

Comparison of the performance of innovative deep learning and classical methods of machine learning to solve industrial recognition tasks

K. Anding^{*a}, L. Haar^a, G. Polte^a, J. Walz^a, G. Notni^a

^a Ilmenau University of Technology, Gustav-Kirchhoff-Platz 2, 98693 Ilmenau, Germany

* katharina.anding@tu-ilmenau.de

ABSTRACT

Artificial intelligence and machine learning are becoming increasingly important in science and society. In image processing, they are mainly used for object classification. The aim of this paper is the comparison of classical supervised machine learning methods with innovative deep learning (DL) approaches in terms of performance, which is described by the calculated accuracy. Classifiers of different characteristics are used. These are the Support Vector Machines, Random Forest, k-Nearest-Neighbor, and Naive Bayes. They are compared to two not pre-trained and four pre-trained neural networks. The former neural network are based on LeNet, the second ones include AlexNet, GoogleNet and ResNet provided by Matlab as well as a pre-trained neural network provided by MVTec HALCON. Comparisons were made using the recognition rates achieved with five real data sets from industrial applications. The results showed that not pre-trained neural networks produce worse results than classical classifiers with the given small amounts of data for training. On the other hand, the pre-trained networks achieved surpassing recognition rates. However, if there are features that describe the classes very well, the recognition performance of classical machine learning methods little differs from that of deep learning algorithms.

Keywords: artificial intelligence, machine learning, deep learning, pre-trained neural network, Convolutional Neural Network

1. INTRODUCTION

Image processing can be automated with the help of artificial intelligence and, in particular, machine learning methods. These topics have long been a focus of science and society. The considerable increase in interest is mainly due to the development of new algorithms and increasingly powerful computer technology. Therefore, these technologies are currently attracting a great deal of attention, which can also be attributed to the recent successes with deep neural networks (deep learning). These have some advantages compared to classical classification methods. So, for example some steps of the image processing chain can be omitted, like feature extraction and feature selection. However, they also show disadvantages such as long processing times, many parameters to be set and no traceability of the decision-making process. Deep Learning algorithms are often called as “black box approach”¹, which can be viewed in terms of its inputs and outputs but without full knowledge of its internal workings.

In order to investigate the differences in recognition performance, classical supervised machine learning methods and innovative deep neural networks will be compared in the context of this paper. The evaluation of the methods is carried out using several data sets from industrial applications with different characteristics and the achieved recognition rate (accuracy). To make the investigations as comprehensive as possible, classifiers with different characteristics, complexity and sensitivity were selected. These are Support Vector Machines (SVM), Random Forest, k-Nearest-Neighbor, and Naive Bayes. The Convolutional Neural Networks (CNN) are used as DL methods. A CNN represent a special case of standard neural networks and is especially suited for image processing tasks. Transfer Learning enables the training of a network with some data and expand it by the own dataset. The pre-trained neural networks AlexNet, GoogLeNet and ResNet provided by Matlab as well as a pre-trained neural network provided by MVTec HALCON are used.

2. STATE OF THE ART

Comparative studies on DL and classical machine learning methods are commonly performed. This is done, for example, with regard to predicting the development of electricity prices². Support Vector Machines and deep neural networks are also compared with respect to the processing of remote sensing images³. Other studies are dedicated to the comparison of natural language understanding⁴ or four different neural networks⁵. The performance is also compared with regard to the evaluation of wrist accelerometer⁶, with the focus on DL and random forest. This paper, on the other hand, will focus on the comparison of recognition and classification performance of deep neural networks and classical methods of machine learning in the processing of images from industrial applications.

2.1 Classical methods of machine learning

Classical methods of machine learning are used in image processing for the classification of objects. These are compared in different studies, such as by Fernández-Delgado et al.⁷. One of the most powerful classifiers is the **Support Vector Machine (SVM)**. If a linear separability of classes is given, a dividing line can be drawn in the two-dimensional feature space, or a linear hyperplane in the multi-dimensional feature space, which delimits the classes and enables classification of unknown objects⁸. Such linear separability is not always given. If there is a non-linearly separable data set, the SVM uses the so-called kernel trick. The data is transferred to higher-dimensional feature spaces using a kernel function until linear separability is possible⁹. Subsequently, the linear separation plane is determined and the data is transferred back to the original feature space. Now the dividing plane no longer appears linear and is able to separate the classes from each other¹⁰. In order to obtain satisfactory results with the SVMs, the parameters must be optimized. The difficulty is that the parameters have to be redefined for each new application. In addition to SVM, **Random Forest** is also a powerful classifier. It consists of a large number of decision trees, a so-called decision forest, and thus belongs to the ensemble methods¹¹. This classifier achieves very good results in a short time, even on complex, highly nonlinear recognition tasks^{12, 13}. Furthermore, it has only a few parameters to be set. The **k-Nearest-Neighbor** classifier uses the nearest objects and a distance measure (e.g. euclidean distance) in the feature space to determine the class membership¹⁴. The **Naive Bayes** classifier belongs to the group of statistical or probability-based classifiers¹⁵. A detailed description of the methods can be found in the cited publications.

All these classifiers are part of supervised machine learning algorithms and they work on the basis of feature vectors. This means that initially significant image regions must be segmented, features of the regions have to be calculated and, if necessary, transformed. The steps required to process a classification task in image processing are shown in Figure 1¹⁶.

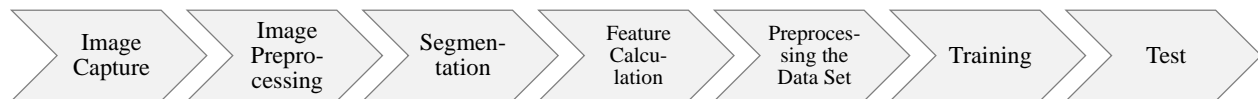


Figure 1: Image processing and machine learning chain¹⁶

So classic classifiers are based on manually selected image features (shape, color, texture and spectral features), which can be physically described and extracted from the image, e. g. circularity as a feature of form, average gray value as a color feature, Laws features as texture features, and statistical characteristics extracted from the point spectrum of a hyperspectral image. The shape, color and texture features are extracted for each existing image channel (e. g. RGB) and made available for classifier training in a summarized physically describable feature vector of limited dimension (usually several hundred features per object example). The image features are selected on the basis of the characteristics of the underlying recognition problem and examined with regard to their discriminant ability, relevance and redundancy (feature selection) before they are used as the basis for training of the classification algorithm. Due to the limited feature set, in contrast to the DL method, significantly fewer data set instances are required in the training data set in order to achieve sufficiently good recognition results. An optimal application-specific selection of the image features is decisive for the achievable quality and generalization capability of the classification algorithm. In contrast to CNN, the selection of suitable image features is not the responsibility of the classification algorithm but of the recognition expert. Here it becomes clear that manually designed feature extraction can cause important features to remain unnoticed if unsuitable feature selection takes place, which can impair the performance of the following classic classifier¹⁷.

2.2 Convolutional Neural Networks – a method of DL

The human brain is capable of processing information in an impressive way. Artificial neural networks are supposed to emulate this remarkable ability. Biological neural networks consist of neurons which are connected by synapses¹⁵. Artificial neural networks are also built according to this model. They consist of an input layer, one or more hidden layers and an output layer. The presence of a large number of layers is referred to as deep neural networks or DL¹⁰. Deep neural networks are present in a wide variety of applications. There are different implementations with regard to structure and sequence. Each of these has a special suitability for different tasks¹⁸. The Feedforward Neural Network is used in particular for computer vision and speech recognition. Radial Basis Function Neural Network can be found especially at power restoration systems. The Recurrent Neural Network is implemented in text to speech applications. The Convolutional Neural Network (CNN) is especially suited for image processing tasks. This is due to the ability to process data arranged in matrices. Since the pixels within an image form such two-dimensional matrices, they can be directly processed¹⁹. In contrast to classical machine learning methods the intermediate steps of the image processing chain, such as feature calculation and preprocessing the data set, are not necessary by using CNNs. They are done automatically within the algorithm. This means that feature extraction is already integrated in the algorithm of CNNs and takes place in implicit form directly within the layers of the neural network¹⁷.

The input data is transformed many times and thus complex features are developed. First, only edges are recognized, later individual parts, which are finally combined into an entire object and viewed as a whole. Figure 2 shows a possible structure of a CNN¹⁶. It consists of an input layer, a series of alternating successive convolutional layers and pooling layers, which calculate and clean up the features, followed by a fully connected layer, which corresponds in its structure to a multilayer perceptron and is used for classification. The transfer of the results is done through the output layer.

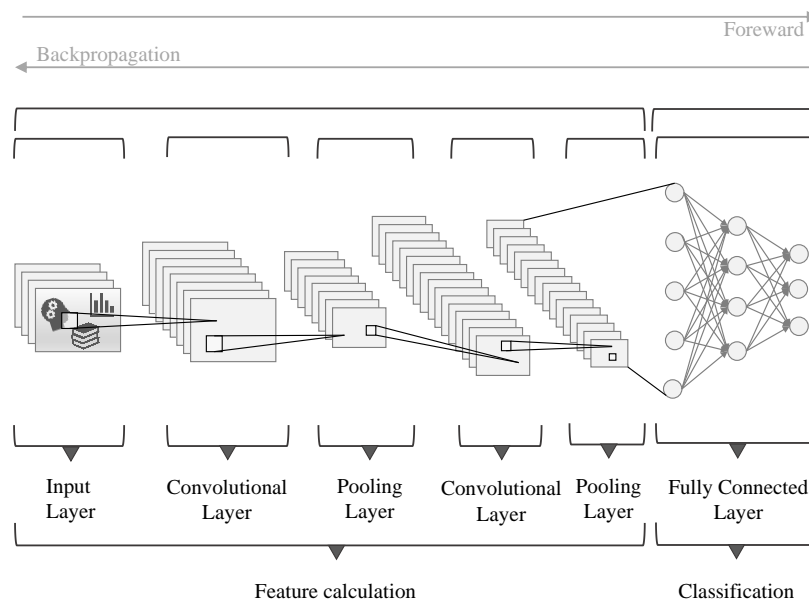


Figure 2: Exemplarily illustration of a Convolutional Neural Network¹⁶

With a sufficient number of layers, CNNs are able to create very complicated functions of the input. Often these are insensitive to large but not important changes, such as background, lighting or environment, but respond to small details that are crucial to decision-making. They are also able to process non-linear data sets, are flexible and can be well adapted to existing problems²⁰.

Nevertheless, there are some disadvantages and restrictions in the application of these algorithms. Deep neural networks require an extremely large amount of data for training in order to develop a good generalization capability and thus deliver good results²¹ (it depends on the intrinsically calculated several thousands of features / curse of dimensionality). Here, the number of examples can rise to millions. However, obtaining such a large number of pre-classified images is time-consuming and cost-intensive. The so-called transfer learning can help to solve this disadvantages²². Here a network with arbitrary pictures is pre-trained and extended by the own images of the selected industrial applications.

Even if many training objects are available, deep neural networks remain susceptible to overfitting. Due to the large amount of data to be processed, the training also takes up a lot of time and can take days or even weeks¹⁰. Very high computing power is required for the effective use of these methods²². For this reason, such applications should be processed by the Graphics Processing Unit (GPU) instead of the Central Processing Unit (CPU) for saving time¹⁰. In addition, many parameters have to be set and optimized, e.g. initial weights, activation function, learning rate, batch size²³. According to Bengio²¹, the quality of the results depends on the initial values. Since neural networks are black boxes, the decision-making process is not user-comprehensible. Low level features of the CNN show certain similarities to classically used features (e. g: edge filters in certain orientations and scales or certain color characteristics²⁴), but in contrast to classical image features low level features cannot be described physically with comparable mathematical formulas. However, there are approaches to investigate the fundamental interpretability of intrinsic features of CNNs or the basis of the underlying decision making of CNNs^{24, 1, 25}. Furthermore, it is difficult to predict how a change in architecture or initial parameters will affect the results. This fact makes the design of a network much more difficult¹⁴. An extremely serious disadvantage is that DL methods can be fooled and the results can be influenced, leading to critical misclassification²⁶⁻²⁹.

3. INVESTIGATIONS

3.1 Data sets

Five different real data sets from industrial applications were used in the investigations of this paper. The first three data sets consist of light scattering images. These represent reflective, industrially produced surfaces. Scattered_Light_1_3Cl contain three classes: one class includes images of surfaces without defects and the other two classes represent scratches or point defects. Scattered_Light_2_4Cl and Scattered_Light_3_4Cl have an additional defect class (hybrid defects) and thus four classes in total. It should be mentioned that Scattered_Light_2_4Cl consists of images of various materials and processing types, which is why it has a large innerclass variability and a small interclass variability (definitions of innerclass and interclass variability are given in Anding³⁰). This data set contains an outlier.

Furthermore, a data set with images of metal surfaces is used with a total of 274 objects examples (instances). In the past we have presented a method for surface defect recognition of metal surfaces using the example of analyzing a counter sunk drilling at heat-treatable steel. The classification was done using color and texture features as input parameters for classical machine learning algorithms like a SVM. This is a small image data set, which is why there may be a low generalization capability of the classifier. We discriminating between the classes: defect-free metal surfaces and two defect classes: longitudinal rills and chatter marks. This data set also contains an outlier.

The Autopetrography data set is based on a developed method for automatic recognition of mineral aggregates to solve the analysis for all petrography classes according to legal requirements. This data set consists of 18.596 objects and is therefore the largest used data set. Because of the similarity of the external appearance of some objects, the interclass variability is quite low. This leads to an overlapping between several classes in feature space and thus to classes, which are difficult to separate. In Table 1, all data sets are listed with information about number of objects, number of features and number of classes. Figure 3, Figure 4 and Figure 5 show example images of the different data sets.

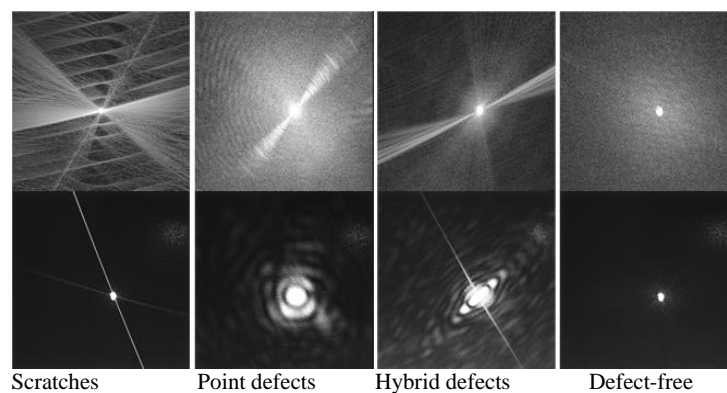


Figure 3: Example images of light scattering image data set.

Some investigations were done based on the given data sets in the last years in our research group by using classical machine learning algorithms 27–30. These investigations were complemented with an analysis of further optimized classical classifiers and innovative DL approaches. Their results are compared and discussed in this paper.

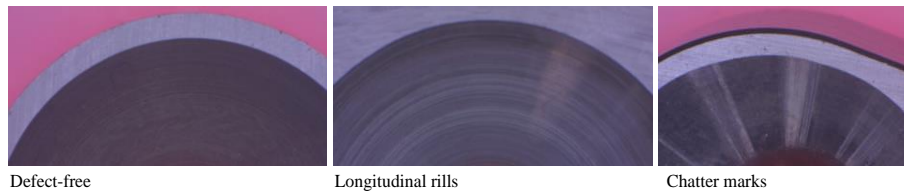
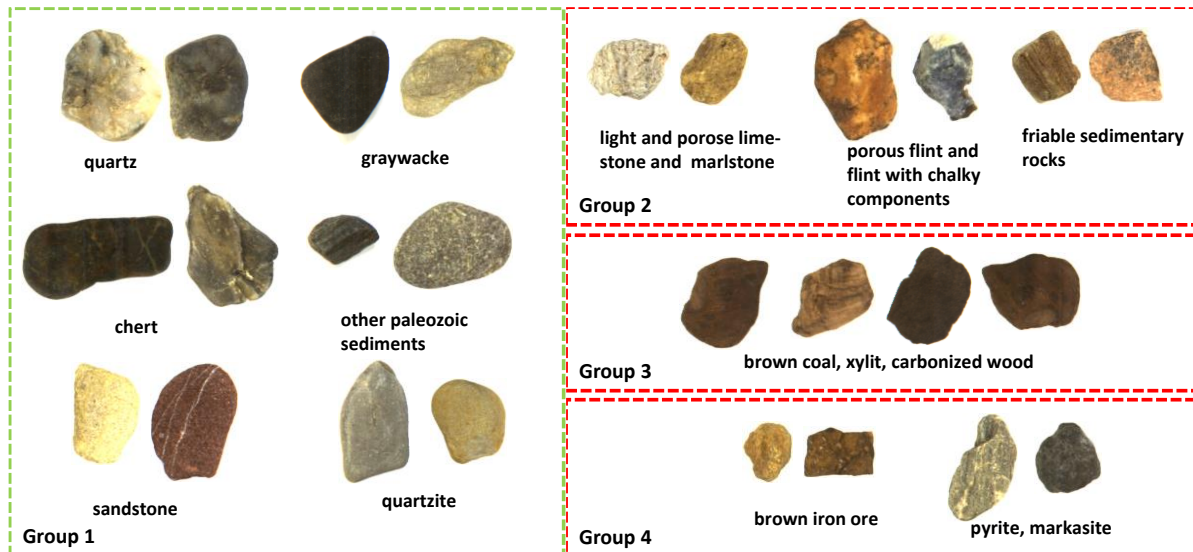


Figure 4: Example images of metal surface data set.

Table 1. Overview of used data sets.

Name	Number of Objects	Number of Features	Number of Classes
Scattered_Light_1_3Cl	300	182	3
Scattered_Light_2_4Cl	719	182	4
Scattered_Light_3_4Cl	1193	182	4
Metal Surfaces	274	123	3
Autopetrography	18.596	234	4



Group 1: uncritical for the production of beton
 Group 2 - 4: critical for the production of beton

Figure 5: Example images of Autopetrography data set.

3.2 Classical supervised machine learning methods

In order to include as many different processes as possible, the investigations were made with four classical classifiers with different characteristics. The k-Nearest-Neighbor classifier is distance-based and was used with a number of three considered objects (neighbors), lying next to the unknown one. Random Forest builds a decision forest with 100 trees. Naïve Bayes belongs to statistical classifiers and was used with the default settings. For the investigations with the SVM, the choice fell on an rbf-kernel. Furthermore, it was subjected to parameter optimization in Matlab with the help of "Pattern Search". In this context, the recognition rate was maximized for the parameters C and γ ³⁵. The LibSVM-library in Matlab and Weka was used for this purpose. Otherwise, the software KNIME and especially the Weka-Plug-In were used. Some of the classification methods rely on the pre-processing of the feature expressions within the data set, in order to obtain good results. Before using k-Nearest-Neighbor, the features have to be normalized so that the values lie

within an interval of [0;1]. This serves to prevent the distance measure from being influenced by different orders of magnitude between the characteristics. Normalization is also performed before using the SVM. In literature, it is recommended to discretize the feature values for Naive Bayes. Only the decision tree Random Forest receives the data in unchanged form. Table 2 summarizes the above.

Table 2. Overview of used classifiers.

Classifier	Data pre-processing	Software
k-Nearest-Neighbor	Normalize	KNIME
Random Forest	None	Weka-Plug-In in KNIME
Naive Bayes	Discretize	Weka-Plug-In in KNIME
Support Vector Machines (SVM)	Normalize	LibSVM in Matlab and Weka-Plug-In in KNIME

3.3 Innovative DL methods

The Convolutional Neural Networks (CNN) are used as DL methods in this paper because they are especially suited for image processing tasks. In the following, the used methods are examined in more detail.

LeNet in KNIME

KNIME enables the use of deep neural networks by integrating "DeepLearning4J", "Keras" and "TensorFlow" whereby the former library is used for investigations with this method³⁶. It is based on the LeNet architecture³⁷ and consists of two convolutional layers, two pooling layers and is terminated with a fully connected layer. Rectified Linear Unit (ReLU) is used as the activation function. The network is not pre-trained and was therefore trained individually for each data set. No data augmentation was carried out either. Before the training, the images were subjected to some pre-processing measures, such as reduction of the channels, reduction of the picture size and conversion into a square format. Furthermore, the pixel values were normalized and then standardized. In the literature, this procedure is also recommended for the use of feature vectors in combination with deep neural networks. In order to improve the results, a complex optimization can be carried out. Thus, the architecture of the network and thus the number and constellation of hidden layers, the parameter values or the activation function could be adapted. Such a specially created and trained network was investigated using the Metal Surface data set³⁸.

Adapted network to the Metal Surfaces data set without pre-training

In the previous section, the deep neural network adapted to the Metal Surfaces data set by the authors Anding, Kuritcyn and Garten, was already mentioned³⁸. The published results will be used for comparative investigations in the context of this paper. Theano, Lasagne and nolearn were used for the realization of the network. Before the application, the images were pre-processed, reducing the number of channels to one and reducing the image size significantly, which leads to lower computational effort. The network architecture is based on LeNet³⁷, whereby a different number and constellation of layers was investigated. Exponential Linear Unit (ELU) was used as activation function. In part, a data set extension was performed. The first problem we have faced in our investigations with CNN on Metal Surface data set was the small size of the given data set. Usually, neural networks require big datasets to achieve good results. There are many techniques to expand data sets. This process is called data augmentation and uses operations like scaling, translation, rotation, flipping, adding noise, changing lighting conditions or using perspective transformation. At first, we flipped random chosen images³⁸. On the next stage we added also random rotation, horizontal and vertical shift for a few pixels and zooming. For training our network we also used small batches from the original dataset images. These batches were augmented on the fly by an implemented generator. It generates new samples at "no cost" and allow a training of the CNN without pre-training. In chapter - 4. Experimental Results – this CNN is called "Adopted CNN".

Pre-trained CNN in HALCON

The software MVTec HALCON 18.05 contains two pre-trained networks, which were trained using pictures from industrial applications. They can be extended by images of own data sets. With such a procedure, the problem of too few data quantities is counteracted. Depending on the problem, nevertheless, hundreds to thousands of images per class must be present according to the reference manual of HALCON²². In this paper, the "pretrained_dl_classifier_enhanced.hdl"

is used. The aim of the investigations is to analyze the added value of this pre-trained network. So far, it can only be executed using the GPU. The parameters preset by HALCON for batch, epochs and learning rate proved to be appropriate and were used. The batch size is thus 64 and the number of epochs 16. An adaptive learning rate is used, which starts at 0.001 and is reduced by 1/10 every four epochs. However, the exact structure of the neural network cannot be seen, which is why no information about the number and sequence of the layers is given by MVTec, the developer of HALCON.

Pre-trained CNNs in Matlab

Matlab also offers pre-trained CNNs. The difference to HALCON, however, is the use of images of different everyday objects, such as keyboards, pens or coffee cups, as well as many animals taken from ImageNet, instead of images from industrial applications. The neural networks AlexNet, GoogLeNet and ResNet-101 are used for the investigations in this paper. They are pre-trained with over one million images, which are divided into 1,000 categories, and can be supplemented with own data ³⁹.

AlexNet – This network architecture was created by Alex Krizhevsky and consists of eight layers, of which five are convolution layers with max-pooling layers and three are fully connected layers ⁴⁰. ReLU is used as the activation function. The images transferred to the network must have a size of 277x277 pixels ⁴¹. A batch size of 10 was selected for the investigations and the number of epochs was set to 6. The initial learning rate is 0.0001.

GoogLeNet – This network is based on the LeNet-architecture and uses a so-called inception ²⁶. Here filters of different sizes are used in the convolution layers. According to the developers, this network makes better use of computational conditions. It consists of 22 layers. ReLU is used as the activation function. The prescribed image size is 224x224 pixels. Again, the batch size is set to 10 and the number of epochs is set to 6. The initial learning rate differs minimally from the one used with AlexNet. Here it assumes a value of 0.0003.

ResNet-101 – This Residual Neural Network (ResNet) consists of 101 layers and is able to skip one or more layers ^{42, 43}. This should facilitate learning and avoid a decrease in the recognition rate with the increase of the depth of the network. In some deep neural networks, such a behavior has been observed. The images must have a size of 224x224 pixels. A value of 10 was selected for the batch size, 6 for the number of epochs and 0.0003 for the initial learning rate.

4. EXPERIMENTAL RESULTS

For a comparison of the results of the procedures described above the achieved average recognition rates (RR) and standard deviations (Stdev) were used. These were determined with the help of cross-validation. Table 3 summarizes the results.

In order to deliver satisfactory results, deep neural networks require an enormous amount of data. This is not possible with the data sets used, which is why high recognition rates could not be achieved without the use of pre-trained networks. Therefore, the CNN from KNIME with a result of 80.47 % only surpasses the simple classifier Naïve Bayes in the case of the Scattered_Light_3_4Cl ³⁶. Otherwise, the recognition rates are lower than those of the classical methods. To counteract this, the architecture, the number and constellation of hidden layers, the parameter values or the activation function could be adapted. Such a specially created and trained network was examined by Anding, Kuritcyn and Garten using the Metal Surface data set, whereby a maximum recognition rate of 83.5 % was achieved ³⁸ (see Adapted CNN in Table 3). This exceeds the performance of the simpler classifiers k-Nearest-Neighbors and Naive Bayes but does not match Random Forest or SVM. The adopted CNN is not pre-trained with other big data sets before subsequent training with Metal Surface data set with only a total of 274 objects (instances), which were captured by a color line scan camera. In the past our first tries, using the given data set without data augmentation, showed a recognition rate of only 48.0 % ³⁸. It was an unacceptable performance and so we used data augmentation to expand the given image data set. Nevertheless data augmentation is not sufficient to reach an adequate amount of training instances, which is required for training of CNNs. So the unsatisfactory final result of 83.5 % is well justified by this given fact.

Table 3. Comparison of the recognition performance of classical classification methods with DL methods. ¹⁶

RR and Stdev data in percent	Data sets									
	Scattered_Light						Metal Surfaces		Autopetrography	
	1_3Cl		2_4Cl		3_4Cl		RR	Stdev	RR	Stdev
Type of classifier	RR	Stdev	RR	Stdev	RR	Stdev	RR	Stdev	RR	Stdev
k-Nearest-Neighbor	85.00	3.93	64.12	5.15	81.81	2.07	74.80	4.51	86.94	0.58
Naïve Bayes	76.67	4.44	50.35	5.02	75.86	2.50	78.46	2.38	73.10	0.57
Random Forest	92.67	3.06	71.78	4.14	84.92	2.39	86.49	0.68	92.90	0.52
LibSVM	89.67	5.62	70.93	6.32	85.25	2.03	85.03	1.32	93.05	0.14
SVM (SMO) (Haralick features)	/	/	/	/	/	/	98.83	2.44	/	/
CNN in KNIME	69.60	6.08	46.07	0.91	80.47	0.21	59.80	0.90	68.80	0.00
Adopted CNN	/	/	/	/	/	/	83.50	0.87	/	/
HALCON-Network	93.00	5.26	72.70	5.09	93.97	2.31	93.45	3.25	93.34	0.34
AlexNet	94.47	5.26	78.06	5.65	93.63	2.17	90.46	2.62	90.90	0.97
GoogLeNet	96.00	3.78	78.60	5.21	93.80	2.75	95.22	3.37	92.58	0.78
ResNet-101	96.67	3.51	77.05	4.40	94.22	2.03	96.32	2.30	94.11	0.95

The situation is different with the pre-trained networks from HALCON or AlexNet, GoogLeNet and ResNet-101 from Matlab. The results of the HALCON-Network and Random Forest are very similar for the data sets Scattered_Light_1_3Cl and Scattered_Light_2_4Cl. However, the results of both the classical methods and the deep neural networks are accompanied by very high standard deviations, which indicates an unstable process. This could be caused by the small number of objects in Scattered_Light_1_3Cl and thus a very small training data set, whereby the classifier cannot achieve a good generalization capability. With Scattered_Light_2_4Cl it could be due to the heterogeneity of the data set. The average recognition rates are higher with GoogLeNet and ResNet-101 than with Random Forest, which has the best result among the classical methods.

The performance of the deep neural networks in Scattered_Light_3_4Cl and Metal Surfaces clearly exceeds that of the other classifiers in Table 3 with the exception of SVM (SMO) and the use of Haralick features. LibSVM achieves 85.25 % with the first data set, whereas ResNet-101 achieves 94.22 %. This makes a difference of about 9 %. For Metal Surfaces and the classical features (without Haralick) and classical methods, the highest average recognition rate of 86.49% was achieved with Random Forest and is exceeded by ResNet-101 with 96.32 %. The reason for the enormous difference could be the outlier, which is known to be in the feature vector of this data set. DL networks analyze the available images so that no anomaly was present. However, even after removing the outlier, the deep neural networks outperform the results of the powerful Random Forest. It achieves an average recognition rate of 89.54 %. Since classification-relevant information is contained in the texture of the Metal Surface data set, neither color nor shape features make a contribution. It appears that the used CNNs can describe the texture better than the manually selected classic features. Classical classifiers, like SVM or Random Forest, working on basis of the extracted feature vectors and not on the original information content of the sample images like CNNs. For this reason, classical supervised machine learning algorithms only work as well as the quality of the represented information content in the feature vectors is. For these reasons in a parallel investigation a new well-adapted and more specific feature vector of Metal Surface data set was calculated by using specific Haralick texture features. Then an optimized SVM (SMO) (using Sequential Minimal Optimization for training of SVM) was used for classification. For illustration the standard feature vector of the Metal Surface data set consists of 123 features and the Haralick feature vector consists of 54 more specific texture features. Using the more specific Haralick feature vector the SVM (SMO) achieved a total recognition rate of 98.83 % compared to 85.03 % of LibSVM with using more unspecific features. It is noteworthy that the conventional classifiers show equally good results as the deep neural networks with the complex Autopetrography data set or with the well-adapted and more specific feature vector of Metal Surface data set. Although ResNet-101 achieves the highest recognition rate for Autopetrography (94.11 %), this is not significant in context to 93.05 % for LibSVM, especially if the standard

deviations are taken into account (0.95 % for ResNet-101 and 0.14 % for LibSVM). The reason for these results could be the careful selection of the well-adapted and more specific features by machine learning experts, whereby they are able to describe the classes as well as the CNNs.

5. CONCLUSION

Deep artificial neural networks have shown considerable development in recent years. They enable very good results in many areas, especially in speech recognition and machine vision. The Convolutional Neural Network, which is a representative of DL techniques, has the ability to process images directly and therefore does not require a classical feature vector. Thus, some steps of the image processing chain, like segmentation, feature calculation or feature selection are omitted because they are done automatically within the algorithm. Sometimes only a segmentation as step before classification is necessary, this depends on the grade of separation of objects in the image. If objects are well separated (only one object in a scene), than a segmentation is not necessary. Otherwise, a segmentation step have to be done before training and testing the CNNs. In the context of the present work, on the one hand not pre-trained networks and on the other hand pre-trained networks were used and compared with classical methods.

In order to deliver satisfactory results, deep neural networks require an enormous amount of data. That is not given with the used data sets, which is why no high recognition rates could be achieved with not pre-trained networks. The situation is different with the pre-trained networks. For some data sets insignificant improvements, for other major improvements were achieved by using DL. For Metal Surfaces with Haralick features and Autopetrography, the results of classical methods and deep neural networks are almost the same. This shows that a careful selection of features leads to a high recognition performance with the classical methods. However, making such a selection is not trivial. Furthermore, it should be noted that e.g. ResNet-101 requires significantly more time for training than Random Forest or SVM. For large data sets such as Autopetrography, this can take several days. On the other hand, the use of pre-trained deep neural networks is quite simple, since no segmentation or feature calculation is required. In addition, they do not have to be adapted to each new data set and still deliver very good results.

6. ACKNOWLEDGEMENTS

The current research projects, which form the basis of this paper, are OptoCheck (project no.: 2017 FE 9111), funded by the Free State of Thuringia and co-financed by the European Union under the European Regional Development Fund (EFRE), and Qualimess next generation (03IPT709X), funded by the federal ministry of education and research. We thank the funding authorities for the strong financial support of this work. The responsibility for the content of this paper lies with the authors. Many thanks go also to our project partners the society for production engineering and development (GFE - Gesellschaft für Fertigungstechnik und Entwicklung Schmalkalden e.V.) in Schmalkalden (Germany), the testing company for road construction and civil engineering mbH & Co. KG (Prüfunternehmen für Straßenbau und Tiefbau mbH & Co. KG) in Bernburg (Germany) and the Fraunhofer Institute for Applied Optics and Precision Engineering IOF in Jena, which provided pre-classified object samples or captured image data sets.

7. REFERENCE LINKING

- [1] Zhang, Q., Cao, R., Shi, F., Wu, Y. N. and Zhu, S.-C., "Interpreting CNN Knowledge Via An Explanatory Graph," AAAI 2018, <<https://arxiv.org/pdf/1708.01785.pdf>> (06 August 2019).
- [2] Lago, J., Ridder, F. de and Schutter, B. de, "Forecasting spot electricity prices. Deep learning approaches and empirical comparison of traditional algorithms," *Applied Energy* 221, 386–405 (2018).
- [3] Liu, P., Choo, Kim-Kwang Raymond, Wang, Lizhe and Huang, F., "SVM or deep learning? A comparative study on remote sensing image classification," *Soft Computing* 21(23), 7053–7065 (2017).
- [4] Sarikaya, R., Hinton, G. E. and Deoras, A., "Application of Deep Belief Networks for Natural Language Understanding," *IEEE/ACM Transactions on Audio, Speech and Language Processing* 22(4), 778–784 (2014).
- [5] Oliveira, T. P., Barbar, J. S. and Soares, A. S., "Computer network traffic prediction. A comparison between traditional and deep learning neural networks," *International Journal of Big Data Intelligence* 3(1), 28–37 (2016).
- [6] Gjoreski, H., Bizjak, J., Gjoreski, M. and Gams, M., "Comparing Deep and Classical Machine Learning Methods for Human Activity Recognition using Wrist Accelerometer", *Proceedings of the IJCAI 2016 Workshop on Deep Learning for Artificial Intelligence*, New York, USA (2016).

- [7] Fernández-Delgado, M., Cernadas, E. and Barro, Senén, Amorim, Dinani, "Do we Need Hundreds of Classifiers to Solve Real World Classification Problems?," *Journal of Machine Learning Research* 15, 3133–3181 (2014).
- [8] Cleve, J. and Lämmel, U., [Data Mining], De Gruyter Oldenbourg, München (2014).
- [9] Cortes, C. and Vapnik, V., "Support-vector networks," *Machine Learning* 20(3), 273–297 (1995).
- [10] Witten, I. H., Frank, E., Hall, M. A. and Pal Christopher J., [Data mining. Practical Machine Learning Tools and Techniques], Morgan Kaufmann, Cambridge, MA (2017).
- [11] Breiman, L., "Random Forests," *Machine Learning* 45(1), 5–32 (2001).
- [12] Anding, K., Garten, D. and Linß, E., "Application of intelligent image processing in the construction material industry," *ACTA IMEKO* 2(1), 61–73 (2013).
- [13] Kuritcyn, P., Anding, K., Linß, E. and Latyev, S., "Increasing the Safety in Recycling of Construction and Demolition Waste by Using Super-vised Machine Learning" *Journal of Physics: Conference Series* 588 (2015).
- [14] Han, J., Kamber, M. and Pei, J., [Data mining. Concepts and Techniques], Elsevier/Morgan Kaufmann, Amsterdam (2012).
- [15] Aggarwal, C. C., [Data Mining. The Textbook], Springer, Cham (2015).
- [16] Haar, L., "Research of different machine learning methods," Report (2019).
- [17] Sesselmann, M., Stricker, R. and Eisenbach, M., "Einsatz von Deep Learning zur automatischen Detektion und Klassifikation von Fahrbahnschäden aus mobilen LiDAR-Daten (Deep Learning for Automatic Detection and Classification of Road Damage from Mobile LiDAR Data)," *AGIT - Journal für Angewandte Geoinformatik* vol. 5-2019, 100–114 (2019).
- [18] Maladkar, K., "6 Types of Artificial Neural Networks Currently Being Used in Machine Learning", <<https://www.analyticsindiamag.com/6-types-of-artificial-neural-networks-currently-being-used-in-todays-technology/>> (24 April 2019).
- [19] LeCun, Y., Bengio, Y. and HINTON, G., "Deep Learning," *nature* 521(7553), 436 (2015).
- [20] Le, Q. V., "A Tutorial on Deep Learning Part 2. Autoencoders, Convolutional Neural Networks and Recurrent Neural Networks," Report, Mountain View, CA (2015).
- [21] Marcus, G., "Deep Learning. A Critical Appraisal," Report, (2018).
- [22] MVTec Software GmbH, "HALCON/HDevelop. Operatorreferenz (de)," Manual, (2017).
- [23] Bengio, Y., [Practical Recommendations for Gradient-Based Training of Deep Architectures], *Neural networks: Tricks of the trade*, Springer, Berlin, 437–478 (2012).
- [24] Eisenbach, M., Seichter, D. and Gross, H.-M., "Are Color Features Important for Person Detection? - Insights into Features Learned by Deep Convolutional Neural Networks", *Proc. Workshop Farbbildverarbeitung (FWS)*, Ilmenau, 169–182 (2016).
- [25] Montavon, G., Samek, W. and Müller, K.-R., "Methods for Interpreting and Understanding Deep Neural Networks," *Digital Signal Processing* vol. 73, 1–15 (2018).
- [26] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V. and Rabinovich, A., "Going Deeper with Convolutions," *Proceedings of the IEEE conference on computer vision and pattern recognition*, IEEE, 1–9 (2015).
- [27] Moosavi-Dezfooli, S.-M., Fawzi, A. and Frossard, P., "DeepFool. A Simple and Accurate Method to Fool Deep Neural Networks," *Proceedings of the IEEE conference on computer vision and pattern recognition*, IEEE, 2574–2582 (2016).
- [28] Nguyen, A., Yosinski, J. and Clune, J., "Deep Neural Networks are Easily Fooled. High Confidence Predictions for Unrecognizable Images," *Proceedings of the IEEE conference on computer vision and pattern recognition*, 427–436.
- [29] Papernot Nicolas, McDaniel, P. and Jha, S., "The Limitations of Deep Learning in Adversarial Settings," *IEEE European Symposium on Security and Privacy (EuroS&P)*, IEEE, 372–387 (2016).
- [30] Anding, K., [Automatisierte Qualitätssicherung von Getreide mit überwachten Lernverfahren in der Bildverarbeitung], Dissertation, Ilmenau (2010).
- [31] Linß, G., Anding, K., Garten, D., Göpfert, A., Rückwardt, M. and Reetz, E., "Automatic petrographic inspection by using image processing and machine learning," *XX IMEKO World Congress, Metrology for Green Growth*, Busan, Korea (2012).

- [32] Garten, D., Anding, K. and Lerm, S., "Image Based Recognition - Recent Challenges and Solutions Illustrated on Applications," *Machine Learning and Applications: An International Journal (MLAIJ)* ISSN: 2394-0840, Vol. 3, No. 3 (2016).
- [33] Garten, D., Anding, K., Polte, G. and Trambitckii, K., "Automatic design of classification systems for visual quality control of metallic surface," 14th IMEKO TC10 Conference on Technical Diagnostics New Perspectives in Measurements, Tools and Techniques for system's reliability, maintainability and safety Milan, Italy (2016).
- [34] Haar, L., Anding, K., Schröder, S., Hauptvogel, M. and Notni, G., "Image Processing of Light-Scattering Images for the Qualitative Surface Analysis," 14th IMEKO TC10 Conference on Technical Diagnostics New Perspectives in Measurements, Tools and Techniques for system's reliability, maintainability and safety Milan, Italy, 36–41 (2016).
- [35] Forschergruppe Prozessbegleitende Qualitätssicherung, "Abschlussbericht im Thüringer Zentrum für Maschinenbau, Thüringer Aufbaubank-gefördert" (2016).
- [36] Walz, J., [Vergleich klassischer und innovativer Klassifikationsverfahren unter Einbezug der Ausreißerdetektion], Masterthesis, Ilmenau (2018).
- [37] LeCun, Y., Bottou, L., Bengio, Y. and Haffner, P., "Gradient-based learning applied to document recognition," *Proceedings of the IEEE* 86(11), 2278–2324 (1998).
- [38] Anding, K., Kuritcyn, P. and Garten, D., "Using artificial intelligence strategies for process-related automated inspection in the production environment," *Journal of Physics: Conference Series*. Vol. 772. No. 1. (2016).
- [39] MathWorks, I., "Pretrained Convolutional Neural Networks,"
<<https://de.mathworks.com/help/deeplearning/ug/pretrained-convolutional-neural-networks.html>> (4 March 2019).
- [40] Krizhevsky, A., Sutskever, I. and Hinton, G. E., "ImageNet Classification with Deep Convolutional Neural Networks," *Advances in neural information processing systems*, 1097–1105 (2012).
- [41] MathWorks, I., "alexnet. Pretrained AlexNet convolutional neural network,"
<<https://de.mathworks.com/help/deeplearning/ref/alexnet.html>> (5 March 2019).
- [42] He, K., Zhang, X., Ren, S. and Sun, J., "Deep Residual Learning for Image Recognition," *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770–778 (2016).
- [43] MathWorks, I., "resnet101. Pretrained ResNet-101 convolutional neural network,"
<<https://de.mathworks.com/help/deeplearning/ref/resnet101.html>> (5 March 2019).