



Deep Anomaly Detection for Endoscopic Inspection of Cast Iron Parts

Ole Schmedemann^(✉) , Maximilian Miotke , Falko Kähler ,
and Thorsten Schüppstuhl 

Institute of Aircraft Production Technology, Hamburg University of Technology, TUHH,
Denickestr. 17, 21073 Hamburg, Germany
ole.schmedemann@tuhh.de

Abstract. Detecting anomalies in image data plays a key role in automated industrial quality control. For this purpose, machine learning methods have proven useful for image processing tasks. However, supervised machine learning methods are highly dependent on the data with which they have been trained. In industrial environments data of defective samples are rare. In addition, the available data are often biased towards specific types, shapes, sizes, and locations of defects. On the contrary, one-class classification (OCC) methods can solely be trained with normal data which are usually easy to obtain in large quantities. In this work we evaluate the applicability of advanced OCC methods for an industrial inspection task. Convolutional Autoencoders and Generative Adversarial Networks are applied and compared with Convolutional Neural Networks. As an industrial use case we investigate the endoscopic inspection of cast iron parts. For the use case a dataset was created. Results show that both GAN and autoencoder-based OCC methods are suitable for detecting defective images in our industrial use case and perform on par with supervised learning methods when few data are available.

Keywords: Anomaly detection · Surface inspection · Endoscopy · Deep learning · Defect detection · Convolutional autoencoder · OCC · GAN · CNN

1 Introduction

A higher degree of automation of endoscopic inspection procedures by machine vision (MV) is often desirable, as cost and time savings are expected. In addition, the quality of inspection can be increased with MV, as manual inspection is often monotonous and therefore prone to fatigue-related errors. When inspecting cavities with visual endoscopy, the miniaturized imaging hardware (sensory, illumination) results in low image quality. Furthermore, endoscopic images are characterized by high variance due to varying relative positioning between the probe and the surface of a part. This makes the setup of classic MV systems challenging because they are based on manually engineered features.

Machine learning approaches promise to reduce the effort needed for setting up a MV system by learning relevant features from provided image data. In addition to reducing complexity in the setup phase, the generalizability of these approaches allows them to respond to high variance in the scene under investigation.

The performance of machine learning methods is highly dependent on the available data. In industrial environments, the availability of anomalous data is often limited or of insufficient quality. Collecting a large number of defective samples is costly [1]. In addition, expert knowledge is required to manually annotate the data. Anomalous data are often biased because rare defects are less common in certain positions or shapes. Others have demonstrated the applicability of deep supervised machine learning methods for automatic endoscopic inspection tasks [2, 3]. Martelli et al. [3] address the lack of suitable anomalous training data by manually creating anomalous samples. However, this is a tedious process and not reproducible for every defect type. Normal data from defect-free samples can be easily obtained in large quantities. Therefore, anomaly detection methods based solely on normal training data have proven to be a promising alternative for visual inspection tasks [4]. These methods learn the structure of the normal data. Images with defective surfaces are recognized as deviating from the learned structure.

In this work, the applicability of these methods for findings of endoscopic images in industrial surface inspection is investigated. A dataset is created from a real inspection task showing defective and normal images with surface defects such as crazes and voids. State-of-the-art anomaly detection methods are used and compared with the performance of supervised methods through the use of CNNs.

2 Related Work

2.1 One-Class Classification (OCC) for Visual Anomaly Detection

Anomaly detection refers to finding outliers patterns in data that do not correspond to a defined notion of normal data. When only normal data is used to train a classifier, then these methods are referred to as one-class classification. In this work, we focus on methods that use image data as an input. One can differ between shallow anomaly detection methods, i.e. Semi-Supervised One-Class Support Vector Machines [5], or deep anomaly detection methods. While deep methods are an active field of research, they have already shown the potential to outperform shallow methods [6]. Therefore, we focus on deep methods in this work. Often used methods are Convolutional Autoencoder (CAE) [7] or Generative Adversarial Networks (GANs) [8–10].

2.2 One-Class Classification Anomaly Detection for Industrial Inspection

Other have shown the applicability of CAEs and GANs for several inspection tasks. Liu et al. [11] used CAEs for an automated optical quality inspection task. They constructed a CAE to inspect surface defects on aluminum profiles. Tang et al. [12] investigated the applicability of OCC methods for the inspection of x-ray images of die castings. They successfully adopted a CAE and achieved a high classification accuracy of 97.45%. Kim et al. [13] uses a CAE with skip connections to inspect printed circuit boards. They achieved a high detection rate of 98% while keeping the false pass rate below 2%. GAN-based approaches have also been studied recently for industrial use cases.

However, our endoscopic inspection task differs from the presented tasks. There is more variance in our image datasets. Additionally, due to the miniaturized sensory the

image quality is reduced. We intuitively assume that this implies a more challenging anomaly detection. Thus, we want to investigate the performance of state of the art OCC methods on the endoscopic use case.

3 Use Case ‘Endoscopic Cavity Inspection of Cast Iron Part’

For this work, real images of a turbocharger housing from the automotive industry have been acquired. The casting must be visually inspected, including the cavities. Today, the part is manually inspected by a human worker. The complex free-form surfaces of the component make automatic visual inspection difficult because of variable distance and angle between sensor, illumination and part. Additionally, the cavity poses challenges to the inspection, such as difficult accessibility, low position accuracy, miniaturized sensory, and insufficient illumination.

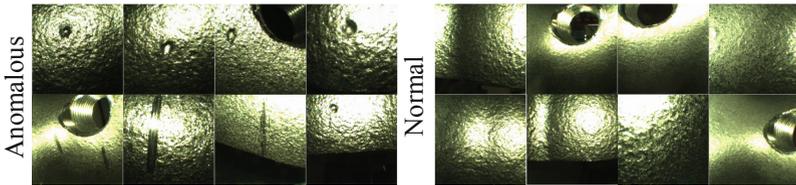


Fig. 1. Examples of the real-world images of the endoscopic images.

For the dataset, the endoscope was inserted by hand through the openings of the component. The dataset¹ consist of 1075 images split in anomalous and normal or defect-free images. Examples can be seen in Fig. 1. Both datasets show endoscopic images of the parts cavities. A 4 mm wide 90°-side view chip-on-tip endoscope is used with a 400x400 pixel resolution and an integrated LED illumination. The images in the dataset were randomly split in training, validation, and testing datasets, see Table 1.

Table 1. Split of endoscopic image surface defect dataset for model training.

	Training	Validation	Test	Sum
Anomalous	/	107	108	215
Normal	645	107	108	860

4 Proposed Approach

4.1 Experimental Setup

Multiple GAN-based approaches in the field of deep OCC anomaly detection have been published, i.e. AnoGAN [8], GANomaly [9], Skip-GANomaly [10], or DAGAN [14].

¹ The datasets may be requested from the corresponding author.

Others have investigated performance differences between the different architectures by investigating the performance on different datasets [4, 9, 10, 14]. But a general statement about the best performing architecture could not be derived. For this work, GANomaly and Skip-GANomaly are used. Firstly, they have achieved some of the best results in several comparative studies and secondly, the program code was provided by the authors via a source code host for reproducibility.

GANomaly. Akcay et al. [9] developed GANomaly. An Autoencoder is used as the generator. The encoder uses Leaky ReLU, Convolutional layers, and batch normalization and the decoder uses ReLU, Transposed-Convolutional layers to reconstruct the original image. Then, both the generated and the original image are mapped again into the latent representation using an encoder network. For training, the Adversarial Loss and the Contextual Loss are formed. The Adversarial Loss is used to improve the reconstruction abilities during the training. To explicitly learn this contextual information and thus capture the underlying data structure of the normal data instances, the $L1$ -norm is applied to the input and the reconstructed output. This normalization ensures that the model is able to produce contextually similar images to normal data instances. The third and last part of the training function is the encoder loss. The encoder loss aims to minimize the distance between the feature representations from the input image and the generated image. The higher-level training function is ultimately composed of a weighted sum of the three distinct sub-training functions.

To identify anomalous data instances with the model in the test phase, an anomaly score is calculated. It is based on the reconstruction error, which measures the contextual similarity between the real and the generated image, as well as the similarity of the latent representation of the real and the generated image. Depending on the threshold chosen, sample data instances are classified as normal or anomalous.

Skip-GANomaly. Skip-GANomaly [10] is an improved version of GANomaly. Main modifications are added skip connections between the encoder and decoder network. Due to the direct information transfer between the layers, both local and global information is preserved and thus an overall better reconstruction of the input data is possible.

CAE. The CAE used in this work consist of the decoder core network of the Skip-GANomaly architecture. Therefore, our CAE uses likewise skip connections between each down-sampling and up-sampling layer. In total the architecture consists of five blocks with each having a Convolutional and batch-normalization layers as well as Leaky ReLU activation function. With a symmetrical setup the outputted latent representation is up sampled back to the original dimension and the image is reconstructed. For calculating the reconstruction error an $L1$ loss between the input and the reconstructed output is used.

CNN. We used a Convolutional Neural Network to create a benchmark and compare the OCC methods with supervised methods. We used an 18-layer ResNet [15] architecture as a binary classifier with one class being the anomalous and one class being the normal images.

4.2 Experimental Results

Firstly, we investigate the performance of the three aforementioned anomaly detection methods on the dataset. Anomaly scores are used to classifier a sample depending on a chosen threshold, see Fig. 2. We use the receiver operating characteristic curve (ROC-curve) to evaluate the performance of a classifier independent from the threshold. In the ROC-curve, the True Positive Rate (TPR) or recall is mapped above the False Negative Rate (FNR) by forming these metrics for different thresholds (yellow values in the right figure). The area under the ROC-curve (AUC) is used as a measure of the performance of the classification model.

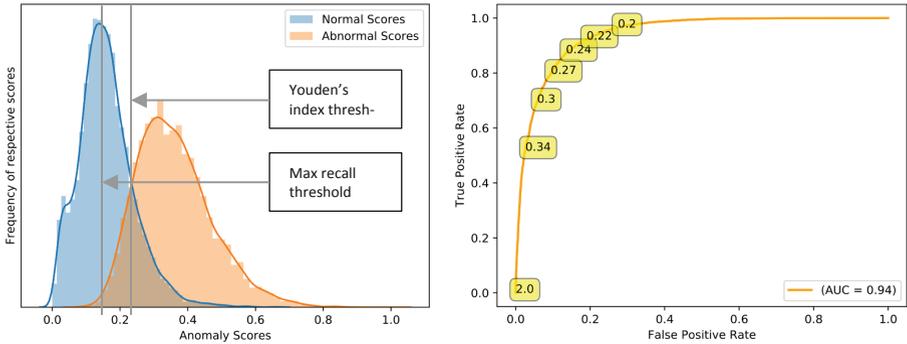


Fig. 2. Left: Histogram of normalized anomaly scores for normal and anomalous samples from the validation dataset. The two methods for threshold selection are plotted. Right: Corresponding receiver operating characteristic curve (ROC-curve) used to calculate the area under the curve (AUC). Yellow: Thresholds.

We conducted a parameter search and evaluated combinations of learning rates ($2e-2$, $2e-3$, $2e-4$), batch sizes (32, 64, 128), and input image sizes (32^2 , 64^2 , 128^2). The models are trained with the training data set and evaluated with the validation data set respectively, see Table 1. The highest AUC values for each method and dataset are listed in Table 2.

Table 2. Area under the ROC curve (AUC) for three anomaly detection methods.

Method	GANomaly	Skip-GANomaly	CAE
AUC	0,771	0,966	0,973

High AUC results indicate that the anomaly scores for the two classes differ significantly. Best results are achieved with the CAE. The GANomaly method performs worst and is not considered further in this work.

4.3 Threshold Selection and Comparison with Supervised Learning

While the AUC is a measure for the overall efficiency of a binary classification model, the choice of threshold is decisive for the use of the model in an application. We use two approaches to determine the threshold value, (a) Youden’s index and (b) maximization of recall. Both approaches are plotted qualitatively in Fig. 2. The Youden’s index, see Eq. 1, is defined for all points of the ROC curve, and the maximum value is used as a criterion for selecting the threshold.

$$J = \max(\text{recall} + \text{specificity} - 1) \quad (1)$$

When using the Youden’s index, the assumption is made that recall and specificity are of equal importance. When it is more important that all anomalous samples are found, the threshold can be selected by maximizing the recall. Among the thresholds with the highest recall, the one with the highest specificity is selected.

Table 3. Threshold selection.

Method	Skip-GANomaly		CAE		CNN
	Youden’s index	Max recall	Youden’s index	Max recall	
Accuracy	87.04%	88.89%	92.59%	87.96%	93.98%
Recall	80.56%	100%	91.67%	100%	88.89%
Specificity	93.52%	77.78%	93.52%	75.93%	99.07%

For the best performing model according to Table 2, both thresholds are set on the validation dataset. Subsequently, the anomaly scores for the test dataset, see Table 1, are determined with the model. Based on the selected thresholds, the quantitative measures for the assessment of a classifier are determined: accuracy, recall, and specificity, see Table 3. In this way, it is ensured that the threshold is tested on data that did not play a role in the determination of the threshold.

In order to benchmark the anomaly detection methods to a supervised method we trained a CNN. Therefore the 214 images in the validation dataset are split in an 80/20 ratio in a train and a validation dataset. We conducted a small hyperparameter study and evaluated combinations of three learning rates (1e-2, 1e-3, and 1e-4) and three batch sizes (8, 16, and 32). We trained for 25 epochs with an SGD optimizer. The best performing network was then tested on the test dataset with the results being presented in Table 3.

4.4 Discussion

The CAE performs on par with the CNN when Youden’s index is used for threshold selection. This is astonishing considering that the CAE has not seen anomalous samples during training. Therefore we believe that CAEs are reasonable alternative to CNNs when few anomalous samples are available. Threshold selection is key for the application of

anomaly detection methods. With both Skip-GANomaly and CAE methods we were able to train a classifier with a 100% recall or true positive rate while the specificity decreased to 77.8% and 75.9% respectively. These models could be used to presort acquired images reducing the overall scope of images that need to be examined.

Using GANs for anomaly detection has not achieved any added value for our use case. On the one hand, the trained GAN classifier performed worse or almost on the same level as the CAE. On the other hand the training effort in order to train GANs was significantly higher, because GANs are more difficult to converge, making the training process more challenging.

5 Conclusion and Outlook

In this work deep anomaly detection methods were applied to detect defects on the surface in the cavity of casting part using endoscopes. To counteract the challenge of insufficient anomalous training samples, one-class classification methods were investigated that rely solely on normal images for training. Two GAN based methods and one Convolutional Autoencoder were trained on our endoscopic dataset.

Results show high accuracy of 92.6% for the Autoencoder which outperformed the GAN-based approaches. When compared to supervised trained models, we could show that the Autoencoder performs on the same level as the trained CNN considering the small dataset used.

Our results are promising but do not indicate that OCC methods can be used without human support for the presented use case. We can identify two application potentials for the investigated anomaly detection methods. On the one hand, the models trained in this way can support a human worker during endoscopic inspection as an assistance system due to their already high correct classification rate. It should be emphasized that only a few defect-free components are required for data acquisition and that the models can therefore be trained and adapted to a task quickly and easily. A second application scenario is the partial automation of the inspection process. The demonstrated ability to pre-sort captured images can significantly reduce the number of images to be inspected by a human in the real process.

Acknowledgements. This research was funded by the German Federal Ministry for Economic Affairs and Climate Action under grant number ZF4736301LP9.

References

1. Peres, R.S., Guedes, M., Miranda, F., et al.: Simulation-based data augmentation for the quality inspection of structural adhesive with deep learning. *IEEE Access* **9**, 76532–76541 (2021). <https://doi.org/10.1109/ACCESS.2021.3082690>
2. Aust, J., Shankland, S., Pons, D., et al.: Automated defect detection and decision-support in gas turbine blade inspection. *Aerospace* **8**, 30 (2021). <https://doi.org/10.3390/aerospace8020030>

3. Martelli, S., Mazzei, L., Canali, C., et al.: Deep endoscope: intelligent duct inspection for the avionic industry. *IEEE Trans. Ind. Inf.* **14**, 1701–1711 (2018). <https://doi.org/10.1109/TII.2018.2807797>
4. Bergmann, P., Batzner, K., Fauser, M., Sattlegger, D., Steger, C.: The MVTEC anomaly detection dataset: a comprehensive real-world dataset for unsupervised anomaly detection. *Int. J. Comput. Vision* **129**(4), 1038–1059 (2021). <https://doi.org/10.1007/s11263-020-01400-4>
5. Muñoz-Marí, J., Bovolo, F., Gómez-Chova, L., et al.: Semisupervised one-class support vector machines for classification of remote sensing data. *IEEE Trans. Geosci Remote Sensing* **48**, 3188–3197 (2010). <https://doi.org/10.1109/TGRS.2010.2045764>
6. Chalapathy, R., Chawla, S.: *Deep Learning for Anomaly Detection: A Survey* (2019)
7. Zhou, C., Paffenroth, R.C.: Anomaly detection with robust deep autoencoders. In: Matwin, S., Yu, S., Farooq, F. (eds): *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, New York, NY, USA, pp. 665–674 (2017)
8. Schlegl, T., Seeböck, P., Waldstein, S.M., et al.: *Unsupervised Anomaly Detection with Generative Adversarial Networks to Guide Marker Discovery* (2017)
9. Akcay, S., Atapour-Abarghouei, A., Breckon, T.P.: *GANomaly: Semi-Supervised Anomaly Detection via Adversarial Training* (2018)
10. Akcay, S., Atapour-Abarghouei, A., Breckon, T.P.: Skip-GANomaly: skip connected and adversarially trained encoder-decoder anomaly detection. In: *2019 International Joint Conference on Neural Networks (IJCNN)*. IEEE, pp. 1–8 (2019)
11. Liu, J., Song, K., Feng, M., et al.: Semi-supervised anomaly detection with dual prototypes autoencoder for industrial surface inspection. *Optics and Lasers in Eng.* **136**, 106324 (2021). <https://doi.org/10.1016/j.optlaseng.2020.106324>
12. Tang, W., Vian, C.M., Tang, Z., Yang, B.: Anomaly detection of core failures in die casting X-ray inspection images using a convolutional autoencoder. *Mach. Vis. Appl.* **32**(4), 1–17 (2021). <https://doi.org/10.1007/s00138-021-01226-1>
13. Kim, J., Ko, J., Choi, H., et al.: Printed circuit board defect detection using deep learning via a skip-connected convolutional autoencoder. *Sensors (Basel)* **21** (2021). <https://doi.org/10.3390/s21154968>
14. Tang, T.-W., Kuo, W.-H., Lan, J.-H., et al.: Anomaly detection neural network with dual autoencoders gan and its industrial inspection applications. *Sensors (Basel)* **20** (2020). <https://doi.org/10.3390/s20123336>
15. He, K., Zhang, X., Ren, S., et al.: *Deep Residual Learning for Image Recognition*

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

