

Deep-learning-based computer vision system for surface-defect detection

Domen Tabernik¹, Samo Šela², Jure Skvarč³, and Danijel Skočaj¹

¹ University of Ljubljana, Faculty of Computer and Information Science,
Večna pot 113, 1000 Ljubljana, Slovenia
`domen.tabernik@fri.uni-lj.si`

² Kolektor Group d. o. o., Vojkova 10, 5280 Idrija, Slovenia

³ Kolektor Orodjarna d. o. o., Vojkova 10, 5280 Idrija, Slovenia

Abstract. Automating optical-inspection systems using machine learning has become an interesting and promising area of research. In particular, the deep-learning approaches have shown a very high and direct impact on the application domain of visual inspection. This paper presents a complete inspection system for automated quality control of a specific industrial product. Both hardware and software part of the system are described, with machine vision used for image acquisition and pre-processing followed by a segmentation-based deep-learning model used for surface-defect detection. The deep-learning model is compared with the state-of-the-art commercial software, showing that the proposed approach outperforms the related method on the specific domain of surface-crack detection. Experiments are performed on a real-world quality-control case and demonstrate that the deep-learning model can be successfully used even when only 33 defective training samples are available. This makes the deep-learning method practical for use in industry where the number of available defective samples is limited.

1 Introduction

Reliable visual inspection is one of the key elements of the production processes for ensuring an adequate quality of the manufactured products. Replacing the manual inspection with the automated machine-vision systems has been a trend for many years. By adopting the Industry 4.0 paradigm, the need for advanced machine-vision inspection systems even increases [10]. Increasing demands for customisation of the products, small product series, more complex products and constantly higher quality requirements aiming at zero-defect production call for more general, flexible and complex machine-vision systems.

Machine vision is a well-established engineering discipline that has led to numerous successful machine-vision applications in industrial production lines. A typical machine-vision system is composed of an adequate hardware and software to perform the inspection task that is integrated with the rest of the production line. An appropriate mechanism for positioning the object to be observed is required, and a choice of a suitable illumination and acquisition devices plays a very important role as well. The hardware part of the system should provide as good visual data as possible, so that the software part can reliably extract the required information about the quality

of the product. In a classical machine-vision approach to defect detection, an engineer would hand-craft the features adapted to a particular problem at hand based on his previous experience with similar problems. However, this leads to several weaknesses. The hand-crafted features tend to be quite problem-specific so when a new problem arises, an engineer would have to manually adapt the features to specifics of the new problem domain. Additionally, there are several problems that seem to be too difficult or impossible to be solved using the established hand-crafted solutions.

In this paper, we present a different approach to implementation of the software part of the machine-vision system. It is based on deep learning, which has proven to be very successful approach for solving numerous computer vision tasks. Modern computer vision systems heavily rely on deep-learning-based perception, which utilizes more advanced modelling capabilities. Compared to classical machine-vision methods, deep learning can directly learn features from low-level data, and has a higher capacity to represent complex structures. Such an approach addresses previously mentioned weaknesses of the classical machine-vision systems. It enables more general development of machine-vision systems, since they can be adapted to new problems by retraining the systems on the corresponding images, without the need of reprogramming the software. And, due to a better representation capacity, deep learning models are more successful in solving very complex problems as well.

Although the underlying principles are general, we will present a computer vision system for detecting a particular surface defect on a particular industrial product, an electrical commutator shown in Fig. 1. Commutators are an integral part of electrical motor, so the production of such an important component is completely automated. As such, each produced commutator undergoes through complete in-line optical inspection in the acceptable production cycle time. In this paper, we will present the part of the system that inspects the product for the most challenging visual defect.

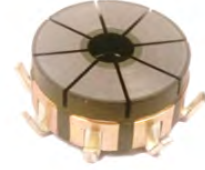


Fig. 1. Electrical commutator.

The remainder of the paper is structured as follows. The related work is presented in Section 2, with details of the optical-inspection system presented in Section 3 and evaluation in Section 4. The paper concludes with a discussion in Section 5.

2 Related work

A classical machine-vision approaches for defect detection follow more or less the same paradigm. Hand-crafted features are developed for the particular problem domain and classifiers, such as SVM, kNN, decision trees, or similar established computer vision techniques, are utilized to extract the discriminative information from the features. Various filter banks [12], histograms, wavelet transforms [5], morphological operations [6] and others techniques are used to hand-craft the appropriate features, since the classifiers are less powerful than deep-learning methods.

In contrast to the classical machine-vision approach the deep-learning approach directly learns the features. Several different works also employed deep-learning methods

for optical inspection, which is the main focus of this paper. The work by [7] showed that five-layer convolutional network can outperform classic hand-engineered features on image classification of steel defect. A similar architecture was used by [4] for the detection of rail-surface defects. A more modern network architecture was employed by [2]. They applied the OverFeat [9] network to detect five different types of surface errors and identified a large number of labeled data as an important problem for deep networks. Although they proposed to mitigate this using an existing pre-trained network, however, their method does not learn the network itself on the target domain and is therefore not using the full potential of deep learning.

Full network learning was performed in [11], where authors evaluated several deep-learning architectures with varying depths of layers for surface-anomaly detection. They applied networks ranging from having only 5 layers to a network having 11 layers, and, although they showed deep network to outperform any classic method, they demonstrated this only on synthetic dataset. Their method has also shown to be fairly inefficient as it extracted small patches from each image and classified each individual image patch separately. A more efficient network for explicitly performing the segmentation of defects was proposed by [8]. They implemented a fully convolutional network with 10 layers, using both ReLU and batch normalization to perform the segmentation of the defects. Furthermore, they proposed an additional decision network on top of the features from the segmentation network to perform a per-image classification of a defect's presence. This allowed them to improve the classification accuracy on the dataset of synthetic surface defects. As opposed to some related works [11,8], the proposed network is applied to real-world examples with small number of defective samples instead of using large number of synthetic ones.

3 Deep-learning-based optical inspection system

Commutators are an integral part of electrical motors and as such are often used in various mechanical systems. The production of such an important component is today completely automated, and is subject to a complete in-line optical inspection. In this section, we present the system for fully automated inspection of the compound-body part of the commutator, where a defect is manifested as a surface fracture of the material. The whole inspection process is depicted in Fig. 3 and consists of the automatic image acquisition process and the defect detection using the deep learning.

3.1 Hardware and image acquisition

The optical inspection is performed at the end of the production line, where the commutator enters into the automated machine for a complete optical inspection. The optical inspection machine, as depicted in Fig. 2, consists of 6 measuring stations that inspect 55 distinctive features with 23 different cameras. Manipulation of commutators inside the optical



Fig. 2. Optical inspection system

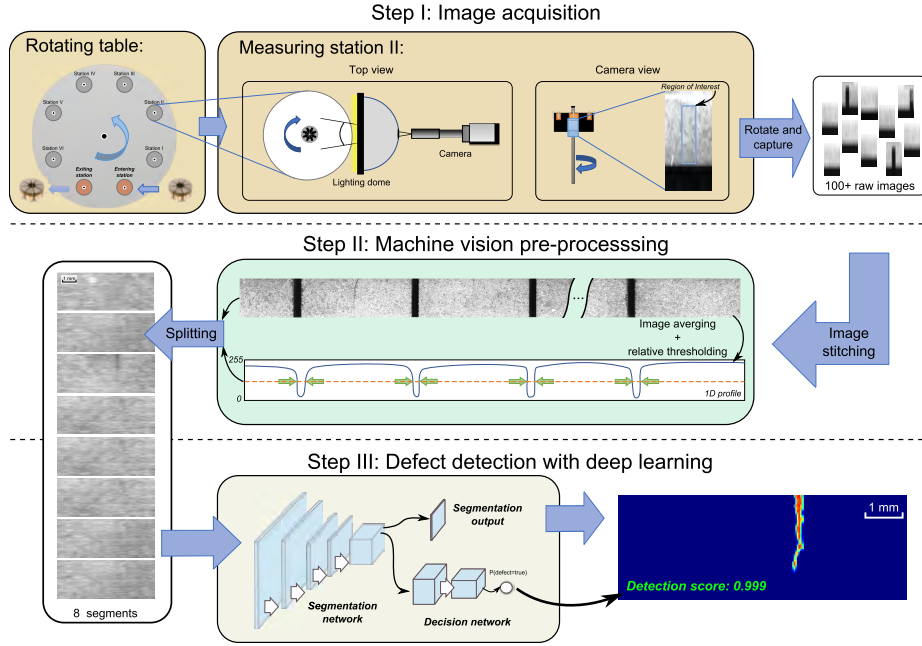


Fig. 3. Surface-defect detection system overview.

inspection machine is carried out by means of a rotating table with eight separate stations. Two stations are reserved for exit and entry point for the commutators, while six remaining stations perform active measurements of the item in a synchronous manner. Each station has 1.5 seconds of available time to complete its process. The inspected features range from 2D and 3D measurements of the physical properties of the product to various defects, such as missing material or porosity, mechanical damage on different parts of the commutator or the presence of residue from the production process.

In this work, we focus on the second active measurement station where the circumference of the compound-body of the commutator is being processed for the surface cracks. The whole process of the surface-defect detection consists of acquiring a high-resolution raw image parts of the circumference, combining them into full high-resolution image and then splitting them into eight segments to preserve only the relevant areas for final surface inspection.

In step one, the commutator is rotated in-place for 360 degrees to acquire the whole surface area of the compound-body. The image optics and cameras are synchronized with the rotation of the commutator and the high-power LED strobe-light source. The surface of the commutator is being illuminated with the dome light source that has an opening at the top for the camera. The camera is positioned perpendicular to the circumference of the inspected surface and is viewing the object in a lateral direction as is illustrated in Step I in Fig. 3. The camera observes a larger area of the image as shown in the camera view in Fig. 3, but only smaller region-of-interest area is used in

remaining steps. The smaller raw inputs are during the acquisition phase progressively combined to form the entire surface area. The final image size is 11700×500 pixels.

In step two, the system removes the dark notches in the image that represent the gaps in the material. Since the inspection is specialized for examining the defects on the surface of the material from which the commutator housing is made, the 11700×500 large image is divided into individual segments that do not contain these areas. To detect the horizontal edges of the material, we compute a 1D profile of the image by averaging the gray values in the lateral direction. This results in 11700×1 large vector. High values in this vector represent the material, while the low values represent the gaps. The abrupt change in those extreme values then represent the precise edge of the material, which we detect by applying a relative threshold level set at 30% to the computed 1D profile. The edges are used to horizontally cut the image to obtain 8 segments of size 1250×500 pixels, which contain only the material needed for the inspection.

3.2 Surface inspection

Next, each image segment is passed through a deep convolutional network that detects surface defects. We utilize a two-stage network architecture that follows the designs presented in [8], with the first stage implementing a segmentation network to localize the surface defect, while the second stage implementing the binary image classification. The overview is depicted in Fig. 4. The first-stage network is referred to as *the segmentation network*, while the second-stage network, as *the decision network*. We provide a brief description of both stages here and refer a reader to [10] for a more detailed description.

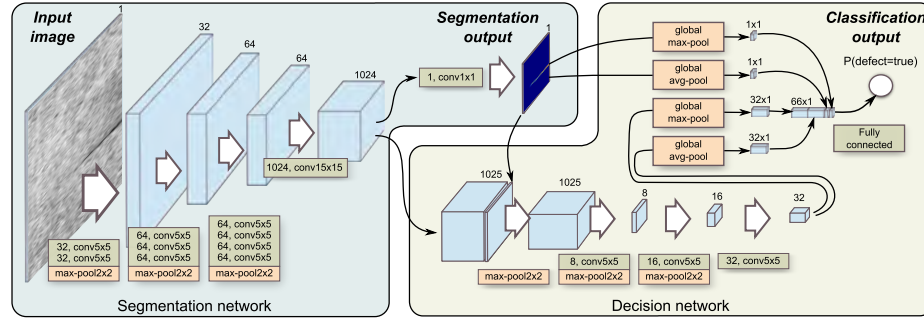


Fig. 4. The proposed architecture with the segmentation and decision networks.

Segmentation network The design of the segmentation network focuses on the detection of small surface defects in a large-resolution image. To accomplish this the network contains 11 convolutional layers and three max-pooling layers that each reduce the resolution by a factor of two. Each convolutional layer is followed by a feature normalization and a non-linear ReLU layer, which both help to increase the rate of convergence during the learning. Feature normalization normalizes each channel to a zero-mean distribution with a unit variance. The number of channels and kernel sizes

used in each layer are shown in Fig. 4. The final output mask is obtained after applying 1×1 convolution layer that reduces the number of output channels. The resolution of the output map is 8-times smaller than of the input image and is not interpolated back to the original image since this resolution suffices for the problem at hand.

Decision network The architecture of the decision network builds on the output from the segmentation network. As the input the decision network takes the output of the last convolutional layer of the segmentation network (1024 channels) concatenated with a single-channel segmentation output map. The input features are processed by a combination of a max-pooling layer and a convolutional layer that are repeated 3 times as shown in Fig. 4. This design effectively results in a 64-times-smaller resolution of the last convolutional layer than that of the original image. Finally, the network performs global maximum and average pooling, resulting in 64 output neurons. Additionally, the result of the global maximum and average pooling on the segmentation output map are concatenated as two output neurons, to provide a shortcut for cases where the segmentation map already ensures perfect detection. This design results in 66 output neurons that are combined with linear weights into the final output neuron.

Learning Both segmentation and decision networks are trained using the cross-entropy loss, however, the loss is calculated per-pixel for the segmentation network and per-image for the decision network. Both models are initialized randomly using a normal distribution. Networks are trained separately by first training only the segmentation network independently, then freezing the weights for the segmentation network and finally training only the decision network layers. This avoids the issue of overfitting from the large number of weights in the segmentation network. This is more important for the decision layers than for the segmentation layers due to limited GPU memory constraining the batch size to one. The segmentation layers are not effected by this due to pixel-wise loss that effectively increases the number of samples in a batch.

4 Evaluation

In this section, we present the evaluation of the proposed system. The whole system has been evaluated by first utilizing image acquisition and machine-vision pre-processing steps to collect the data. The collected data has then been used to train and evaluate the deep-learning model. Moreover, the deep-learning model has been compared against the state-of-the-art commercial product, the Cognex ViDi Suite [3].

4.1 Evaluation setup

Evaluation data To collect the data for the evaluation, fifty defective items were passed through the first two stages. This resulted in a total of 399 images; 52 with a visible defect and 347 without any visible surface defects. The defective samples were annotated with a pixel-wise segmentation mask that was additionally dilated with morphological kernel of size 5×5 . Several examples of the defective and non-defective surfaces are depicted in Fig. 5. The evaluation is performed with a 3-fold cross validation, while ensuring all the images of the same physical product are in the same fold.

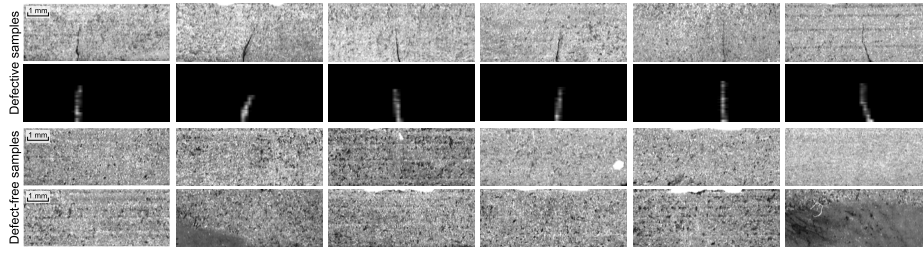


Fig. 5. Several examples of surface images with visible defects and their annotation masks in the top, and defect-free surfaces in the bottom.

Evaluation metrics Three different classification metrics were measured in the evaluation: (a) average precision (AP), (b) number of false negatives (FN) and (c) number of false positives (FP). Note, the positive sample is referred to as an image with a visible defect, and the negative sample, as an image with no defects. We focus mostly on the average precision, since it is more appropriate metric than FP or FN, as it accurately captures the performance of the model under different thresholds in a single value. The number of miss-classifications (FP and FN) are dependent on the specific threshold applied to the classification score. We report FP and FN at a threshold value where the best F-measure is achieved.

Implementation and learning details The network architecture was implemented in the TensorFlow framework [1], using stochastic gradient descend without momentum, a learning rate of 0.1, a batch size of one, and training for 100 epochs. Additionally, positive and negative samples were balanced during the learning by taking images with defects for every even iteration, and images without defects for every odd iteration.

Commercial software We compared the presented model against the Cognex ViDi Suite v2.1 [3]. The training and evaluation model was set to mirror the training and evaluation of the segmentation and decision network. This included using a gray-scale image, learning for 100 epochs and using the same train/test split. This configuration and hyper-parameter setup resulted in the best possible performance that we could achieve with the commercial software on this domain. For more details on the hyper-parameter setup for both the proposed deep-learning model and for the Cognex ViDi Suite the reader is referred to [10].

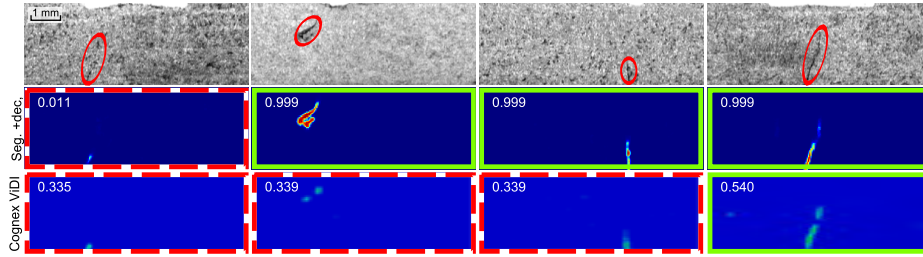
4.2 Results

The results are presented in Table 1. The segmentation and decision network outperformed the commercial product in all metrics. Observing the number of miss-classification at the ideal F-measure reveals that the segmentation and decision network missed only one defective sample, while the commercial product had 5 miss-classifications. Several miss-classified images are presented in Fig. 6. Both methods performed well on non-defective samples and did not detect and false positives.

Table 1. Comparison of the defect-detection methods.

	Average precision	False positive	False negative	FP at 100% recall
Seg. and dec. network	99.9 %	0	1	3
Cognex ViDi Suite [3]	99.0 %	0	5	7

Table 1 also shows the number of false positives that would be obtained if a zero-miss rate would be required, i.e., if a recall rate of 100% is required. These false positives then represent the number of items that would be needed to be manually verified by a skilled worker and directly point to the amount of work required to achieve the desired accuracy. Comparing the results in this metric reveals that the presented model introduces only 3 false positives at a zero-miss rate out of all 399 images. This represents 0.75% of all images. On the other hand, the commercial product achieved worse results, requiring the manual verification of 7 images.

**Fig. 6.** Examples of true-positive (green solid border) and false-negative (red dashed border) detections with the segmentation output and the corresponding classification (the actual defect is circled in the first row).

4.3 Computational cost

Due to requirements of the production process the processing time for the inspection of the whole item is restricted to 1.5 seconds. The processing time is composed of the raw image acquisition, cropping of the ROI and combining it into the whole image. This process takes 0.6 ms per raw image part, and 61 ms for all 101 parts. Next, the processing time also accounts for the image splitting, which accounted for 4 ms.

The computationally most demanding part is the actual defect detection with deep-learning. Using the same GPU as used for the learning (NVIDIA GTX 1080 TI 11GB) the processing of a single image takes 110 ms, thus resulting in 880 ms for all eight images and 945 ms for the whole system. Although, deep-learning-based defect detection is computationally most demanding, it is still efficient enough to meet the required criteria. Even with a more cost effective GPU, such as GTX 1080 8GB, the system can still manage to complete the task in the required time, taking 1112 ms for the defect detection on all eight images, and 1177 ms for the whole system.

Table 2. Computational cost for the individual processing stages.

<i>Per item processing time</i>	Image capture	Pre-processing and split	Defect detection	Total
GTX 1080 TI 11 GB	61 ms	4 ms	$110 \cdot 8 = 880$ ms	945 ms
GTX 1080 8 GB	61 ms	4 ms	$139 \cdot 8 = 1112$ ms	1177 ms

5 Discussion and conclusion

In this paper, we presented a complete optical inspection system for automated detection of surface defects on electrical commutator. We presented the image acquisition system, which captures the surface of the whole item and converts it into eight non-overlapping images using classical machine-vision processes. We then employed a more powerful deep-learning approach to detect surface-defects using a segmentation and decision networks. We evaluated the deep-learning approach on the problem of surface-defect detection where defects appear as fractures on the compound-body of the electrical commutator. The segmentation and decision network was demonstrated to achieve significantly better results than the related state-of-the-art commercial product, with only one miss-classification for the segmentation and decision network, and five miss-classifications for the commercial product.

The performance of the presented method was achieved by learning the network from only 33 defective samples. This indicates that the presented deep-learning approach is suitable for the studied industrial application with a limited number of detected samples available. The system has also proven to be ready for use in the industrial environment with the required manual inspection rate as low as 0.75% (three images out of all 399 images) when all defective samples need to be found.

Although the main innovation of the presented system is the deep learning approach that has been used for detecting the surface defects, a part of the developed system still uses the classical machine-vision techniques. In the pre-processing step the classical methods are used to stitch and split the images, therefore to produce high-quality images that are latter on used by the learning-based defect-detection method. This part is relatively easy to implement, and robust solutions have already been established, so there is no need to replace it with the data-driven approach. In principle, this could be done, but probably at a cost of a higher required number of training images, which could often be difficult to obtain. A good advice is therefore to use simple machine-vision methods in the pre-processing step to prepare the training data as well as possible, and to use them in the deep-learning approach to solve the more difficult part of the visual inspection problem. Such an approach provides the best results, and it is expected that a reasonable combination of both, classical machine-vision and data-driven learning-based approaches will very often be used in the future machine-vision systems.

Acknowledgements: This work was supported in part by the following research programs: GOSTOP program C3330-16-529000 co-financed by the Republic of Slovenia and the ERDF, ARRS research project J2-9433 (DIVID), and ARRS research programme P2-0214.

References

1. Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G.S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mané, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viégas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., Zheng, X.: TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems (2015), <https://www.tensorflow.org/>
2. Chen, P.H., Ho, S.S.: Is Overfeat Useful for Image-Based Surface Defect Classification Tasks? In: IEEE International Conference on Image Processing. pp. 749–753 (2016)
3. Cognex: VISIONPRO VIDI: Deep learning-based software for industrial image analysis (2018), <https://www.cognex.com/products/machine-vision/vision-software/visionpro-vidi>
4. Faghieh-Roohi, S., Hajizadeh, S., Núñez, A., Babuska, R., Schutter, B.D.: Deep Convolutional Neural Networks for Detection of Rail Surface Defects Deep Convolutional Neural Networks for Detection of Rail Surface Defects. In: International Joint Conference on Neural Networks. pp. 2584–2589. No. October (2016)
5. Ghazvini, M., Monadjemi, S.A., Movahhedinia, N., Jamshidi, K.: Defect Detection of Tiles Using 2D-Wavelet Transform and Statistical Features. *International Scholarly and Scientific Research & Innovation* **3**(1), 773–776 (2009)
6. Mak, K.L., Peng, P., Yiu, K.F.: Fabric defect detection using morphological filters. *Image and Vision Computing* **27**(10), 1585–1592 (2009). <https://doi.org/10.1016/j.imavis.2009.03.007>, <http://dx.doi.org/10.1016/j.imavis.2009.03.007>
7. Masci, J., Meier, U., Ciresan, D., Schmidhuber, J., Fricout, G.: Steel defect classification with Max-Pooling Convolutional Neural Networks. In: Proceedings of the International Joint Conference on Neural Networks (2012). <https://doi.org/10.1109/IJCNN.2012.6252468>
8. Rački, D., Tomažević, D., Skočaj, D.: A compact convolutional neural network for textured surface anomaly detection. In: IEEE Winter Conference on Applications of Computer Vision. pp. 1331–1339 (2018). <https://doi.org/10.1109/WACV.2018.00150>
9. Sermanet, P., Eigen, D.: OverFeat : Integrated Recognition, Localization and Detection using Convolutional Networks. In: International Conference on Learning Representations (2014)
10. Tabernik, D., Šela, S., Skvarč, J., Skočaj, D.: Segmentation-based deep-learning approach for surface-defect detection. *Journal of Intelligent Manufacturing* pp. 1–18 (2019). <https://doi.org/10.1007/s10845-019-01476-x>, <http://link.springer.com/10.1007/s10845-019-01476-x>
11. Weimer, D., Scholz-Reiter, B., Shpitalni, M.: Design of deep convolutional neural network architectures for automated feature extraction in industrial inspection. *CIRP Annals - Manufacturing Technology* **65**(1), 417–420 (2016). <https://doi.org/10.1016/j.cirp.2016.04.072>, <http://dx.doi.org/10.1016/j.cirp.2016.04.072>
12. Zheng, H., Kong, L.X., Nahavandi, S.: Automatic inspection of metallic surface defects using genetic algorithms. *Journal of Materials Processing Technology* **125-126**, 427–433 (2002). [https://doi.org/10.1016/S0924-0136\(02\)00294-7](https://doi.org/10.1016/S0924-0136(02)00294-7)