DATA ACQUISITION CHALLENGES IN AI-DRIVEN SURFACE INSPECTION: A PROVEN SOLUTION PROPOSAL ON COATED SHEET METAL PARTS

Sebastian Hunger
Miele & Cie. KG
Carl-Miele Straße 21, 33332 Gütersloh, DEU
sebastian.hunger@miele.com

Michael Breiter
Eyyes GmbH
Im Wirtschaftspark 4, 3494 Gedersdorf, AUT
michael.breiter@eyyes.com

Claudia Klein
Eyyes GmbH
Im Wirtschaftspark 4, 3494 Gedersdorf, AUT
claudia.klein@eyyes.com

ABSTRACT

This paper introduces an advanced AI-based automated surface inspection system for enhanced quality control in the manufacturing of white-coated sheet metal parts. The study emphasizes overcoming the challenges in AI-driven inspection, such as the need for large datasets corresponding to numerous physical components, which presents storage and logistical issues. By integrating Convolutional Neural Networks (CNNs) and a novel annotation process, the system can be trained effectively on various surface defects. The paper discusses three big data acquisition challenges and provides a solution approach, including large data volumes equivalent to numerous physical components, a novel division of the training process to reduce the workload for domain experts and the relevance of previously clearly defined defect classes.

Keywords AI Surface Inspection · Data acquisition challenges · AI-driven Quality Control · Quality Assurance in Manufacturing · AI in Industrial Process · Convolutional neural networks (CNN) · EfficientNet · VGG16

1 Introduction

The increasing complexity of industrial production processes requires innovative solutions such as automated surface inspection. This technology offers greater precision and repeatability than conventional methods [1]. In today's manufacturing industry, quality assurance is crucial, particularly for customer-relevant surfaces in the production of household appliances. With the aim to eliminate the strenuous, subjective and repetitive process of manual inspection of the components during production [2], this paper introduces an automated surface inspection system, representing a significant advancement in quality inspection. Utilizing artificial intelligence (AI) and advanced camera technology, this system facilitates fully automated assessments of white-coated sheet metal parts formed directly from the coil. The challenge of reliably identifying even the smallest defects with a thickness of a tenth of a millimeter on the approximately 0.5 square meter white-coated sheet metal components is paramount, as these directly impact customer satisfaction and brand perception.

1.1 Related work

Several studies have explored the domain of surface defect detection and classification through various machine learning models. The authors of [3] demonstrate the training of a ResNet model to recognize three different types of defects on steel raw surfaces, achieving a recall rate between 30% and 75% and precision between 72% and 90%, depending on the defect type. In [4] a convolutional auto encoder based anomaly detection method is shown to distinguish between defective and defect-free metal part regions. The authors show the whole process of model selection and parameter optimization but can't find a sufficiently well working result. In that work, no defect type classification is performed. In contrast, [5] outlines a method for training neural networks specialized in anomaly detection. They use a special approach to learn convolution layers for domains with low number of defect samples available. They show promising results on the public DAGM dataset. Furthermore, [6] introduces a new benchmark dataset for the classification of defects on raw steel surfaces. They further train a fast and efficient CNN model targeting inline-inspection, though the practical requirements for the deployment of the application are not discussed. There are many valuable insights into good camera/light settings in [7]. However information to automatic detection of defects and/or algorithmic classification is missing. To our knowledge there is a clear lack of scientific discourse and empirical studies in the field of inline surface inspection of pre-coated raw material with analogue dimensions and transport speeds.

1.2 The Initial situation

The pre-coated raw material is processed directly from the coil, starting with the uncoiling of the coated strip. Before the initial forming step, the material is straightened using leveling rolls and then cut into sheets approximately 600x700 mm using an automatic shear. Subsequently, the material undergoes six forming steps in a tool on a 1000-ton press, creating functional holes, cutouts, beads, and flanges in the component. The finished component is then placed on a conveyor belt and transported with a speed of 11.5 m/min to a manual inspection and handling station. Here, employees inspect the components for permissible defects and prepare them for assembly on the final product.

1.3 AI-driven surface inspection

The new AI-driven surface inspection system is placed on the conveyor belt between the outfeed conveyor of the press and the manual inspection and handling station. This is achieved by having the components pass through an inspection tunnel mounted on the conveyor belt (see fig. 2 no. 1.; movement from bottom right to top left). This makes it possible to inspect the front walls during the run without additional handling and to provide the result of the automated inspection when the respective component arrives at the manual handling station. Inside the test tunnel, 25 light strips and 30 high-resolution monochrome cameras are installed. The cameras are arranged in three rows and allow multiple recording directions at different angles of incidence to the inspected surface. They are mounted in a height-adjustable manner. The particular obstacle posed by the required overlap of the cameras is overcome through precise measurement of the overlap and dedicated calibration methods. The resolutions of up to 16.2 megapixels (5328 x 3040 pixels) ensures that the smallest unacceptable defect (a tenth of a millimeter in the highest quality zone) is displayed with at least 4 pixels. This value has proven to be effective in practical testing and can be used as a basis for the selection of cameras in further inspection systems. A complete raw data recording of one of the 600x700 mm components generates an image dataset of 2.7 GB (lossless compression as PNG). At full production capacity of the press, the live system thus needs to process approximately 1700 megapixels per second. For this reason, the computing and evaluation unit is located directly adjacent to the inspection tunnel (see fig. 2 no. 2.).

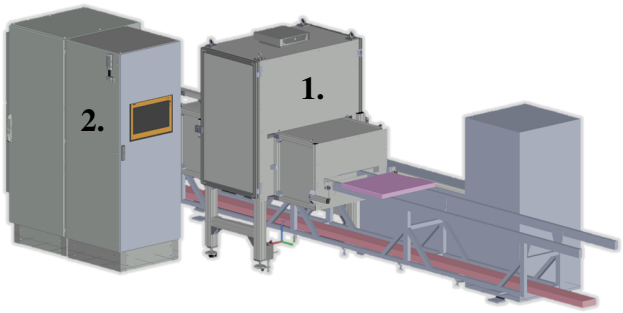


Figure 1: AI-driven surface inspection system

1.4 Why using AI for inspection ?

In the overall assessment of whether a component is acceptable, the type of surface defect (error or defect **class**) and its position (error or defect **localization**) play a crucial role. The localization of a detected anomaly is determined by an underlying CAD model of the component, light gate and a high-resolution encoder (5000 pulses/rotation) on the conveyor belt, which measures rotational motion and translates it to electrical signals. Since the components are always placed in the same orientation at the press outlet, the absolute position of a detected anomaly can be determined by rule-based algorithms. The detected anomaly is then classified by a convolutional neural network (CNN) into one of the following defect classes: scratch, impression, pore/prick, inclusion, bump/dent, contamination, streak, flitter, delamination from the base material and unknown defect. Whether the anomaly then becomes a defect depends on the combination of its position on the component and the defect class. This is why the defect classification capability of the CNN is of the greatest importance for the system. It is therefore not enough to

simply recognise an anomaly. For example, in the highest quality zone (direct visual area of the customer), defects such as scratches, pimples/inclusions, etc., are only permissible up to a size of a tenth of a millimeter. Shape-changing defects that alter the base material and coating, such as bumps and dents, are generally not permissible in this quality zone. In one of the lower quality zones, however, smaller bumps and dents with a diameter lying on the surface of five-tenths of a millimeter are allowed. Here, the robustness, accuracy, and classification capability of the CNN are crucial to avoid misclassifications. Otherwise, an anomaly will be classified as an impermissible defect due to incorrect classification, although the component should have been classified as acceptable (pseudo-error) or, worse, faulty components that should not be installed are rated as acceptable (slip).

1.5 CNN

Convolutional Neural Networks (CNNs) are known since 1989 as highlighted in LeCun´s work on "Backpropagation Applied to Handwritten Zip Code Recognition" [8]. Their ongoing success story was marked by a significant milestone in 2012, when they triumphed in the ImageNet Large Scale Visual Recognition Challenge, surpassing various other methods [9]. Since this breakthrough, CNNs have been instrumental in addressing numerous complex problems through Deep Learning. The most important property of CNNs is their ability to autonomously learn salient features directly from raw data during the training process. This characteristic distinguishes them from other machine learning approaches, which typically rely on pre-defined features. This given, CNNs can easily be scaled to use a higher volume of features and also more complex features. This adaptability and learning capability make CNNs a versatile and powerful tool in the realm of deep learning.

Convolutional Neural networks use convolution layers for deep feature learning. A convolution layer is the mathematical convolution operation, represented as neural network node. In computer vision, 2D spatial convolution over the image pixels with a third dimension for inter-channel influence is used to extract features in images. Several convolution layers as a feed-forward network are forming a deep neural network, where deep features gain a higher receptive field and can represent occurrence of more complex structures or shapes in the original image. CNNs are inspired by the early visual cortex of animals [6].

A single convolution operation consists of a square 2D spatial size N and the number of input channels C, giving N * N * C learnable parameters. In a convolution layer, M such convolutions are used to create M outputs, which are used as inputs for the next layer.

In formula 1 the computation of a single convolution layer output is shown, where $in(x', y', c)$ is the input layer with the 2D spatial index x' and y' and the channel index c, out is the output layer with channel index m and w_m is the m-th convolution filter with its convolution parameters indexed by x,y and c. In addition, $\alpha = N/2$ is an offset to center the convolution window around the input index position.

$$\mathrm{out}(x', y', m) = \sum_{c=1}^{C} \sum_{y=1}^{N} \sum_{x=1}^{N} w_m(x, y, c) * \mathrm{in}(x + x' - \alpha, y + y' - \alpha, c) \qquad (1)$$

In a depthwise convolution, the 3D convolution is reduced to pure 2D convolution ($C = 1$), without inter-channel influence, followed by a 1 x 1 x C convolution, bringing

back some inter-channel influence. This reduces the number of parameters and the number of necessary operations.

1.6 VGG16 vs. EfficentNet

The VGG-16 (Visual Geometry Group-16) network architecture is a highly efficient model for image classification, developed by the Visual Geometry Group at the University of Oxford. It consists of 16 layers, including 13 convolutional layers and three fully connected layers, interspersed with five max-pooling layers to reduce spatial dimensions and improve computational efficiency. The input layer is configured for 224x224 pixel images. Each convolutional layer uses 3x3 pixel filters, increasing in depth from 64 to 512. The model uses Rectified Linear Unit (ReLU) activation functions and ends with a softmax function for class probability outputs [10].

The EfficientNet B0 network architecture uses mobile inverted bottleneck layer with depth-wise separable convolution (mbconv) as building blocks, which are a more efficient approximation for general convolution filters. It consists of 16 such layers, two general convolution layers, a global average pooling layer and a fully connected layer. The channel depth is growing from 32 in the first layer to 320 in the last mbconv layer. The input resolution is 224x224. The architecture is designed to be scalable to reach best quality for any processing complexity target. The findings of Mingxing Tan and Quoc V. Le led to an alternative training approach utilizing the architecture of EfficientNet B0. The authors demonstrated that the EfficientNet architecture achieves comparable quality with reduced computational complexity, compared to other architectures [11]. However, a direct comparison with the VGG architecture was not explored in their study.

2 Challenges in Data acquisition

For the industrial application of AI-driven surface inspection three key challenges in the acquisition of training data can be identified:

1. **Uniform Anomaly Recording:** Anomalies identified by human inspectors must be uniformly and systematically recorded. These anomalies should be classifiable as either usable or defective, as much as possible independently of the inspector. Here, the personal perception of a defect plays a significant role.

2. **Unbalanced Dataset:** The defects detected on sheet metal parts measuring 600x700mm are predominantly less than one square millimeter in size. For this, a concept must be developed that allows for the rapid and targeted preparation of numerous small defects on a relatively larger defect-free surface for training the neural network.

3. **Data Management:** As mentioned in section 1.5 substantial data volumes are required for training purposes. These extensive data sets are equivalent to a considerable number of physical components, which present specific challenges in terms of storage and logistics.

The first challenge can be practically addressed by clearly defining error classes and incorporating additional well-defined meta information such as high- or low-contrast defects, their size or length (especially in the case of scratches) and position. This can be an excellent aid for an independent assessment of defects. It is advisable to clearly define in advance which anomalies (and which meta information) are to be classified as acceptable or unacceptable and not to leave this decision to the personal perception of surface quality. For this purpose, it is recommended that a small core team of quality

experts establish a clear distinction between the defect classes, e.g. a differentiation according to characteristics such as penetration depth, whether the coating penetrates completely or whether the defect is located on the substrate and the coating remains at least locally intact. This can also provide an additional decision-making aid in borderline cases. However, this is also not a trivial matter and is heavily dependent on individual circumstances, so that a detailed description is not provided here.

The second and third challenge was addressed in the project through the introduction of a specially developed "coarse" and "fine" annotation process. In the so-called coarse annotation process, the general position of the detected anomaly is marked using defect stickers, as shown in Figure 2. Each of these annotation stickers has a size of 15x10 mm (code 8x8 mm) and is assigned with a unique numerical identifier. After applying the sticker, the characteristics described under the first challenge, such as error class, size, length, and assessment regarding in-spec or out-of-spec (defective) anomalies, can be noted behind the sticker's identifier. The coarsely annotated component is then fed into the system in a special training mode. The system identifies the stickers and saves the image section with a recognized sticker for further processing. This allows only the raw images around an embroidered defect to be saved in a targeted manner (50-120 MB). Because of the extensive data associated with a part, manually searching for a marked defect area and extracting the images would be a time-consuming and resource-intensive task prone to errors. Although re-annotations are a natural part of the process in any AI project the time dedicated to them has been significantly reduced. Using the sticker number also allows for rapid location of noted defects at any time. From this point onwards, the physical component is no longer required. Subsequently the targeted fine annotation process can begin, in which the exact defect region is marked on the image series taken independently of the physical component and can be used for subsequent training.

A significant advantage arises from this training method, particularly crucial for the training process in industrial practice. Components can be immediately marked during the production process by quickly applying an annotation sticker. The second advantage is that no highly trained quality expert with domain knowledge is required to perform the coarse annotation process. This creates a decoupling in the data acquisition process between pure anomaly detection and the necessary classification of anomalies for training and eliminates the need for long-term storage, entry, and repositioning of physical components, which is particularly beneficial for larger sheet metal parts.

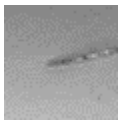
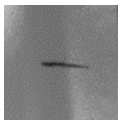
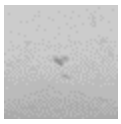
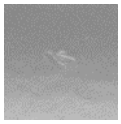
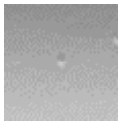
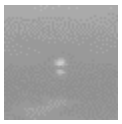
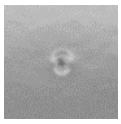
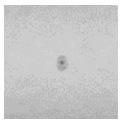
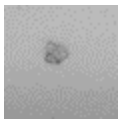
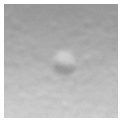
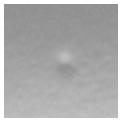
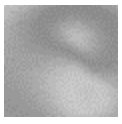
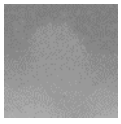
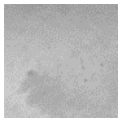
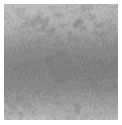
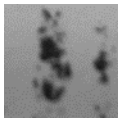
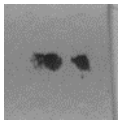
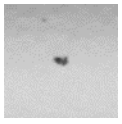
Figure 2: Positioning an annotation sticker near an anomaly

3 Data and Results

The initial data was collected as described above and annotated in 9 defect classes, both coarse and fine. Some examples of the most common and currently best represented error classes are shown in Table 1. In practice, it is recommended to provide a corresponding table with sample images during manual training data preparation, which enables a quick assignment to the in advance defined error classes.

3.1 Data

Table 1: defect class overview

Scratch			
Impression			
Pore/prick			
Inclusion			
Bump/dent			
Streak			
Contamination			

In addition, regions without any defects are collected by recording parts that don't have any defects and considering every region as "no defect" sample further grouped by "background" and "part". Collecting of the defects is done by the annotation process described above. Meta information like type and size is assigned to the defect as a

database entry per annotated defect. Manual corrections and data quality checking is done by machine learning experts with deep learning knowledge. After a few iterations of data collection, prototypical training and hard negative mining, a few defect classes were removed from training, because of low number of occurrences or ambiguity. Table 2 shows the current number of samples per class. It's essential to acknowledge that the research is still in the training phase and the noted imbalance in the number of test samples is expected to correct itself over time. In subsequent analyses the reasons for any deviations that occur will be discussed.

Table 2: samples per class

Class	Defects	Samples	Augmented	Training	Validation	test
No defect:						
Background	-	∼58500	∼58500	∼46800	∼12000	0
Part	-	∼196000	∼196000	∼156000	∼40000	∼7300
Defect:						
Scratch	176	639	5112	4368	744	13
Impression	155	488	4264	3580	684	31
Pore/prick	38	43	404	340	64	35
Inclusion	254	263	2104	1892	312	42
Bump/dent	81	117	936	792	144	33
Contamination	103	535	3026	2305	721	35

A single defect (column 2) on the part results in multiple samples (column 3) because an image is captured every 4.5 mm, so multiple images of the same defect are available. Data augmentation is used to create multiple additional samples (column 4). Only data augmentation that preserves the original image quality such as position jitter, mirroring and multiple of 90°rotations are used, so no interpolation is used which would introduce artificial features (or remove available features). Neither noise addition, nor blurring or resizing is used as augmentation by purpose, because specific cameras in their full pixel-resolution are used by design. Training set is used to learn the CNN weights. Validation set is used to observe the training and to select the best training stage result before overfitting begins.

3.2 Results

There is a strong class-imbalanceness in the data, which is handled by a special batch-balancing, where underrepresented classes are oversampled within a training epoch. Data was split into disjunct training and validation sets. No defect samples that belong to the same defect are put in separate sets. Training of VGG-16 and EfficientNet B0 were performed with an input resolution of 160x160[1] pixel grayscale images. Evaluation shows an overall accuracy of 99% (EfficientNet) and 98% (VGG), which is dominated by the high number of no-defect samples. The accuracy weighted by the number of samples per class is significantly lower with 53% (EfficientNet) and 58% (VGG). In detail (see Tabel 3), both classifiers perform well on the defect classes, scratch, inclusion and contamination and is weaker for impression, pore/prick and bump/dent. Due to the relatively small amount of data, a random effect can be assumed and no further analysis of the differences is carried out at this stage.

[1]Due to the generally very small defect size, the input layer was reduced from 224x244 to 160x160, resulting in improved processing time and lower storage volumes for the defect sections.

Defects	EfficientNet	VGG16
Scratch	0.69	0.85
Impression	0.39	0.32
Pore/prick	0.26	0.17
Inclusion	0.64	0.67
Bump/dent	0.18	0.42
Contamination	0.77	0.66

Table 3: Accuracy comparison of the detection performance of EfficientNet and VGG

From analyzing the confusion matrix and actual image content it looks like bump/dent and pore/brick share some typical patterns with other defect classes, which makes training harder. For these classes, a much higher number of samples will be needed, so that the classifier can learn how to distinguish fine details in the pattern, without overfitting to a few specific defect samples. It is important to mention that the strong imbalance and classifier focus on giving a very strong performance on the no-defect samples has some benefit in practice. In the production system, a very sensitive anomaly detection is used to generate defect candidates. After accepting only candidates that are detected at least 3 out of approx. 8 times at nearly the same part position (+/- 0.5 mm), while theoretically visible in the camera field of view, the classifier decides whether it is still a non-defect region like dust or textile fibers, or whether it is one of the defect classes. On an average real part, over 5000 candidates have to be classified, most of them no-defect regions. For that reason, the performance on no-defect regions has to be outstanding good, otherwise every single part would be overall classified as defective.

4 Conclusion

In conclusion, this paper introduces an novel AI-driven automated surface inspection system for the manufacturing of white-coated sheet metal parts, marking a substantial advancement in quality control. By integrating Convolutional Neural Networks (CNNs) and a new annotation process, the system adeptly addresses several key challenges in AI-based surface inspection, particularly the need for substantial datasets and the accompanying storage and logistical complexities.

The implementation of a structured and uniform anomaly recording process, coupled with a new approach to manage unbalanced datasets, underlines the system's capability to handle a wide range of surface defects with precision. The dual annotation method, integrating both coarse and fine processes, streamlines the training of the neural network, ensuring rapid preparation of large volumes of data with minimal human intervention.

However, it is crucial to acknowledge that this paper presents only the initial stages of the CNN training process. There remains a substantial need for additional training data to further refine the CNN algorithms and enhance their defect detection capabilities (probably around eight thousand more samples). Nevertheless, the system can already be used for initial support during surface inspection. The comparison between the EfficientNet and VGG16 models provides valuable insights into their respective strengths and areas for improvement in defect detection, which will guide future enhancements and adaptations of the AI-driven surface inspection system.

In summary, while this study offers a robust solution to current challenges in manufacturing quality control and sets a precedent for future research, it also highlights the ongoing journey in the realm of AI-driven surface inspection. The continual evolution of

training data and model refinement is essential for the advancement of quality assurance in manufacturing processes, promising a future of increasingly efficient and accurate automated inspection systems.

References

[1] Neogi, N., Mohanta, D.K. & Dutta, P.K. "Review of vision-based steel surface inspection systems". J Image Video Proc 2014, 50 (2014). https://doi.org/10.1186/1687-5281-2014-50

[2] Xie, Xianghua. “A Review of Recent Advances in Surface Defect Detection using Texture analysis Techniques.” Electronic Letters on Computer Vision and Image Analysis 7 (2008): 1-22. url: https://api.semanticscholar.org/CorpusID:8516210

[3] Konovalenko, Ihor, et al. "Steel surface defect classification using deep residual neural network." Metals 10.6 (2

[4] Kähler, Falko, Ole Schmedemann, and Thorsten Schüppstuhl. "Anomaly detection for industrial surface inspection: Application in maintenance of aircraft components." Procedia CIRP 107 (2022): 246-251.

[5] Staar, Benjamin, Michael Lütjen, and Michael Freitag. "Anomaly detection with convolutional neural networks for industrial surface inspection." Procedia CIRP 79 (2019): 484-489.

[6] Fu, Guizhong, et al. "A deep-learning-based approach for fast and robust steel surface defects classification." Optics and Lasers in Engineering 121 (2019): 397-405.

[7] Mike Muehlemann; "Standardizing Defect Detection for the Surface Inspection of Large Web Steel". Illumination Technologies, Inc. 2000

[8] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, L. D. Jackel; "Backpropagation Applied to Handwritten Zip Code Recognition". Neural Comput 1989; 1 (4): 541–551. doi: https://doi.org/10.1162/neco.1989.1.4.541

[9] Krizhevsky, Alex and Sutskever, Ilya and Hinton, Geoffrey E.; "ImageNet classification with deep convolutional neural networks". Commun. ACM 2017; 6 (60): 84-90. doi: https://doi.org/10.1145/3065386

[10] Karen Simonyan, Andrew Zisserman; "Very Deep Convolutional Networks for Large-Scale Image Recognition". Visual Geometry Group, Department of Engineering Science, University of Oxford; ICLR 2015 doi: https://doi.org/10.48550/arXiv.1409.1556

[11] Mingxing Tan, Quoc V. Le; "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks". ICML 2019 doi: https://doi.org/10.48550/arXiv.1905.11946020): 846.