Transfer learning in recognition of drill wear using convolutional neural network

Jaroslaw Kurek, Grzegorz Wieczorek, Bartosz Swiderski Michal Kruk, Albina Jegorowa

University of Life Sciences, Warsaw, Poland

Stanislaw Osowski

Faculty of Electrical Engineering, Warsaw University of Technology and Electronic Faculty, Military University of Technology, Warsaw, Poland

***Abstract*—The paper presents an application of transfer learning using convolutional neural network (CNN) in recognition of the drill state on the basis of hole images drilled in the laminated chipboard. Three classes are recognized: red, yellow and green, which correspond with 3 stages of drill state. Red class indicates the drill, which is worn out and should be replaced immediately in drilling process. Yellow class corresponds to the state in which warning should be sent to the operator to check manually state of the drill. The last class corresponds to the green state indicating good condition of drill, enabling further use in production. The important advantage of transfer learning approach is possibility of training classification model using only small portion of data. This is in contrast to the classical deep learning methods of convolutional neural networks, which require very large data base to achieve acceptable accuracy of class recognition. The results of numerical experiments in drill state recognition have confirmed suitability of this novel method to accurate class recognition at small population of available learning data.**

Keywords—deep learning; convolutional neural networks; tool condition monitoring

I. INTRODUCTION

The sharpness of the of drill is an important aspect of production quality in furniture manufacturing industries. Not enough sharp drill can have bad impact on produced furniture elements and lead to budget losses of company. It is important to catch the production point when the operator should replace drill by the new one. The simple manual observation of drill is not effective and the research to automatize this stage by the computerized system is carried out.

Three classes of drill states are usually considered by specialists: red, yellow and green in an automatic assessment to the drill. Red class means that drill should be replaced immediately and not used in drilling process (to many damaged furniture elements). Yellow state corresponds to the drill of suspected state. In this case the assessment of drill state should be taken up by an expert. The manual retrospection by an operator will answer the question whether drill should be replaced or was only falsely messaged by the automatic system. The last, green state indicates good condition of drill, leaving production process onward without any changes.

Many dedicated sensors are often used in automatization of this process, called tool condition monitoring (TCM). Such automatic system of drill assessment applies usually signals measured by many devices, on the basis of which the diagnostic features are defined. The most commonly used signals are these related to feed force, cutting torque, noise, vibration and acoustic emission [4]. However, to register them we have to install quite expensive arrangement of different sensors. Moreover, it needs a lot of preprocessing stages, including selection of proper dedicated sensors, registering appropriate signals, generation and selection of the best diagnostic features and finally building the classification model applying the diagnostic features as the input attributes. All this means, that TCM in classical approach is quite complex and expensive. The papers [1,2,3] have taken into accounts many features generated on the basis of these registered signals. However, the accuracy of 3 class recognition are below the level of 90%.

To reduce the complexity and cost of drill monitoring we suggest to rely our assessment on the images of the drilled holes. In such case only camera is needed to take a picture of holes after drilling process. On the basis of the hole image the drill condition will be assessed. Recently, the most efficient approach in image processing is deep learning using the convolutional neural network (CNN) [5,8]. Thanks to its application there is no need to elaborate the specialized diagnostic features. They are generated automatically in an unsupervised way by using many hidden layers.

The first attempt to apply deep learning to this problem has been presented in [5], where CNN and deep learning were engaged in recognition of 2 classes of drill (sharp enough and worn out). On the basis of 900 images (300 images representing the first class and 600 images of the second class) the system was able to achieve 66.6% accuracy. After increasing this number (through rotation, scaling and adding some noise) to 11700 and then to 33300 images the recognition accuracy has been increased to 89% and 95.5%, respectively. However, the process of learning the whole system was very long and lasted over 20 hours.

This paper will propose different approach to building the classification system, which is significantly less resistant to the number of samples used in learning. It applies so called transfer learning [10,11,13].

The classification process will consider 3 classes of drill states, called red, yellow and green, similarly to traffic lights in production. Red class indicates the drill, which is worn out and should be replaced immediately in drilling process. Yellow class warns operator to check the state of drill in manual way. The green class corresponds to the drill state indicating good condition of drill, enabling its further use in production. The aim of the work is to show, that application of transfer learning at application of even small number of originally measured samples allows getting good results of image recognition. The numerical experiments will be performed on the data set significantly smaller than presented in paper [5]. It contains only 242 samples of images, representing these three classes.

II. DATABASE

The samples of data in the form of drilled hole images have been collected in cooperation with Faculty of Wood Technology at Warsaw University of Life Sciences using standard Buselatto JET 100 CNC vertical machining centre.

The drilling processes have been performed on standard laminated chipboard (Kronopol U 511 SM) of the dimension 150x35x18mm, used typically in furniture industry. The acquisition of images has been done by means of regular 12 mm FABA drill equipped with a tungsten carbide tips.

The database consist of three subsets of samples. The first subset represents green class, where drill is sharp enough to continue drilling process. An example of such drill is presented in Fig. 1a. 102 samples of the hole images made by this drill have been acquired. After making these 102 drilling processes the drill has been manually blunted by the experts using microscope. The typical microscopic view of this drill is presented in Fig. 1b. The drill prepared in this way was used in the next 60 drilling processes, producing images representing yellow class. The last subset of data, called red, has been generated after blunting the drill to the state leading to the unacceptable drilled holes. Totally the database consists of only 242 samples (hole images) with the following distribution: green class - 102 samples/images, yellow class – 60 samples/images and red class – 80 samples/images. Figure 1 depicts examples of the typical hole images representing these three classes of data.

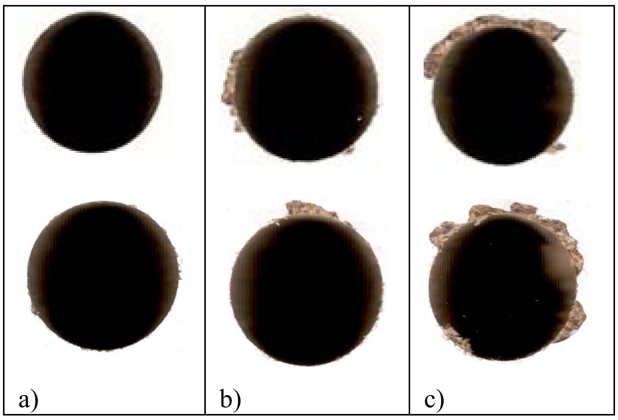


Fig. 1. Examples of hole images representing three classes of data: a) green b) yellow, c) red.

III. DEEP LEARNING APPROACH USING CNN

Deep learning, called also deep structured learning or deep machine learning, is currently very popular classification approach, especially efficient for the images [6, 7, 8, 9]. In such case the convolutional neural network (CNN) is typically used. The hidden layers of nonlinear processing perform the function of extracting the diagnostic features, starting from small elements in the images, like blobs, points, edges, corners, etc. Each successive layer uses the output signals from the previous layer as its excitation. In this way the higher level features are derived from lower level features and create the hierarchical representation of data. Thanks to this the user does not need to elaborate the specific diagnostic features for the particular problem. Deep learning finds the features themselves in the training process without user manual intervention. Moreover, the basic features created sequentially from layer to layer are universal and serve reasonable well for different types of images.

Unlike the regular methods where pixel by pixel is considered in the process of feature generation, deep learning algorithm analyses blocks of pixels, filtering the subsequent regions of different, however, fixed size, for example 5x5 pixels in small images up to 15x15 pixels in large images. The linear filtration is followed by the nonlinear transformation of the filter output signals, usually rectified linear unit (ReLU), and finally by pooling operation, which reduces the size of the actually processed blocks.

The most important difference between the classical machine learning and deep learning approaches in image recognition is the way of extracting diagnostic features. In classical machine learning it is a separate stage of processing, usually involving many different stages. By contrast, in deep learning approach the feature generation is done automatically in hidden layers as an embedded internal process of learning the whole multilayer structure. The extracted features may be used as the input signals to the softmax classifier representing the integral part of CNN or serve as the input attributes to the external classifier, for example support vector machine.

The CNN network used in this paper applies 5 convolution layers responsible for generation of diagnostic features. Each convolution layer performs local filtering, followed by the ReLU nonlinear transformation and normalization, as well as pooling operation performed on the resulting sub-images. The filtration is done by the neuron of the specified weights connected locally with the pixels forming the small mask moving vertically and horizontally along the whole image with some step (stride). The convolution layer is composed of many such neurons, specializing in extracting different types of details, like points, edges, crossings, etc.

The rectified linear unit provides the output signal to be nonnegative, since $\mathrm{ReLU}(x)=x$ for $x>0$ and 0 for $x\leq 0$. Each image produced by the particular neuron in the convolutional layer is subject to pooling. This operation is aimed on replacing the input image by its reduced size statistical representation. The most popular functions include Max pooling or average pooling.

The important operation in CNN is zero padding. The chosen number of zeros is added to the border of image to prevent quick shrinking of the successive images in the following layers. Zero padding the input allows controlling the kernel width and the size of the output image.

Last few layers in CNN are called fully connected, since their neurons are fully connected with the neurons of the next layer, like in classical neural networks. The output layer contains as many neurons as is the number of recognized classes (in our case 3).

IV. PRETRAINED CNN ALEXNET IN RECOGNITION OF HOLE IMAGES

The data base used in our experiments contains only 242 images of the drilled holes. It is significantly too small base to start training CNN from scratch. In such case the only possibility to obtain good results of classification is to use the special model pretrained on very large set of absolutely different images. This approach is known as transfer learning [11,12].We will use here the AlexNet model created by Krizhevsky, Sutskever, and Hinton [10,11,12] and implemented in Matlab [13]. This model was pretrained using more than one million images representing 1000 different classes [13,14]. The AlexNet structure used in our task consists of 9 layers composed of 25 sub-layers.

Softmax classifier.

The softmax function takes the vector **u** and calculates the *i*th component of output vector representing *M* classes in the form

$$\mathrm{softmax}(\mathbf{u})_i = \frac{\exp(u_i)}{\sum_{j=1}^{M} \exp(u_j)} \qquad (1)$$

This function is a normalized exponentially and takes the values from 0 to 1. They are treated as probability of particular class. The highest value of the component indicates the recognized class. In calculation of softmax function the cross entropy has been used. In our application for three classes *M*=3.

To apply the pretrained AlexNet all images should be transformed to the same size required by this network 227x227x3. In adaptation of the model to our purposes we have applied the first six trained layers without any change. It means that mechanism of creating the diagnostic features has been directly taken from different classes of images used in pretraining stage of AlexNet. The last three, fully connected layers have been adapted by us in the learning process to obtain the best results of our image recognition.

V. SUPPORT VECTOR MACHINE (SVM)

As the alternative of last of layer in deep learning structure Support Vector Machine (SVM) has been applied in form of the Gaussian kernel [13,17]. SVM is a simple circuit structure of one hidden kernel layer and one output unit performing the weighted summation followed by sign function (positive summed signal means class 1 and negative– class 2). The hyperparameters (the regularization constant C and Gaussian kernel width) have been adjusted by repeating the learning experiments for the set of their predefined values and choosing the best one for the validation of data set. The learning process of SVM network is relatively easy and effective since the whole learning task is simplified to the solution of the quadratic optimization problem with linear constraints. In our experiments we have used the modified Platt algorithm, implementing the modified sequential optimization [17].

VI. RESULTS OF NUMERICAL EXPERIMENTS

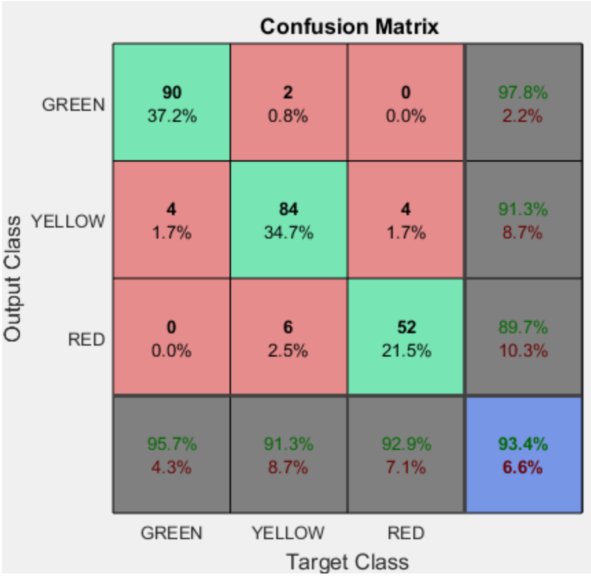
The data set used in numerical experiments contained only 242 images representing three classes. This set was split randomly into 10 subsets: 9 subsets were used in training process and the remaining one in testing mode. The experiments have been repeated 10 times in 10-fold cross validation mode, each time exchanging the testing part of data. The statistical results in the form of the mean value of class recognition accuracy obtained in 10 repetitions of classification experiments are presented in Table I. The results are related to the standard CNN, learned from scratch (the row "Standard CNN") and application of the pretrained CNN corresponding to AlexNet with softmax as a classifier (the row "Pretrained CNN").

TABLE I. RESULTS OF APPLYING DEEP LEARNING ALGORITHMS IN DRILL CONDITION CLASSIFICATION TO 3 CLASSES (GREEN, YELLOW, RED) ON THE BASIS 242 IMAGES OF DRILLED HOLES.

Deep learning algorithm	**Accuracy[%]**
Standard CNN	35%
Pretrained CNN	85%
Pretrained CNN+SVM (C=1000, gamma=0.01)	93.4%

The advantage of using pretrained CNN is evident. However, the final accuracy of class recognition is still not satisfactory from practice point of view. These results have been compared to much more complex and one of the best classical approach - Support Vector Machine (SVM) [4,14,15],. Therefore, in the third numerical experiments we have changed the final classifier to the Support Vector Machine (SVM). We have applied hyperparameters such as C=100, =10.

TABLE II. CONFUSION MATRIX FOR THE TESTING DATA IN 10-FOLD CROSS VALIDATION APPROACH USING SVM AS A FINAL CLASSIFIER



The input attributes to SVM were formed by the features extracted from 8th layer. It means 500 input signals generated by CNN AlexNet. The SVM applied the Gaussian kernel of gamma=0.1 and regularization constant C=1000. The results of this solution are presented in the last row of Table I. It is evident, that application of SVM as a classifier has led to the best accuracy (93.4%).

Table II depicts the obtained confusion matrix for the testing data in Matlab format [13], related to the application of SVM as the final classifier. The sensitivity of recognition of the green class was 95.7%. Slightly worse were the sensitivities of discovering the yellow and red class: 91.3% and 92.9%, respectively. The green and red class have been never misclassified. The only problems have appeared in recognition of the yellow class, which was sometimes misinterpreted as the green or red class. These results have proved the applicability of this simplified drill diagnostic system in practical application. By increasing the population of learning data (especially the yellow class) we will be able to reduce the misclassification rate to the value close to zero.

VII. Conclusion

The paper has presented the application of the pretrained AlexNet CNN network to the recognition of the state of the drill on the basis of the set of images of the drilled holes. The main advantage of such approach is that limited number of learning samples might be used to train the CNN network to the required task.

The presented results of numerical experiments confirmed good performance of the pretrained network in recognition of three sharpness states of the drill. The results have been compared to these obtained by traditionally learned CNN, in hich the population of learning data has been increased using such operations as rotation and scaling of the original images.

The artificial images of different types used in large amount in training process of AlexNet (over 1 million different images in recognition of 1000 different classes) have allowed obtaining good set of diagnostic features for general use in recognition of arbitrary images. The pretrained network needed only little interference in the last three full layers of CNN. This process was relatively quick and resulted in a good generalization property of the class recognition system.

References

[1] K. Jemielniak, T. Urbański, J. Kossakowska and S. Bombiński, “Tool condition monitoring based on numerous signal features”, *Int. J. Adv. Manuf. Technol.*, vol. 59, pp. 73–81, 2012.

[2] S. S. Panda, A. K.Singh, D. Chakraborty and S. K. Pal S.K., “Drill wear monitoring using back propagation neural network”, *Journal of Materials Processing Technology*, vol. 172, pp. 283–290, 2006.

[3] R. J. Kuo, “Multi-sensor integration for on-line tool wear estimation through artificial neural networks and fuzzy neural network”, *Engineering Applications of Artificial Intelligence*, vol. 13, pp. 249-261, 2000.

[4] J. Kurek, M. Kruk, S. Osowski, P. Hoser, G. Wieczorek, A. Jegorowa, J. Górski, J. Wilkowski, K. Śmietańska and J. Kossakowska, “Developing automatic recognition system of drill wear in standard laminated chipboard drilling process”, *Bulletin of the Polish Academy of Sciences. Technical Sciences,* vol.64, pp. 633—640, 2016.

[5] J. Kurek, B. Swiderski, A. Jegorowa; M Kruk and S. Osowski, “Deep learning in assessment of drill condition on the basis of images of drilled holes”, *Proc. SPIE 10225*, Eighth International Conference on Graphic and Image Processing (ICGIP 2016), 102251V (February 8, 2017); doi:10.1117/12.2266254.

[6] L. Deng, and D. Yu, "Deep Learning: Methods and Applications", *Foundations and Trends in Signal Processing*, vol. 7, pp. 3–4, 2014.

[7] Y. Bengio, "Learning Deep Architectures for AI", *Foundations and Trends in Machine Learning*, vol. 2, No 1, pp. 1–127. 2009.

[8] I. Goodfellow, Y. Bengio and A. Courville, *Deep learning*, MIT Press, 2016.

[9] J. Schmidhuber, "Deep Learning in Neural Networks: An Overview", *Neural Networks*, vol. 61, pp. 85–117, 2015.

[10] A. Krizhevsky, I. Sutskever and G. Hinton, “Image net classification with deep convolutional neural networks”, Advances in Neural Information Processing Systems, vol. 25, pp. 1-9, 2012.

[11] O. Russakovsky, J. Deng, H. Su, et al. "ImageNet Large Scale Visual Recognition Challenge." *International Journal of Computer Vision* (IJCV). Vol 115, Issue 3, 2015, pp. 211–252.

[12] *BVLC AlexNet Model.* https://github.com/BVLC/caffe/tree/master/models/bvlc_alexnet

[13] *Matlab 2017a, user manual*, The MathWorks, Inc. Natick, MA, USA, 2017.

[14] ImageNet. http://www.image-net.org

[15] B. Scholkopf and A. Smola, “Learning with Kernels”, MIT Press,cambridge, (2002).

[16] M. Kruk, B. Świderski, S. Osowski, J. Kurek, M. Słowińska and I. Walecka, “Melanoma recognition using extended set of descriptors and classifiers”, Eurasip Journal on Image and Video Processing 43, 1-10, (2015).

[17] Vapnik, Statistical Learning Theory, Wiley, New York, 1998.