Comprehensive Quantitative Quality Assessment of Thermal Cut Sheet Edges using Convolutional Neural Networks

Janek Stahl[1]
janek.stahl@ipa.fraunhofer.de
Andreas Frommknecht[1]
andreas.frommknecht@ipa.fraunhofer.de
Marco F. Huber[1,2]
marco.huber@ieee.org

[1] Department of Machine Vision and Signal Processing
Fraunhofer Insitute for Manufacturing Engineering and Automation IPA, Stuttgart, Germany

[2] Institute of Industrial Manufacturing and Management IFF, University of Stuttgart, Germany

Abstract

In this study, we present a novel holistic approach to assess the quality of thermal cut edges using images of the cut edges. Using deep learning techniques, we estimate quality criteria such as roughness, edge slope tolerance, groove tracking, and burr height. Our approach significantly surpasses the current state of the art in evaluating thermal cut edges using 2D images. To the best of our knowledge, this study presents the first image-based groove tracking evaluation for thermal cut edges. Our results show that a comprehensive, accurate, and fast prediction of edge quality can be effectively achieved by implementing a simple image acquisition system combined with a convolutional neural network (CNN).

1 Introduction

In sheet metal production, the quality of the edges of the thermal cut sheets is crucial, as it significantly influences the performance and reliability of the final product. Manufacturers confront the challenge of attaining consistent, high-quality results due to the wide variation in material properties and the complexity of the cutting process. This complexity includes phenomena such as focus shift [25], which can alter cut quality during machine operation mainly caused by thermal lensing [10]. Identifying parameters that consistently yield high-quality results is a complex task.

A primary concern is to maintain a high-quality edge while minimizing production costs and waste. This challenge has spurred the demand for efficient and reliable real-time evaluation methods for thermal cut sheet edges. In this paper, we present a comprehensive approach that employs the capabilities of CNNs to assess the quality of thermal cut sheet edges, offering rapid and precise quality evaluations based on key criteria such as averaged roughness depth [11], slope tolerance, groove tracking, and burr height.

Our approach does not rely on complex measurement systems. Instead, we use an image acquisition system consisting of an industrial RGB camera in conjunction with an incident light source and a transmitted light, making the process more accessible and cost effective. Using this simple setup, we can effectively determine the quality criteria of thermal cut edges, resulting in a more efficient and reliable production process.

2 State of the Art

Evaluating sheet metal quality by human visual inspection requires years of experience and expertise and remains a subjective method. Analytical measurements and calculations of quality criteria, on the other hand, do not suffer from subjectivity, but often require expensive equipment and laboratory conditions that are impractical in sheet metal production environments. As a result, industry demand for a fast, accurate and cost-effective method of quality assessment is growing.

The emergence of deep learning for visual tasks since 2012 [13] has enabled researchers to explore new methods for evaluating the quality of thermally cut sheet edges simply by using images of the cut edge. Before the rise of deep neural networks, rule-based methods dominated such image processing tasks. Particularly in our use case of sheet metal quality assessment, these traditional methods would require impractical and economically unfeasible feature engineering tailored to each specific criterion, sheet thickness, and material. However, deep neural networks introduced a paradigm shift in computer vision by automating feature engineering. This shift was underlined by the high accuracy demonstrated by breakthrough models such as AlexNet. This approach is proving to be much more efficient than measurements, as it allows for quick and easy data acquisition, making it particularly suitable for use in sheet metal production environments, while providing high accuracy and reduced subjectivity. In recent years, several studies have been published that have addressed the challenges of evaluating the quality of thermal cut sheet metal edges. Stahl *et al.* [21] demonstrated the potential for quick and robust quantitative evaluation of thermal cut sheet edge roughness using images by training a modified AlexNet [13] on them. Tatzel *et al.* [24] extended this work by demonstrating the ability of CNNs to predict multiple cut edge roughness values with a larger dataset.

While these studies focused primarily on roughness evaluation, Stahl *et al.* [22] investigated the effect of edge illumination on the evaluation. The authors used a low-budget setup with simple LED light patches and captured images with a mobile device, incorporating edge slope tolerance and burr height into the quality prediction. De Mitri *et al.* [4] explored the implementation of cut edge segmentation on mobile devices and evaluated the image area of the segmented edge for image sharpness. The analysis of image sharpness is conducted to ensure that the image has sufficient quality, making the evaluation meaningful. Both studies share the common goal of developing a low-cost, app-based quality assessment for cut edges, taking advantage of the widespread use of smartphones and accessible technology to streamline the assessment process. Our proposed method addresses another gap by also evaluating groove tracking, thus providing a comprehensive cut edge quality assessment.

3 Methodology

In this study, we aim to develop a deep learning model capable of predicting the holistic quality of thermal cut edges using images of the cut edges as input. Our methodology consists of several key steps, including dataset generation, image acquisition and preprocessing, and model training. We systematically varied the parameters of a thermal cutting machine to produce a large set of square stainless steel samples, from which we obtained images and relevant quality metrics. We then preprocessed these images and implemented various deep learning architectures, such as VGG16 and Xception, for model training. To optimize the models, we performed hyperparameter tuning and evaluated the performance using cross-validation and quantitative metrics such as the coefficient of determination (R^2).

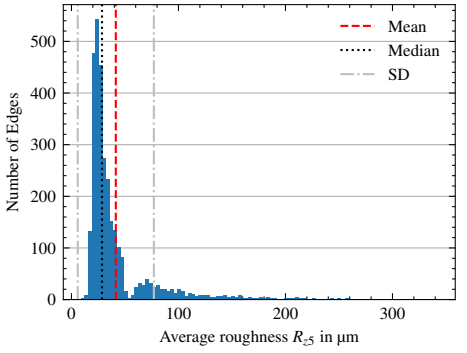
3.1 Creation of the Dataset

In the present work, a comprehensive set of parameter combinations for a thermal cutting machine was used to produce square stainless steel samples with a thickness of 3 mm and a side length of 100 mm. The parameters, including feed rate, gas pressure, nozzle-to-sheet distance, and adjustment value, were methodically varied, resulting in 785 cut samples. Due to the square shape of the samples, four edges per sample were obtained, resulting in a total of 3,140 cut edges.

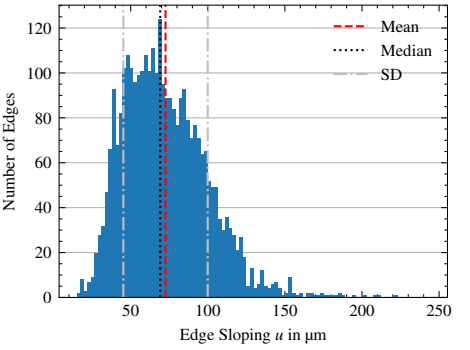
An optical measurement system (3D profilometer VR-3200, Keyence Corporation) was used to scan a 50 mm section of the edge surface on each side. Based on the acquired height data, the average roughness depth (R_{z5}), the edge slope tolerance (u) and the groove tracking (n) were calculated in accordance with the DIN EN ISO 9013 standard [5]. To evaluate roughness, measurement lines are strategically placed on the edge surface parallel to the cutting direction. The first line is placed 300 µm below the top edge, with subsequent lines at 300 µm intervals. The roughness value for all these lines is calculated and the maximum value is selected as the ground truth for the edge. In addition, the underside of the edges was measured to obtain data on the formation of burrs. Since the standard does not provide a calculation scheme for the quantitative determination of the burr height, it is defined for this paper as follows: From the depth information of the scanned underside at the center of the measurement, a 20 mm section is extracted and divided into five segments, each 4 mm wide along the cutting direction. For each segment, the distance between the maximum height value and the underside of the sheet is determined. The ridge height b is then calculated by averaging the five differences.

Data Histograms

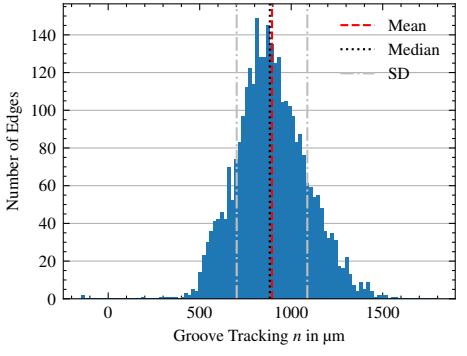
This results in a dataset whose distributions are shown in Figure 1. Figure 1(a) shows the distribution of data points for the averaged roughness depth R_{z5}, which has a distinct right skew. The majority of the data points are in the range of 10 to 50 µm. A noticeable dip at 50 µm is due to the adjustment of the measuring distance when an expected averaged roughness depth of 50 µm is reached (see [5]). The data set contains a limited number of cut edges with elevated averaged roughness depth values. In Figure 1(b), the histogram of the slope tolerance data points shows a distribution that resembles a normal distribution with a slight right skew. The data set has a small number of outliers with an increased edge slope tolerance. Figure 1(c) shows the distribution of the groove tracking data, which also approximates a normal distribution. A subset of the cutting edges has a negative value,



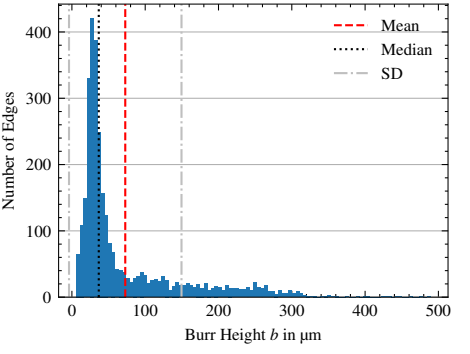
(a) Distribution of data points for the average roughness R_{z5}.



(b) Distribution of data points for the edge slope tolerance u.



(c) Distribution of data points for the groove tracking n.



(d) Distribution of data points for the burr height b.

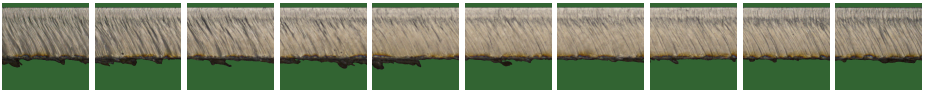
Figure 1: Histograms of the data set for the different criteria for holistic quality assessment of thermal cut edges.

indicating that the grooves run in the cutting direction, in contrast to the majority, which run in the opposite direction. Finally, Figure 1(d) reveals the distribution of the burr height across the dataset. The histogram shows a right-skewed distribution, with a significant portion of the data in a low-value range below 70 µm. These uneven distributions could have a negative impact during model training such as model bias (see Section 4).

Image Acquisition and Image Preprocessing

For image acquisition, a Basler color camera model acA4096-30uc is used in combination with a standard 35 mm fixed focal length MeVis-C lens from LINOS. An 8 mm distance ring is placed between the camera and the lens to achieve the desired working distance and a 50 mm field of view. A coaxial incident light LFV3-70SW from CCS was chosen for the setup. In order to improve the visibility of metal residues at the lower edge of the cut for the determination of the burr height, a transmitted light illumination is integrated into the setup. This transmitted light highlights the contour of the metal residue or burr at the bottom edge of the cut. In addition, the transmitted light image can be used as a mask for segmenting the cut edge surface. The CCS LFL-100SW2 surface illumination is selected as the transmitted light

(a) Region of interest within an image showing the cut edge of a thermal processed sheet.



(b) Disjoint image sections of the region of interest of a cut edge image.

Figure 2: Example of preprocessed images. Using both transmitted and incident illumination allows for simultaneous segmentation, filtering out image regions that are not relevant for quality assessment.

illumination. By binarizing the transmitted light image with the OTSU threshold [16] and applying this as a mask for the incident light image, efficient, fast and robust segmentation is achieved. This allows the segmentation of the relevant image region to be determined automatically. The non-relevant background pixels of the incident light image are replaced by green pixels. Figure 2(a) shows the region of interest in a preprocessed image.

Typically, CNN structures are designed and benchmarked on the ImageNet dataset [5], which uses square image dimensions. As a result, conventional CNN architectures are tuned for square images. The images of a cutting edge, however, have a rectangular shape. To avoid potential changes that could reduce performance, we divide the image of the cut edge into ten separate, non-overlapping segments, as shown in Figure 2(b). During model training, each individual image is labeled with the cut edge information and treated as an independent training data point. Therefore, the number of data points is increased tenfold. This approach requires that images of an edge are fully associated with either the training set, the validation set, or the test set. Overlaps could lead to misleading results due to data leakage. With individual image predictions, multiple predictions are obtained per cut edge. To acquire a single value for each quality criterion per cut edge, the predictions of all individual images of an edge are averaged.

3.2 Model Training

To study the effect of model architecture on prediction quality, we use VGG16 [19] and Xception [2]. VGG16 is a simple deep convolutional neural network with small 3x3 filters in its 16 weight layers. Xception extends the Inception [23] architecture by utilizing a depth-separable convolutional design that allows efficient use of model parameters. To use CNNs for regression, we modified both architectures by replacing the output layer with four fully connected neurons, each tied to an evaluation criterion. We used a linear activation function, linking each final layer parameter linearly to an output neuron. This modification allowed us to predict continuous values tied to our specific evaluation metrics, making the architectures suitable for our regression-based task. By comparing their performance, we aim to understand the impact of architecture on prediction quality and identify the most suitable choice for our application. To accommodate the Xception architecture, we resized the images to 299×299 pixels as it is specifically designed for this input size. For the VGG16, we resize them to 224×224 since it is designed for this size.

To investigate the effectiveness of data augmentation techniques, which have been shown to improve the performance of deep learning models in image classification [17], we applied moderate data augmentation of brightness range variation as well as strong augmentation including brightness range variation, height shift, width shift, horizontal flip, and rotation up to two degrees in each direction. We then compared the results with those of a model trained without data augmentation.

Given our relatively small dataset, we used models pre-trained on the ImageNet dataset, as the effort required to enlarge the dataset is substantial. We implemented a two-step transfer learning approach as described by Chollet [3]. First, we trained the unadjusted weights of the top layers of the model by feature extraction, keeping the feature extraction part of the model frozen. Then, the last block of the feature detector was unfrozen for fine tuning [26].

We used min-max normalization to the interval $[0,1]$ for the multiple output regression model to ensure that all output variables were on the same scale, thereby preventing any single output from dominating the others in terms of magnitude and facilitating more efficient model training.

Before initiating the training process for each model, we conducted hyperparameter optimization utilizing the Keras tuner library [15] in conjunction with the hyperband technique [14]. This enabled us to investigate various hyperparameter combinations and to identify the ideal configurations customized for our specific task. For the hyperparameter tuning, we split the dataset into 80 % for training, 10 % for validation, and 10 % for testing, ensuring a fair evaluation of the model's performance during the optimization process. During the training phase, we employed early stopping [2] with a 20-epoch patience and performed 10-fold cross-validation to ensure a more reliable evaluation of the model's performance. We use mean square error as the loss function and Adam optimization [12] as the optimization method. For training purposes, each edge image is labeled with the corresponding edge quality. To evaluate the edges in the test set, we averaged the predictions for individual images, ignoring the spread of individual image predictions and assigning equal weights to each prediction. In order to evaluate the quality of the model predictions both quantitatively and qualitatively, the coefficient of determination (R^2) is employed [8].

The choice to average predictions from square patches was rooted in the underperformance of alternative methods, such as using rectangular images or an R-CNN sequence approach. This strategy was aimed at amplifying the training data and concentrating the model's attention on significant local features by simplifying the image context. A simple average was employed to collate predictions spanning the entire image due to its straightforwardness and resultant effectiveness. The implementations for this study were carried out using the TensorFlow Python library [1].

4 Results

The results of the model training are shown in Table 1, which presents the accumulated prediction quality of the k-fold cross-validation as the coefficient of determination R^2 for each criterion, along with the average prediction. Without data enrichment, the Xception model achieves the highest average prediction accuracy of 87.2 %.

The best prediction for roughness, with an R^2 of 86.1 %, is obtained by fine-tuning (FT) the VGG16 model and using brightness range variation as an augmentation technique, beating the previous best of 70.9 % by Stahl *et al.* [22]. For edge slope tolerance, our method achieves the highest prediction of 79.2 % with the Xception model without data

Table 1: The accumulated prediction quality of the k-fold cross-validation is presented as the coefficient of determination R^2 for each criterion, along with the average prediction.

Base-Archi.	Training	Aug.	$R^2(R_{z5})$	$R^2(u)$	$R^2(n)$	$R^2(b)$	$R^2(avg.)$
VGG16	FT	no	85.6 %	77.1 %	87.3 %	96.0 %	86.5 %
VGG16	FT	med.	86.1 %	77.4 %	87.9 %	96.3 %	86.9 %
VGG16	FT	strong	84.6 %	74.8 %	85.4 %	95.1 %	85.0 %
Xception	**FT**	**no**	**85.8 %**	**79.2 %**	**87.7 %**	**96.2 %**	**87.2 %**
Xception	FT	med.	85.8 %	78.9 %	87.7 %	96.1%	87.1 %
Xception	FT	strng	85.2 %	77.4 %	87.2 %	95.7 %	86.4 %

augmentation, surpassing the previous best of 64.2 %. In addition, the VGG16 model with brightness range variation beats the previous best burr height prediction of 91.4 % to 96.2 %, and sets the first benchmark for groove tracking with 87.7 % accuracy.

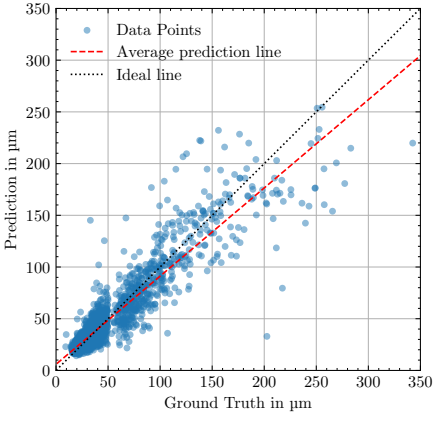
By applying moderate image augmentation during training, such as varying the brightness range, an improvement in average prediction is observed for the VGG16 architecture. In contrast, a marginal 0.1 % decrease in average prediction is observed for the Xception architecture. Interestingly, heavy use of image augmentation techniques during training results in an minor average decrease in prediction accuracy on the test set of 1.5 % for VGG16 and 0.8 % for Xception.

The results indicate that both the VGG16 and Xception models exhibit comparable performance across the evaluated criteria, with minimal differences between them, suggesting that the choice of either model may be dictated by specific requirements or preferences. VGG16 has approximately 138 million parameters, while Xception has a significantly lower parameter count of approximately 22.9 million, making it a lighter and more efficient alternative. As a result, Xception is recommended for applications where reduced computational requirements and improved efficiency should be achieved without sacrificing performance. Furthermore, the deliberately chosen image patch sizes for the models, namely 224×224 pixels for VGG16 and 299×299 pixels for Xception, did not seem to affect performance.

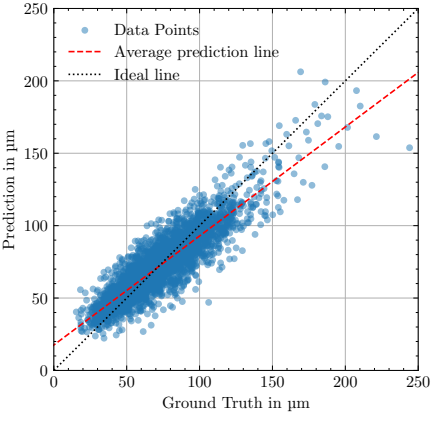
Figure 3 shows the cumulative test set predictions for each criterion from the ten runs using the Xception architecture without data augmentation. A black dotted line, representing the ideal prediction, and a red dashed line, obtained by least-squares fitting a first-degree polynomial function to the blue data points, are included in the plots for reference. The average spread of individual image predictions was observed to be: 5.0 µm for roughness, 6.4 µm for edge slope tolerance, 33.74 µm for groove tracking, and 8.9 µm for burr height.

4.1 Discussion

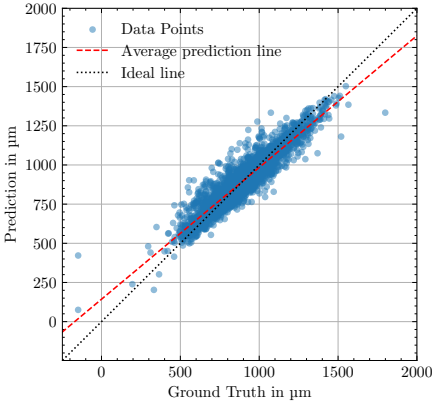
The results of this study show that excessive image augmentation during training can lead to a decrease in prediction quality on the test set. A possible explanation for this decrease is the divergence between the training and test datasets, which subsequently has a negative impact on the predictions. Similar effects were observed by Elgendi et al. [7]. Another possible reason is that certain image augmentations, such as width shifting, may change the semantic meaning by causing the loss of relevant edge regions, which consequently leads to a reduced performance on the test dataset, as pointed out by Shorten and Khoshgoftaar [18]. In addition, the augmentation applied may have been too aggressive, thereby undermining the



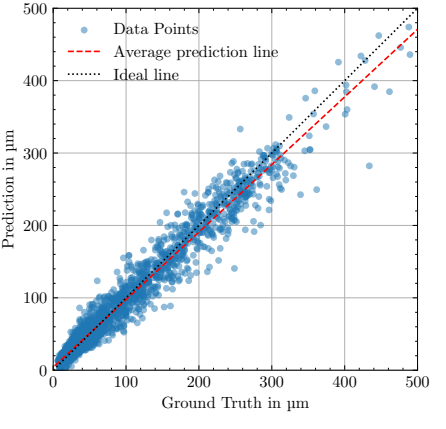
(a) Prediction diagram for the criterion averaged roughness R_{z5}.



(b) Prediction diagram for the criterion edge slope tolerance u.



(c) Prediction diagram for the criterion groove tracking n.



(d) Prediction diagram for the criterion burr height b.

Figure 3: The accumulated predictions of the quality criteria in the test sets of the k-fold-cross validation.

effectiveness of the training process, as noted by Snider [20] on object recognition problems.

Nevertheless, further investigation is needed to determine whether the image augmentation techniques employed in this study improve the model's robustness to variations in image quality during inference in real-world environments, despite the observed performance degradation within the test dataset.

Examining the scatter plots in Figure 3, it is evident that the red line has a less steep slope than the black line across all criteria. In addition, the red and black lines intersect in the range of values where the majority of the data points are concentrated, indicating that the model predictions tend to cluster around the centroid of the data.

In Figures 3(a) and 3(d), the result plots for average roughness and burr height show that the model has integrated the skew present in the data distribution (see Figure 1(a) and 1(d)) and tends to underestimate higher values. Conversely, predictions near the data center, which contains the majority of data points in the lower value range, are closer to the ideal line. A similar behavior can be seen in the result plots of Figure 3(b) and 3(c). The flattening of the red line is due to clustering around the data centroid in the training set. Values below the mean of the data distribution are generally overestimated, while those above are underestimated. Due to the data distribution (see Figure 1(b) and 1(c)) and the bias towards the mean, the model exhibits epistemic uncertainty, as shown by the red line. The greater the model's epistemic uncertainty, the worse its predictions for a given criterion, and vice versa. Despite these biases, it can be stated that with an average prediction accuracy of 87.2%, the model is capable of providing fast and high quality user feedback using 2D images.

A further notable finding from this study was that, for the roughness and burr criteria, we observed a correlation between the spread of individual image predictions and the deviation of the averaged value from the ground truth. This correlation offers the possibility to identify bad predictions and consequently to implement a weighted averaging of the predictions to improve the accuracy of the model predictions. However, the extent of the correlation and the appropriate approaches require further investigation in subsequent studies. Addressing these areas will improve the overall predictive capability of our model and provide a more robust assessment of surface roughness and burr.

5 Conclusion

The results suggest that cut-edge images alone can be used for holistic quality assessment. However, the unpredictability of the cutting process and the challenge of controlling quality variation make it difficult to obtain a balanced data set, particularly with respect to roughness and burr. Samples with high roughness and burr values are particularly challenging to produce, as they indicate a cutting process that is close to its operating limit, with a higher risk of damage to the cutting unit from slack spatter. Such damage can require costly repairs, additional to the cost of producing these samples.

The result plots also show that the roughness prediction is compromised by numerous outliers. These outliers are present in all training results, regardless of architecture or data augmentation, probably due to inaccuracies in the ground truth. This could be due in part to the measurement technology and in part to the subsequent computational method. Since the roughness value depends on the position of the measurement line (see Section 3.1), the sampling distance is critical and could be responsible for missing the true maximum roughness at the edge, resulting in outliers above the ideal line. To address this uncertainty and potentially improve prediction accuracy, future studies could consider increasing the sampling

frequency by decreasing the distance between the measurement lines. On the other hand, the Keyence VR-3200 operates on the fringe projection method. Reflective surfaces, such as the stainless steel edges examined in this study, can pose challenges due to reflections and subsequent overexposure that distort height measurement data. This distorted data results in inflated roughness values that distort the ground truth and explain the outliers below the ideal line. Alternatively, using a different measurement principle, such as depth of field based measurement or confocal microscopy, could avoid this problem altogether.

In the future, models could be developed to understand the complex relationships between quality criteria and process parameters. Such models could then be used to provide machine operators with process optimization suggestions to produce samples that meet desired quality standards. Given the wide variety of cutting processes, materials, and thicknesses in the sheet metal industry, the road to a universal model that can help all machine users is a long one. However, this study represents a first step toward a comprehensive quality assessment of thermal cut edges.

References

[1] M. Abadi. TensorFlow: Learning Functions at Scale. In *Proceedings of the 21st ACM SIGPLAN International Conference on Functional Programming*, pages 1–1, 2016.

[2] F. Chollet. Xception: Deep learning with depthwise separable convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1251–1258, 2017.

[3] F. Chollet. *Deep learning with Python*. Simon and Schuster, 2021.

[4] O. De Mitri, J. Stahl, C. Jauch, and C. Distante. Image acquisition, evaluation and segmentation of thermal cutting edges using a mobile device. In *Multimodal Sensing: Technologies and Applications*, volume 11059, page 110590U, 2019.

[5] J. Deng, W. Dong, R. Socher, L. Li, K. Li, and L. Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255, 2009.

[6] Deutsches Institut für Normung. DIN EN ISO 9013:2017-05: Thermisches Schneiden - Einteilung thermischer Schnitte - Geometrische Produktspezifikation und Qualität, 5 2017. DIN EN ISO 9013.

[7] M Elgendi, M. Nasir, Q. Tang, D. Smith, J. Grenier, C. Batte, B. Spieler, W. Leslie, C. Menon, R. Fletcher, et al. The effectiveness of image augmentation in deep learning networks for detecting covid-19: A geometric transformation perspective. *Frontiers in Medicine*, 8:629134, 2021.

[8] L. Fahrmeir, C. Heumann, R. Künstler, I. Pigeot, and G. Tutz. *Statistik: Der Weg zur Datenanalyse*. Springer-Verlag, 2016.

[9] I. Goodfellow, Y. Bengio, and A. Courville. *Deep learning*. MIT press, 2016.

[10] T. Himmer, T. Pinder, L. Morgenthal, and E. Beyer. High brightness lasers in cutting applications. In *International Congress on Applications of Lasers & Electro-Optics*, volume 2007, pages 87–91. Laser Institute of America, 2007.

[11] International Organization for Standardization. Geometrical product specifications (GPS) – Surface texture: Profile – Part 2: Terms, definitions and surface texture parameters. Standard ISO 21920-2:2021, International Organization for Standardization, Geneva, Switzerland, 2021.

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

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

[14] L. Li, K. Jamieson, G. DeSalvo, A. Rostamizadeh, and A. Talwalkar. Hyperband: A novel bandit-based approach to hyperparameter optimization. *The Journal of Machine Learning Research*, 18(1):6765–6816, 2017.

[15] T. O'Malley, E. Bursztein, J. Long, F. Chollet, H. Jin, L. Invernizzi, et al. Kerastuner. https://github.com/keras-team/keras-tuner, 2019.

[16] Nobuyuki Otsu. A threshold selection method from gray-level histograms. *IEEE transactions on systems, man, and cybernetics*, 9(1):62–66, 1979.

[17] L. Perez and J. Wang. The effectiveness of data augmentation in image classification using deep learning. *arXiv preprint arXiv:1712.04621*, 2017.

[18] C. Shorten and T. Khoshgoftaar. A survey on image data augmentation for deep learning. *Journal of big data*, 6(1):1–48, 2019.

[19] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[20] E. Snider, S. Hernandez-Torres, and R. Hennessey. Using ultrasound image augmentation and ensemble predictions to prevent machine-learning model overfitting. *Diagnostics*, 13(3):417, 2023.

[21] J. Stahl and C. Jauch. Quick roughness evaluation of cut edges using a convolutional neural network. In *Fourteenth International Conference on Quality Control by Artificial Vision*, volume 11172, page 111720P, 2019.

[22] J. Stahl, C. Jauch, M. Tuncel, and M. Huber. Investigation of different illumination scenarios for the evaluation of thermally cut edges with convolutional neural networks using a mobile device. *Electronic Imaging*, 33:1–7, 2021.

[23] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9, 2015.

[24] L. Tatze and F. P. León. Image-based roughness estimation of laser cut edges with a convolutional neural network. *Procedia CIRP*, 94:469–473, 2020.

[25] L. Tatzel and F. P. León. Impact of the thermally induced focus shift on the quality of a laser cutting edge. *Journal of Laser Applications*, 32(2):022022, 2020.

[26] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson. How transferable are features in deep neural networks? *Advances in neural information processing systems*, 27, 2014.