An attempt to detect anomalies in car body parts using machine learning algorithms

Thomas Schromm[1,2], Fabian Diewald[1,2], Christian Grosse[1]

[1]Technical University of Munich, Chair of non-destructive testing, cbm - Center for Building Materials, Baumbachstrasse 7, 81245 Munich, Germany, E-Mail: thomas.schromm@tum.de

[2]BMW AG, 80788 Munich, Hufelandstrasse 5, Germany, E-Mail: thomas.schromm@bmw.de

Abstract

Industries, which produce hundreds of terabyte of CT data per year, demand automated evaluation approaches. This work provides a first glance of an attempt to automatically detect and characterize possible defects and/or anomalies which formed during common joining processes. We investigated a standard riveting process with respect to the resulting final head height of steel self-piercing half-hollow rivets. The methods include conventional image processing algorithms, like edge-detection, thresholding and principle component analysis (PCA) which were used to pre-process the CT data. In order to automatically evaluate the reconstructed volumes, which contained several of the aforementioned rivets, we compared the performance of different, publicly available, convolutional neural network (CNN) architectures. Furthermore, we investigated the impact of data augmentation and showed by means of a *k*-fold cross-validation that the training data causes no overfitting of the network. The obtained results suggest that an automated evaluation of the generated computed tomography scans, with regard to a rivet's final head height, is feasible. However, in order to increase the network's reliability and accuracy, the amount of training data needs to be further enlarged and diversified.

Keywords: Computed tomography, non-destructive testing, neural networks, automation, self-piercing half-hollow rivets

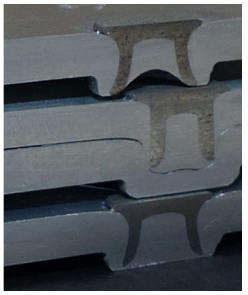
1 Introduction

Modern car bodies consist of several steel, aluminium and/or carbon fibre reinforced polymers (CFRP) components, which are usually joined by riveting, welding, screwing and/or glueing. The quality of the joined components and the joining elements, is in general crucial for the structural integrity of a car. Hence, a frequent and thorough inspection of the joining processes and their outcomes is of utmost importance. Joining processes in the research and development phase (R&D phase) of prototypes are often without any routine. Due to new designs, components, and structural requirements the testing of joints in this field is usually a model-specific testing process. The implementation of modi operandi is therefore impracticable.

The current inspection of car bodies in prototype production is predominantly destructive, decentralized and deals with testing an abundance (thousands) of joining sites. These three points result in the challenges that we aim to overcome: 1) A thorough investigation of every single inspection site is necessary but extremely time-intensive, due to the abundance of connecting components. 2) The joining sites are most often manipulated in the course of their testing, due to destructive testing methods (e.g. micro-sectioning and stress-tests). The use of computed tomography has the potential to eliminate this drawback. However, scanning whole cars [4, 6] is still very time-intensive. Instead, pieces of the body must be cut out before they can be analyzed in conventional industrial CT-systems. This, however, manipulates the sample once again. 3) Testing joining sites with micro-sections is not comprehensive enough, since only a 2D cross-section of the sample's 3D volume is used to evaluate its quality. 4) The testing of the respective joining techniques is often decentralized. Different modalities (micro-sectioning, thermography, ultrasound, CT, eddy-current, etc.) in different locations are used for different joints. This results in an avoidable wastage of resources.

In order to realize a non-destructive, sustainable and efficient testing process for car body prototypes, the points raised above must be amended. To do so, we aim at realizing an automated system which is able to both record and evaluate whole sections of car bodies comprehensively and without the need of destroying them. This work provides a first glance of an attempt to automatically detect and characterize possible defects and/or anomalies which occur during common joining processes. We investigated a standard riveting process with respect to the resulting final head height of steel self-piercing half-hollow rivets in aluminium plates (see Figure 1).

(a) Unprocessed rivets and joined plates displayed as micro-sections.



(b) Joined aluminium plates with differently processed steel rivets displayed as microsections.

(c) CT-reconstruction of joined aluminium plates with steel rivets.

Figure 1: Steel self-piercing half-hollow rivets

2 Samples

In order to create a data set that resembles realistic joining sites, the physical samples used in this work were specifically produced by experts who intentionally induced common defects/anomalies (see [8], [7] and [9]). Exemplary samples with their respective scans can be seen in Figure 1.
Among the different types of rivets used in car bodies and all the characteristic parameters used to determine whether or not a rivet joint is flawed or not, we chose to investigate the final head height of steel self-piercing half-hollow rivets (see Figure 2). In the following, assessing statements like flawed or flawless (correct) refer to the head height only. Other characteristics were not considered at this point. We joined two plates with either exclusively

- proud head heights (sticks too far out),
- flush head heights (penetrates too deep),
- flawless head height, and
- a mixture of proud, flush and flawless head heights.

The head height is expressed with the parameter h_r. This *physical* labelling makes it easier to produce *digitally* labeled training and test data later on.

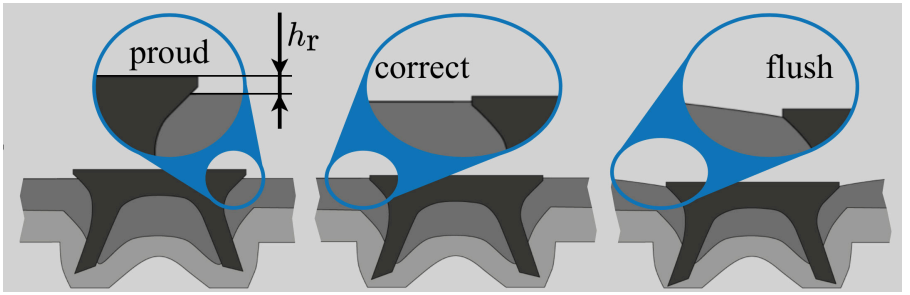


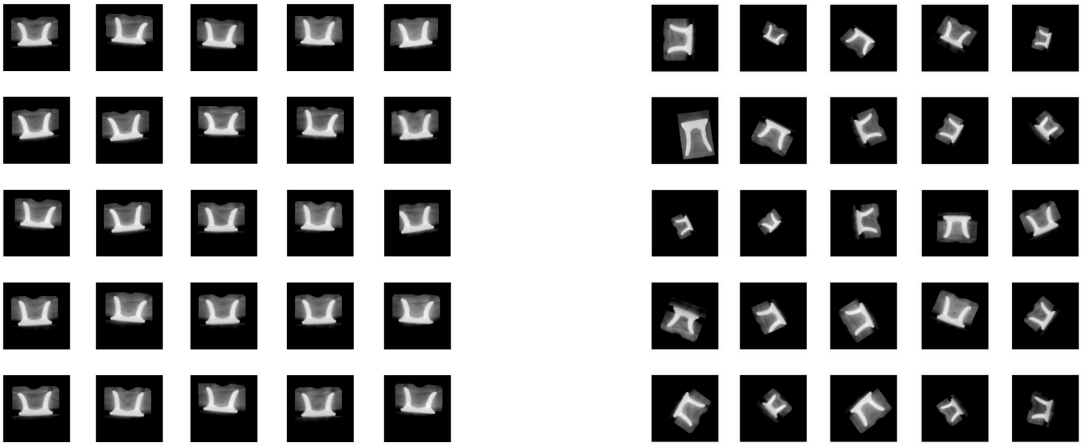
Figure 2: Three classes where investigated: proud, correct and flush head height. For specifications see [9].

The data set was generated by scanning the plates with two CT systems, namely a v|tome|x M240 and a v|tome|x L240, both by GE Sensing & Inspection Technologies. The tomography settings can be seen in Table 1:

Set	System	Current [μA]	Voltage [kV]	Power [W]	Projections	Detector area [px × px]	Averaged	Exp. Time [ms]
1	v|tome|x M240	390	220	85.8	2600	2024×2024	3	2000
2	v|tome|x L240	340	220	74.8	2200	2024×2024	4	1000

Table 1: Tomography settings

The rivet volumes were detected and extracted with conventional image processing algorithms like Otsu-thresholding [17] and region-growing [3]. In order to extract only symmetrical mid-sections through the center of mass of the 3D rivet, a principal



(a) Exemplary data extracted via PCA and image processing.

(b) Exemplary augmented data.

Figure 3: Excerpt from data-set

component analysis (PCA) was employed. Two of the resulting three eigenvectors span the desired mid-sections, that cut the rivets symmetrically in half. By applying these algorithms to the volume of the reconstructed CT-scan, 2D images (224x224 pixel) containing only the mid-section of one rivet were extracted. This approach produced 171 images (34 flush, 99 correct and 38 proud head heights). In order to artificially create a larger, more diverse data-set, data augmentation (translation, rotation and scaling) was employed. This enlarged the data set by a factor of 6, resulting in 1026 (204 flush, 594 correct and 228 proud head heights) images. An excerpt of the data set can be seen in Figure 3.

3 Evaluation methods

Four different pre-trained (with 10^6 images from the ImageNet data-base [16]) and publicly available CNN architectures were tested on the data set:

- vgg19 with a layer depth of 19 [12],
- resnet18 with a layer depth of 18 [13],
- resnet101 with a layer depth of 101 [13], and
- googlenet with a layer depth of 22 [14].

The training set was partitioned in 60% training data, 20% validation data and 20% test data. The trainings were all performed with a mini-batch size of 20, an initial learning rate of 10^{-4} and a maximum number of epochs of 30. To investigate the performance of the networks, correctly and incorrectly classified cases from the augmented data set are visualized in confusion matrices (see section 4).
To make the resulting network more robust against overfitting, a k-fold cross-validation was applied during the training process. It involves randomly partitioning of the original dataset into subsets of roughly equal size. By repeating the process k times, the average cross validation error indicates the performance of the network [15]. In our case, we generated $k = 3$ data sets, each with a different constitution regarding training, validation and test data set.

4 Results

If only the PCA data-set (without augmentation) is used for training, the overall classification success rates of the specific categories (proud, correct, flush) of the confusion matrices reached 62.6%, 66.5%, 66.5% and 63.1% for vgg19, resnet18, resnet101 and googlenet, respectively. The data augmentation improved the networks performance considerably as can be seen in Table 2. The respective improvements are $\uparrow$ 31.6%, $\uparrow$ 18.9%, $\uparrow$ 14.1% and $\uparrow$ 26.2%.
The validation accuracies of the 3-fold cross validation showed only subtle variations (86.0%, 86.7% and 85.5%) compared to the validation accuracy of the initial constellation (87.0%). Hence, the network did not fall victim to overfitting.

Table 2: Performance of different networks evaluated with confusion matrices. Top-left: vgg19, Top-right: resnet18 Bottom-left: resnet101, Bottom-right: googlenet. The sum resulting in each column (proud, correct, flush) represent how many instances of each category were actually in the data set. The distribution among the equally named rows, represent the predictions of the network. With an optimally performing network, the red cells contain only zero (no wrong classification) and the green cells the total number of each category (all instanced of each category were predicted correctly).

PREDICTION \ ACTUAL	proud	correct	flush	∑[%]
proud	45	0	0	100|0
correct	1	119	11	90.8|9.2
flush	0	0	30	100|0
∑[%]	97.8|2.2	100|0	73.2|26.8	94.2|5.8

PREDICTION \ ACTUAL	proud	correct	flush	∑[%]
proud	30	0	0	100|0
correct	15	114	9	82.6|17.4
flush	1	5	32	84.2|15.8
∑[%]	65.2|34.8	95.8|4.2	78.0|22	85.4|14.6

PREDICTION \ ACTUAL	proud	correct	flush	∑[%]
proud	33	7	0	82.5|17.5
correct	13	105	13	80.2|19.8
flush	0	7	28	80|20
∑[%]	71.7|28.3	88.2|11.8	68.3|31.7	80.6|19.4

PREDICTION \ ACTUAL	proud	correct	flush	∑[%]
proud	43	2	1	93.5|6.5
correct	3	108	7	91.5|8.5
flush	0	9	33	78.6|21.4
∑[%]	93.5|6.5	90.8|9.2	80.5|19.5	89.3|10.7

5 Conclusion

We investigated an automatic approach to evaluate the quality of steel self-piercing half-hollow rivets in CT-data. The feature we chose to investigate was the rivets' head height. The principle feasibility of detecting such a subtle characteristic automatically in CT reconstructions by means of image processing and convolutional neural networks was successfully shown. However, in order to increase the network's reliability and accuracy, the amount of training data needs to be further enlarged and diversified. In order to assess the quality of a rivet joint comprehensively, more quality characteristics need to be considered, meaning more samples need to be produced, scanned and used for training.

References

[1] Konopczyński T, Rathore D, Rathore J, et al. Fully Convolutional Deep Network Architectures for Automatic Short Glass Fiber Semantic Segmentation from CT scans. In: 8th Conference on Industrial Computed Tomography (iCT 2018). 8th Conference on Industrial Computed Tomography (iCT 2018). Wels, Austria, 2018.

[2] R. Schweiger, Anwendung von Deep Learning Methoden für die visuelle Inspektion von Oberflächen im Produktionsprozess von Fahrzeugteilen, Master Thesis, Technische Universität München, 2016.

[3] R. C. Gonzalez and R.E. Woods, Digital Image Processing 2nd Edition, Prentice Hall, New Jersey, 2002.

[4] NDT.net, XXL-Micro-CT - Comparative Evaluation of Microscopic Computed Tomography for Macroscopic Objects, `https://www.ndt.net/search/docs.php3?showForm=off&id=18029`, 2019

[5] NDT.net, Robo-CT - Robot based Computed Tomography and its Transition towards the full scale Car Body, `https://www.ndt.net/search/docs.php3?showForm=off&id=21965`, 2019.

[6] NDT.net, Potentials of Full-Vehicle CT Scans Within the Automotive Industry, `https://www.ndt.net/search/docs.php3?showForm=off&id=19225`, 2019

[7] R. Gschneidinger, 12.10.2018, personal interview

[8] T. Stadler, F. Häfner, 09.04.2018, personal interview

[9] BMW Group Standard: GS96001-2, 2017

[10] Scheidhammer Thomas, 07.05.2018, personal interview

[11] BMW Merkmalsklassifizierung von Faserverbundbauteilen. Version 14

[12] Simonyan, K., Zissermann, A., Very deep convolutional networks for large-scale image recognition, International Converence on Learning Representations, San Diego, 2015.

[13] He, K., Xiangyu, Z., Shaoqing, R., Juan, S., Deep residual learning for image recognition, Proceedings of the IEEE conference on computer vision and pattern recognition, 2016.

[14] Szegedy, C., Wei, L., Yangqing, J., Sermanet, P., Reed, S., Anuelov, D., Dumitru, E., Vanhoucke, V., Rabinovich, A., Proceedings of the IEEE conference on computer vision and pattern recognition, 2015.

[15] Goodfellow, I., Bengio, Y., Courville, A., Deep Learning, The MIT Press, Cambridge, 2016.

[16] Russakovsky, O., Deng, J., Su, H., ImageNet Large Scale Visual Recognition Challenge, International Journal of Computed Vision, Vol. 115, Issue 3, 2015.

[17] Nobuyuki Otsu, A threshold selection method from gray-level histograms. IEEE Trans. Sys., Man., Cyber. 9 (1): 62?66 (1979). doi:10.1109/TSMC.1979.4310076