloy when it is being forged to steel, different modes of cooling, heating, and hardening produce variations in steel. Broadly speaking, steel can be classified into austenite, martensite, and under certain circumstances even a mixture of both. Martensite forms when steel is quenched very quickly, whereas austenite forms through a lengthy cooling process. Even within the austenite cooling process, there are many factors that influence the development of microstructures within the steel that contribute to the graining process, i.e. the formation of individual grains. Determining the characteristic grain size of the sample, which is used to determine the grain density, is important for many applications as it relates to the tensile and compressive stresses that the material is able to withstand. These grains and other microstructures of the resulting steel can - through an extensive etching and cleaning process - be made visible under a light microscope [10].

Traditionally, austenitic steel grain density is estimated by costly and labour intensive work done by a metalographer where etched steel samples are manually inspected under a light microscope. Currently the most reliable way to perform this grain density estimate is by including a template, that is projected onto the viewfinder of the microscope. The metalographer then uses this template to determine the grain density by comparing it to the different available templates. Since only a 2-dimensional cross-section of the 3-dimensional material is visible, grains might appear smaller or larger than the average grain size within the material due to the slicing process. It is therefore a requirement that a judgement is made based on the relative distributions of grain density within a single slice of the sample.

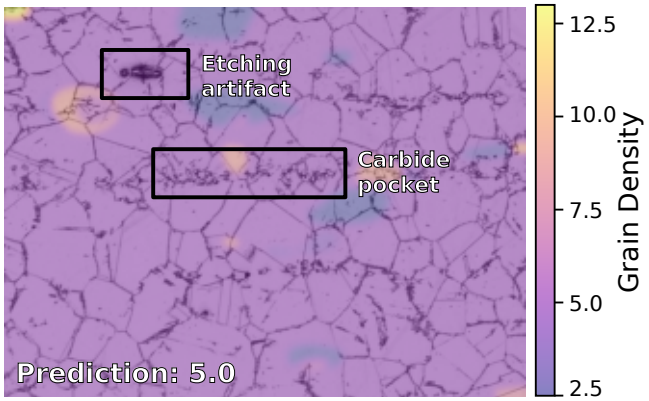


Fig. 1. A single image of an austenitic steel sample taken with a light microscope at 100-fold magnification. Our method estimated the overall density of the sample to be 5.0. The overlayed heatmap is created with a proposed visualization scheme, detailed in Section VI, that aids in understanding the decision process of the network, as a pixel-wise density estimate can show inhomogeneous regions if they are present. Note that the learned representation is robust enough to ignore sample preparation artifacts and carbide pockets that have a similar appearance to higher density grain regions.

In this paper, we propose a deep learning-based approach to estimate austenitic steel grain density from a single image. We explore how classical cross-entropy-based losses allow to learn classification models with state-of-the-art performance. However, we find that classification models - at least in the domain of grain density estimation - come at a price when comparing it to similar, albeit slightly less performant regression models, that show more resilience when dealing with out of distribution samples, and appear to have a more robust and human interpretable feature space. We therefore also propose to use regression-based losses that are capable of predicting a continuous grain density, at the cost of a slight decrease in performance.

It is notoriously difficult to explain the decision making process of deep neural networks, which can often be a source of confusion when applied and deployed in real world applications. It is therefore important, to provide tools to visualize the model's decisions, and understand the failure cases and the reasons for a failure. Recently, some methods

[1] Graz University of Technology, Austria
[2] Silicon Austria Labs, TU-Graz SAL DES Lab, Austria
[3] JOANNEUM RESEARCH Forschungsgesellschaft mbH, Austria
Corresponding Author: filip.ilic@icg.tugraz.at

have been proposed to evaluate the confidence by using class activation maps [32]. We adapt a well known algorithm – GradCAM [27] – that allows us to visualize local regions within a single sample, that can a) show us glimpses of the underlying feature embeddings and whether they really encode relevant information that trained metalographers would look for, and b) give more fine grained analysis of the input image than just the classification label, shown in Fig. 1.

A challenge of using deep learning models in this domain is that deep neural networks require large amounts of data to reach a satisfying degree of robustness. Because the data acquisition and labeling of steel samples requires thorough metalographic knowledge, this data is rather scarce. Therefore, we propose a heavy data augmentation scheme that allows to generate grain densities of continuous granularity, even when only whole grain (i.e. 4.0, 5.0, etc.) austenite data is available, as it often is.

In summary, our contributions are the application of classification and regression based deep learning models to the domain of microstructural analysis of austenite steel, with a focus on the differences in interpretability of the resulting feature representation that the two modes of learning yield. Furthermore, we propose a data augmentation scheme which could be extended to other datasets that are fractal-like or display self-similarity. We show that its usage improves performance across a variety of different models. We present through ablations on classification and regression models that in general classifiers perform better than regressors in this setting. However, this improved performance comes at the cost of a decrease in robustness and interpretability.

II. RELATED WORK

In the past, many insights into material composition and corresponding material properties were derived from expert knowledge and experience. Nowadays, data generated by simulations and measurement systems are becoming more available, thus moving away from physically-based tests. Agrawal and Choudhary [1] introduce the term *deep materials informatics* in the context of data-driven technologies and provide a comprehensive overview of challenges and applications of deep learning with respect to learning chemical compositions of materials, prediction of crystalline structures, (3D) microstructure analysis, and microstructure reconstruction [6]. Furthermore, [12] illustrate opportunities and current paths, where machine learning will have significant influence on material science.

Automated detection or classification of microstructures is the central theme of metallographic studies. Chowdhury et al [7] use image analysis and machine learning to discriminate whether samples have dendritic morphologies or not. DeCost and Holm [13] use a feature-based approach to identify generic signatures of microstructures. These serve as the basis for a Support Vector Machine (SVM) [8] classifier to distinguish 7 microstructure classes. Similarly, Gola et al [17] employ an SVM model for reproducible and objective microstructure classification and achieve classification accuracy greater than 90% for cast iron samples. The morphological data comes from both optical microscopy and electron microscopy images, and the mixed microstructure exhibits a variety of graphite morphologies. An extension of the classification system to deep learning techniques achieved 95% accuracy on unprocessed electron micrographs of low-alloy steels [3], [25]. Here, a combination of CIFARNet, a modification of LeNet [22], and a pretrained VGG16 network [28] were used. Mulewicz et al. [23], [24] distinguish 8 classes of microstructures of different steel grades (C15, C45, C60, C80, V33, X70, and non-hardened steel) from optical microscopy images with the aid of a deep network structure based on ResNet18 [18]. The authors of [15] train models with U-Net architectures with about 30-50 micrograph samples in order to achieve robust segmentation for bainite microstructures. To segment microstructures into four relevant domains ("grain boundary carbide, spheroidized particle matrix, particle-free grain boundary denuded zone, and Widmanstätten cementite"), DeCost et al [11] use pixel-based machine learning [4]. Their segmentation model was compared to the results of microscopic annotation by metalographer using 24 carbon steel samples. Although direct comparison in microstructures (< 5 pixels) was not possible and demonstrated the need for high quality training data, it was still possible to show the effectiveness of deep learning in the analysis of complex microstructures. Albuquerque et al. [9] apply a multilayer perceptron with backpropagation to achieve a microstructure segmentation for cast iron images. Verification on a test set of 60 images showed high correlation to human ground truth. In this line Bulgarevich et al. [5] apply a Random Forest classifier to optical microscopic images of steels for an automated segmentation. Austenite grain density is a significant variable in the AI system of Kuziak [21], which allows the estimation of different phase constituents occurring during the cooling process.

III. DATASET

To perform a density analysis the grains within the steel need to be made visible. Various types of acids are used which etch the weak spots of the metal surface, i.e. the grain boundaries or other impurities of the metal, away first, leaving behind a darkened appearance. The prevailing industry standard to measure the grain density within the material is the ASTM e112 [2] norm. It specifies a 100-fold magnification at which the optical microscope images are captured. Therefore all our images are taken with a 100-fold magnification, and in total consist of 242 images that have a resolution of 1280×960. We split them into 125 train, 53 validation, and 64 test images, keeping the distribution of classes balanced.

The dataset contains images from whole-grade densities ranging from 4.0 to 13.0 with increments of 1.0, and additionally the grain density 2.5. This range of grain densities are provided by the manufacturing process at the steel mill. Fig. 2 shows austenitic steel with various grain densities; it also shows the variation in appearance that is due to the different alloys, and variations in the etching process.

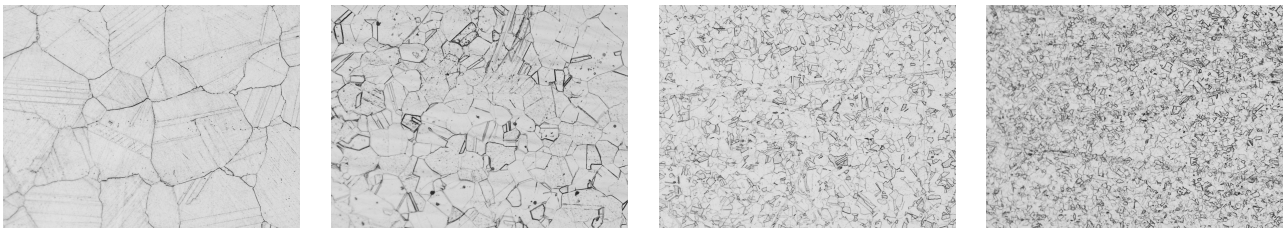


Fig. 2. Samples of varying grain densities. The grain density increases from left to right and exhibits fractal like self similarity at different scales. It does however produce a variety of artifacts due to the etching process depending on size of individual grains and variation among samples. While the grain density is fundamentally a continuum, it is often discretized to whole- or half-grades in practice.

IV. GRAIN DENSITY ESTIMATION

The problem of grain density estimation can be framed as an image classification task, with the increased complexity that naturally occurring grain density variations might not be homogeneously distributed across the entire sample. Image classification has seen a massive shift from hand-crafted feature detectors towards the use of different deep learning techniques. When data is limited, a common technique is to pretrain on a large dataset and only fine-tune the network to the specific domain. This exploits a good initialization of the network parameters to learn an adjusted representation of that smaller domain [26]. Since the amount of annotated data which contains information about microstructures, including grain density, is usually limited due to its acquisition cost, we propose to use fine-tuning with a cross-entropy loss on a pretrained classification network. To the best of our knowledge, this popular technique has not been applied to this setting. The closest work which uses deep learning for microstructural analysis in steel is [3]. Although this work is not applied to austenitic steel, nor evaluated with grain density estimates, we consider it in our comparisons.

Classification allows for the network to learn representations which project the input data into a feature space where different classes can be easily discriminated without any specific ordering. However, due to the nature of steel grains spanning a continuum of different sizes, we also consider to frame the grain density estimation as a regression problem. This allows the network to not only discriminate between different classes, but also maps to a feature space that implicitly preserves grain density order.

Classification. We consider a backbone pretrained feature extractor Φ parameterized by weights θ_Φ and a classifier Ψ parameterized by weights θ_Ψ. We define $\mathbf{o}(\mathbf{x}) = \Psi(\Phi(\mathbf{x};\theta_\Phi);\theta_\Psi)$ as the output logits of the network given an image $\mathbf{x}$. Then, given $\mathbf{y}$ as the one-hot encoding of the ground truth label corresponding to the N classes (grain densities), we consider the cross-entropy loss

$$\mathscr{L}_{\mathrm{CE}}(\mathbf{x},\mathbf{y};\theta_\Phi,\theta_\Psi) = \sum_{k=1}^{N} y_k \log \frac{\exp(\mathbf{o}_k)}{\sum_{i=1}^{N}\exp(\mathbf{o}_i)}. \qquad (1)$$

Regression. We use the same feature extractor and head as in classification, together with output logit $\mathbf{o}(\mathbf{x})$ given an image $\mathbf{x}$. However, given $\mathbf{y}$ as the actual numerical value of the ground truth grain density, and $\mathbf{d} = \mathbf{o}(\mathbf{x}) - \mathbf{y}$, we define the regression loss as a smooth ℓ_1 loss

$$\mathscr{L}_{\mathrm{S1}}(\mathbf{x},\mathbf{y};\theta_\Phi,\theta_\Psi) = \begin{cases} \frac{\mathbf{d}^2}{2\alpha}, & \text{if } |\mathbf{d}| < \alpha \\ |\mathbf{d}| - \frac{\alpha}{2}, & \text{otherwise} \end{cases}, \qquad (2)$$

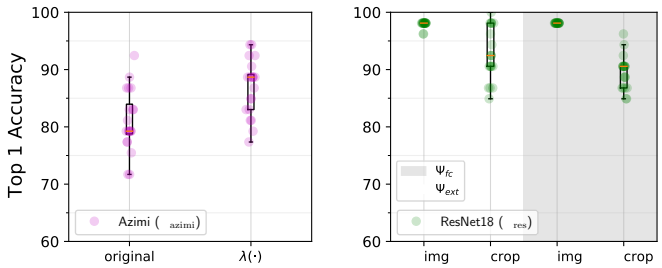
where $\alpha = 1$. This threshold α specifies when the loss function changes between ℓ_1 and ℓ_2^2. This loss is less sensitive to outliers, than the mean squared error and can help prevent exploding gradients [16].

Data augmentation. As stated earlier, austenitic steel data for microstructural analysis is costly to acquire and difficult to annotate correctly. This leads to generally small datasets, which can be an issue for deep learning models. However, apart from fine-tuning on pretrained models, another popular training strategy is data augmentation, which consists of altering and extending samples from the dataset with class preserving transformations. The transformed samples increase the number of images to be learned from and help the model generalize better, and to have a more robust representation of the target domain.

The grain density G is determined by $N = 2^{G-1}$, where N is the number of grains per square inch at $100\times$ magnification. The different grain densities exhibit similar structures and patterns at different scales with self-similar features. Therefore, various magnifications of samples with their corresponding adapted labels can be generated from image patches to simulate larger or smaller grain densities by cropping and resizing them in accordance with the grain density formula. Our proposed data augmentation strategy consists of generating new samples which differ at a maximum of ± 0.5 grades from the original. In the case of classification this is set to a binary ± 1.0 to align with our class labels. In addition we perform the common data augmentation best practices: random rotations between 0°and 360°, horizontal and vertical flips, and contrast jitter to simulate possible changes in the lighting conditions during data acquisition or variations in the etching strength during sample preparation. In the experimental sections we will denote the additional rescaling during data augmentation as $\lambda(\cdot)$, and apply the rest of mentioned transformations to all reported experiments.

Image and crop augmentation. Regression or classification can be performed by passing the whole image or crops of a fixed size to the model. Our proposed data augmentation

Fig. 3. Two different proposed heads for image classification: Ψ_{fc} is defined as a single fully-connected layer, whereas Ψ_{ext} is defined with an extension of 2 more intermediate fully-connected layers. When doing regression, the last layer is replaced by a single output.



a) Azimi et al. [3] b) Ours

Fig. 4. Grain density estimation with classification. We compare Azimi et al. [3] (purple) and our proposed ResNet18-based architecture (green) with various configurations. We also demonstrate the effectiveness of our data augmentation $\lambda(\text{crop})$ on [3].

is nearly identical for both of these scenarios with one small difference. In the case of whole image augmentation, the re-scaling function $\lambda(\cdot)$ does not operate within the bounds of ± 0.5, but between 0.0 and -0.5 – analogously 1.0 for classification. This means that we only generate samples with a smaller grain density. This is because the re-scaling to a higher grain density would require generating or replicating new image regions to fit the space left empty from the re-sizing. Using the whole image will be denoted as *img*, while the use of crops of size 224×224 will be denoted as *crop*.

Architectures. Due to the relatively small size of the dataset, retraining state-of-the-art feature extractors such as ResNet18(Φ_{res}) [18] or AlexNet(Φ_{alex}) [20] architectures from scratch yields worse results than using pretrained models. Therefore, we use pretrained Φ_{res} and Φ_{alex} models on Imagenet [14] as the backbones in our experiments. These two architectures have shown to perform well in different image classification and regression tasks, some of which share the domain of microstructure analysis [3]. The two architectures also represent two paradigms in deep learning; convolutions alone, or incorporating residual blocks. Furthermore, we propose to use two different heads applied on top of the feature extractor: Ψ_{fc} and Ψ_{ext} (see Fig. 3). Ψ_{fc} is a single fully-connected layer on top of the feature extractor, commonly used in fine-tuning from a pretrained model. The other, Ψ_{ext} is an extended head with two intermediate fully-connected layers, to allow for a larger capacity in the classification or regression head.

Metrics. We evaluate image classification performance with Top 1 accuracy. However, for regression, exact prediction of the grade is neither necessary nor effective. A more comparable metric to classification is to allow for a margin of ± 0.5 around the regressed prediction. If the prediction lies within the margin we still consider it to be correct. This relates to the available metalographic data being labeled either in whole-grain or sometimes in half-grain steps.

Experimental setup. Each network architecture is trained with Adam [19] with an initial learning rate of 3e-4. Training spans 1,500 epochs and the final model is chosen from the epoch with the lowest validation loss before evaluating on the test split. Each experiment consists of 20 seeds to measure the robustness to different initializations.

V. EXPERIMENTAL RESULTS

To assess the performance of the proposed strategies, we first compare results on classification, then on regression, and finally we summarize and discuss them together.

Classifying Grain Density. The first approach we consider is using classification networks to solve the problem of grain density estimation. We show our results across the different configurations introduced in Section IV and compare them in Fig. 4b). Furthermore we compare our models directly to the approach proposed in [3] (Φ_{azimi}). We note that our best configuration Φ_{res}^{cls} outperforms Φ_{azimi} by 19.4% on average, if Φ_{azimi} is trained with their proposed scheme. However, if we employ our proposed data augmentation pipeline $\lambda(\text{crop})$ on Φ_{azimi} performance improves and the gap is reduced to 11.3%, yielding an improvement of 8.1% just by using $\lambda(\cdot)$. As Φ_{alex} never exceeds 60% Top 1 accuracy across the various settings it is omitted from the ablation figure.

Regarding our results of Φ_{res}, we show that training on whole images results in better performance than training on image crops. The network heads Ψ show no effect when training on whole images, and a slight increase of performance when using Ψ_{ext} on crops.

Regressing Grain Density. We also investigate using regression networks to estimate the grain density. In Fig. 5, we show an ablation of the regression configurations. Φ_{res} outperforms Φ_{alex} in every configuration that is comparable. This is especially impressive as Φ_{res} has only roughly 11 million parameters, whereas Φ_{alex} has around 60 million parameters. It is easy to conclude that there is neither a gain in performance nor a gain in computational cost in using Φ_{alex}. We find that the best performing model is Φ_{res} with a plain Ψ_{fc} head, using image crops and our λ augmentation. This is also shown and summarized in Table I.

Regarding the heads, Ψ_{fc} in combination with Φ_{res} yields models that have a smaller standard deviation. This is explained by the fact that heads with more capacity tend to over-fit on the limited data, while the pretrained backbone is robust enough to not degenerate. Another interesting finding is that passing the whole image (i.e. global information) through the network generally performs worse across all tested configurations than using crops, except for one outlier. This indicates that local information plays a more important

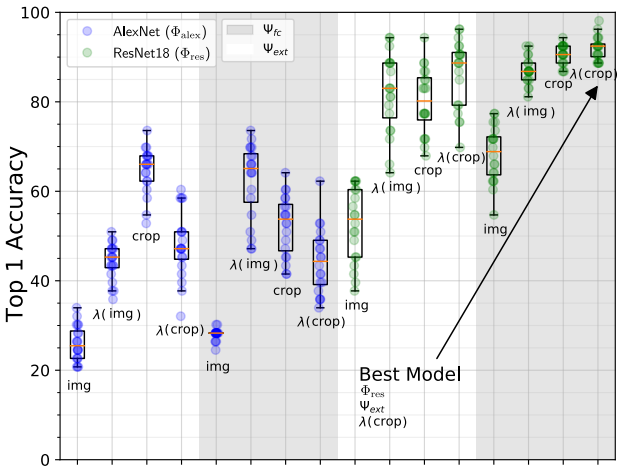


Fig. 5. Grain density estimation with regression. Comparison of Φ_{res} (green) and Φ_{alex} (blue) when training on images (img) or crops (crop), combined with classifiers Ψ_{fc} (gray) and Ψ_{ext} (white), and with data augmentation $\lambda(\cdot)$.

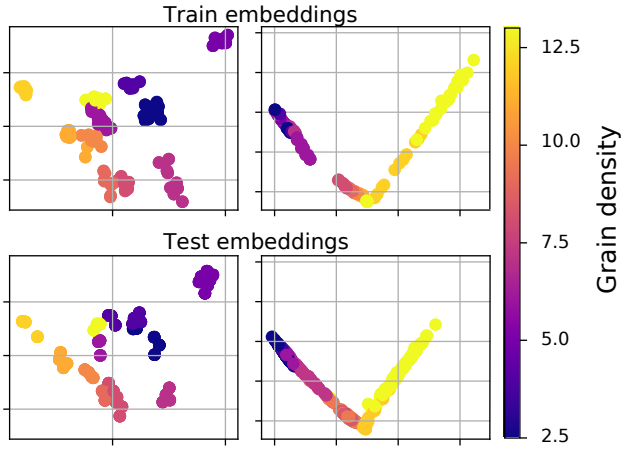


Fig. 6. Classifier Φ_{res}^{cls}(left) and Regressor Φ_{res}^{reg}(right) feature space visualization with PCA. While classification models outperform regression models w.r.t. Top 1 accuracy, it might come at a cost. The learned feature representation of the classifier, while good at separating classes, does not span the grain density space continuously according to their size. This is in contrast to regressors that clearly show a grain-density axis.

TABLE I

SUMMARY OF GRAIN DENSITY ESTIMATION

	Classification		**Regression**	
	Crop	Image	Crop	Image
$\Phi_{\text{res}}, \Psi_{\text{ext}}$	89.91 ± 3.02	98.11 ± 0.09	86.42 ± 7.66	81.98 ± 8.80
$\Phi_{\text{res}}, \Psi_{\text{fc}}$	93.02 ± 4.07	97.92 ± 0.58	91.89 ± 2.68	86.98 ± 2.99

TABLE II

MEAN ABSOLUTE ERROR ON CLASSES OF UNSEEN GRAIN DENSITIES

Model	Train	Test	Unseen Classes
Best Classifier ($\Phi_{\text{res}}, \Psi_{\text{ext}}^{cls}$)	0.000	0.019	1.038
Best Regressor ($\Phi_{\text{res}}, \Psi_{\text{fc}}^{reg}$)	0.135	0.216	**0.646**

role, and that the anticipated inhomogeneities within a sample do not contribute to wrong estimates, which is surprising because Classifiers Φ_{res}^{cls} all performed considerably better with global than with local information. Finally, our data augmentation strategy $\lambda(\cdot)$ increases performance by $\sim$7% on average w.r.t. Φ_{res}. The best model configuration is a combination of Φ_{res}, Ψ_{fc}, and $\lambda(\text{crop})$, as seen in Table I.

Discussion We generally observe that classification models outperform their regression counterparts (see Table I). In contrast to regression which prefers crops to images, we find that classifiers exhibit preference towards whole images. This already hints that the learned feature representation for regression and classification is drastically different, which we explore further in the following section.

VI. FEATURE REPRESENTATION AND VISUALIZATION

Interpretability. Visualizing the feature space of learned image representations is often done to gain insight into the decision making process. When visualizing embeddings $\Phi(\mathbf{x}) \in \mathbb{R}^{256}$ from image $\mathbf{x}$, we need to reduce its high dimensionality to allow for better visual analysis. This step could be done with methods such as Principal Component Analysis (PCA) [30] or t-Stochastic Neighbor Embedding (t-SNE) [29]. We choose PCA since distances in the projection are preserved, unlike in non-linear projections such as t-SNE.

We forward pass our training samples through Φ_{res}^{cls} and Φ_{res}^{reg}, with *cls* and *reg* denoting the best classifier and regressor network backbones. We then apply the same projection to the test set and observe where they end up in feature space. Results are shown in Fig. 6. A drastic difference between the feature representations of Φ_{res}^{cls} and Φ_{res}^{reg} can be observed. The embedding space of the classifier does not arrange classes corresponding to grain densities in any particular order. Instead, classes form clusters where interpolation in the feature space does not equal interpolation in grain density. Contrarily, a very orderly arrangement of grain densities emerges when learning with regression, as shown in the right column of Fig. 6. These results are particularly interesting as we previously show that Φ_{res}^{cls} outperforms Φ_{res}^{reg} by a significant margin, thus one could relate a more structured feature space representation to better performance.

Out-of-distribution robustness. In order to investigate if the ordered grain density feature structures emerging from learning with a regressor is beneficial, we explore inference on unseen and out-of-distribution data that *does* occur in real world scenarios. We further have metalographers annotate 668 new samples belonging to half-grade density austenite steel – which finer partition is commonly used in real world applications – and captured with a similar setup as the data described in Sec. III. Concretely this new dataset consists of austenite steel images with classes corresponding to grain densities ranging from 3.5 to 12.5, in increments of 1.0 – with only the grain density 10.5 missing.

In Fig. 7, we show the embedding plots of these unseen classes, both for Φ_{res}^{cls} and Φ_{res}^{reg}. Once more it can be observed

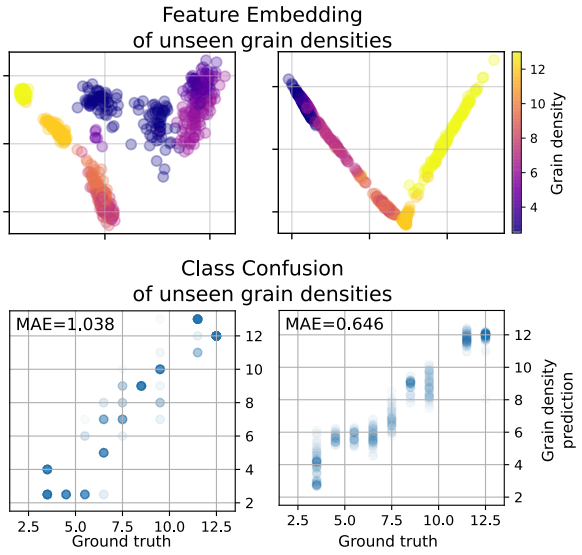


Fig. 7. Feature embeddings of unseen grain densities on a classification (left) and regression (right) model. Bottom row shows the regression plot, in essence a confusion matrix, where each sample is plotted relating its ground truth and predicted value. The Mean Absolute Error (MAE) is considerably lower for the regression model.

that the feature space exhibits mostly a continuous representation of the grain densities in the case of Φ_{res}^{reg}, and Φ_{res}^{cls} exhibits the same clustering behaviour. Not only is this shown qualitatively in the visualization, but is also quantitatively established in terms of Mean Absolute Error (MAE) over the out-of-distribution samples. In conclusion, Table II summarizes our findings by showing that the classifier performs both better on train and test sets, but generalizes worse to out-of-distribution samples. In contrast, regression presents a potential trade-off between the performance of a model and its interpretability at the feature representation, which allows evaluation of intermediate grain densities without re-training.

Grain density attention mapping. Work that focuses on visualization and explainability of convolutional neural networks has been around almost since their inception [31]. A common technique, especially for classification-based methods, is the use of class activation map (CAM) [32] algorithms. Since the grain density estimation is also framed as a classification problem, we can apply GradCAM [27], a popular CAM algorithm, to highlight areas in images that correspond to particular classes. We exploit the fact that an ordering of the grain density classes exists, which enables us to analyse image structures that lead to high activations in the output neurons. This can be used in order to visually perform a grain density homogeneity estimation.

GradCAM generates attention maps based on the gradients of a network w.r.t. a particular class and image. We perform a GradCAM step for every single class given, stack the generated attention maps, and compute the maximally activated class value for each pixel. The resulting scalar field is a pixel-wise class activated discriminative map.

To test the robustness and predictive capabilities of our proposed architectures we splice together an image consisting of 4 individual images of different grain densities. The images used correspond to those in Fig. 2. The resulting class attention maps are shown in Fig. 8, for a model trained on whole images and one trained on crops. We observe that the classifier network that was trained on whole images has difficulties to detect the boundaries of the various grain densities, whereas the network trained on crops shows no such limitation and produces a heatmap delineating the spliced quadrants very well. The crop trained model processes the individual crops separately, which are then assembled to a single attention map. The inhomogeneity detection and visualisation provided by the crop-trained model could be explored in future work, because the homogeneity of austenitic steel is useful for determining its mechanical properties.

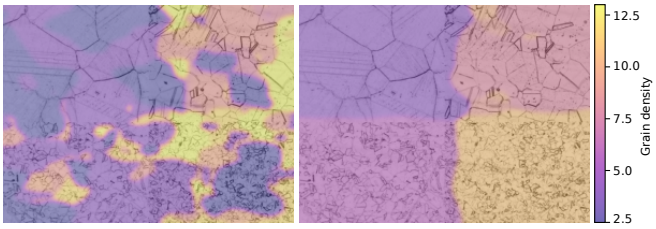


Fig. 8. An artificially spliced image from 4 different grain densities trained with whole images (left) and crops (right). The overlayed heatmap is generated by our proposed argmax GradCAM modification to provide pixel-wise grain density estimates. Best viewed digitally.

VII. CONCLUSION

We explore classification and regression with deep neural networks for estimating the grain density of austenitic steel samples taken with optical microscopes. We show that classification models overall yield better results than comparable regression models. Our findings show that the learned feature representation of classifiers and regressors differs drastically. The feature embedding of regressors yields an interpretable axis that corresponds to the actual grain density, whereas classifiers do not seem to encode the grain density as a major dimension in their feature space, and instead partition it into rigid, easily separable clusters. This is also reflected in the results that compare the performance of both types models on previously unseen grain density samples. Dealing with such out-of-distribution samples is especially important in the context of real-world applications. Since regression is shown to be robust w.r.t. out-of-distribution samples while maintaining accurate grain density estimates, we demonstrate a feasible way of additional quality control in steel mills. We also show the adaptation of a popular CAM algorithm to visualize grain densities and inhomogeneities within a sample, which also provides insight into the learned feature representation. Due to limited data, common in these settings, we introduce a novel data augmentation technique tailored to grain density estimation, which is shown to improve the performance of both classifiers and regressors.

ACKNOWLEDGMENT

This work was supported by *Land Steiermark* within the research initiative "Digital Material Valley Styria". Marc Masana acknowledges the support by the "University SAL Labs" initiative of *Silicon Austria Labs (SAL)*.

REFERENCES

[1] A. Agrawal and A. Choudhary, "Deep materials informatics: Applications of deep learning in materials science," *MRS Communications*, vol. 9, no. 3, pp. 779–792, 2019.

[2] E.-. ASTM, "Standard test methods for determining average grain size," *ASTM International: West Conshohocken, PA, USA*, 2004.

[3] S. Azimi, D. Britz, M. Engstler, M. Fritz, and F. Mücklich, "Advanced Steel Microstructural Classification by Deep Learning Methods," *Scientific Reports*, vol. 8, 02 2018.

[4] A. Bansal, X. Chen, B. Russell, A. Gupta, and D. Ramanan, "Pixelnet: Representation of the pixels, by the pixels, and for the pixels," *arXiv preprint arXiv:1702.06506*, 2017.

[5] D. Bulgarevich, S. Tsukamoto, T. Kasuya, M. Demura, and M. Watanabe, "Pattern recognition with machine learning on optical microscopy images of typical metallurgical microstructures," *Scientific Reports*, vol. 8, 12 2018.

[6] K. Choudhary, B. DeCost, C. Chen, A. Jain, F. Tavazza, R. Cohn, C. WooPark, A. Choudhary, A. Agrawal, S. J. L. Billinge, E. Holm, S. P. Ong, and C. Wolverton, "Recent advances and applications of deep learning methods in materials science," 2021.

[7] A. Chowdhury, E. Kautz, B. Yener, and D. Lewis, "Image driven machine learning methods for microstructure recognition," *Computational Materials Science*, vol. 123, pp. 176 – 187, 2016.

[8] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.

[9] V. H. C. de Albuquerque, P. C. Cortez, A. R. de Alexandria, and J. M. R. Tavares, "A new solution for automatic microstructures analysis from images based on a backpropagation artificial neural network," *Nondestructive Testing and Evaluation*, vol. 23, no. 4, pp. 273–283, 2008.

[10] C. G. de Andrés, F. Caballero, C. Capdevila, and D. San Martın, "Revealing austenite grain boundaries by thermal etching: advantages and disadvantages," *Materials Characterization*, vol. 49, no. 2, pp. 121–127, 2002.

[11] B. L. DeCost, T. Francis, and E. A. Holm, "High throughput quantitative metallography for complex microstructures using deep learning: A case study in ultrahigh carbon steel," *Microscopy and microanalysis: the official journal of Microscopy Society of America, Microbeam Analysis Society, Microscopical Society of Canada*, vol. 25 1, pp. 21–29, 2018.

[12] B. L. DeCost, J. R. Hattrick-Simpers, Z. Trautt, A. G. Kusne, E. Campo, and M. L. Green, "Scientific ai in materials science: a path to a sustainable and scalable paradigm," *Machine learning: science and technology*, vol. 1, no. 3, p. 033001, 2020.

[13] B. L. DeCost and E. A. Holm, "A computer vision approach for automated analysis and classification of microstructural image data," *Computational Materials Science*, vol. 110, pp. 126 – 133, 2015.

[14] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *Conference on Computer Vision and Pattern Recognition*. IEEE, 2009, pp. 248–255.

[15] A. I. Durmaz, M. Müller, B. Lei, A. Thomas, D. Britz, E. Holm, C. Eberl, F. Mücklich, and P. Gumbsch, "A Deep Learning Approach for Complex Microstructure Inference," *Nature communications*, vol. 12, no. 1, 2021.

[16] R. Girshick, "Fast r-cnn," in *International Conference on Computer Vision*. IEEE, 2015, pp. 1440–1448.

[17] J. Gola, D. Britz, and F. Mücklich, "3D-Gefügeforschung und neue Möglichkeiten der zuverlässigen Gefügeklassifizierung durch Kombination mit maschinellem Lernen," *METALL*, vol. 72, pp. 454–456, 2018.

[18] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Conference on Computer Vision and Pattern Recognition*. IEEE, 2016, pp. 770–778.

[19] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[20] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems*, vol. 25, pp. 1097–1105, 2012.

[21] J. Kusiak and R. Kuziak, "Modelling of microstructure and mechanical properties of steel using the artificial neural network," *Journal of Materials Processing Technology*, vol. 127, pp. 115–121, 09 2002.

[22] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," in *Proceedings of the IEEE*, vol. 86, no. 11, 1998, pp. 2278–2324.

[23] B. Mulewicz, G. Korpala, J. Kusiak, and U.Prahl, "Deep convolution neural networks in classification of metals microstructure," in *International Conference on Adaptive Modeling and Simulation, ADMOS 2019*, 2019.

[24] B. Mulewicz, G. Korpala, J. Kusiak, and U. Prahl, "Autonomous interpretation of the microstructure of steels and special alloys," *Materials Science Forum*, vol. 949, pp. 24–31, 03 2019.

[25] M. Müller, D. Britz, and F. Mücklich, "Application of trainable segmentation to microstructural images using low-alloy steels as an example," *Practical Metallography*, vol. 57, pp. 337–358, 04 2020.

[26] M. Oquab, L. Bottou, I. Laptev, and J. Sivic, "Learning and transferring mid-level image representations using convolutional neural networks," in *Conference on Computer Vision and Pattern Recognition*. IEEE, 2014, pp. 1717–1724.

[27] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization," in *International Conference on Computer Vision*. IEEE, 2017, pp. 618–626.

[28] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *International Conference on Learning Representations*. IEEE, 2015.

[29] L. Van der Maaten and G. Hinton, "Visualizing data using t-sne." *Journal of machine learning research*, vol. 9, no. 11, 2008.

[30] S. Wold, K. Esbensen, and P. Geladi, "Principal component analysis," *Chemometrics and intelligent laboratory systems*, vol. 2, no. 1-3, pp. 37–52, 1987.

[31] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *European Conference on Computer Vision*. Springer, 2014, pp. 818–833.

[32] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, "Learning deep features for discriminative localization," in *Conference on Computer Vision and Pattern Recognition*. IEEE, 2016, pp. 2921–2929.