Journal of Machine Engineering, 2023, Vol. 23, No. 2, 100–109
ISSN 1895-7595 (Print) ISSN 2391-8071 (Online)

Received: *10 January 2023* / Accepted *16 April 2023* / Published online: *19 April 2023*

deep learning, assembly, begin-end point

Izabela KUTSCHENREITER-PRASZKIEWICZ[1*]

DEVELOPMENT OF A NEURAL NETWORK STRUCTURE FOR IDENTIFYING BEGIN-END POINTS IN THE ASSEMBLY PROCESS

The paper presents an approach to video-based assembly analysis using machine learning. A neural network is one of the machine learning methods that is widely studied in many engineering fields. The purpose of this paper is to develop a deep neural network structure for identifying begin-end points for a selected component assembly process. A neural network structure that effectively identifies begin-end points is proposed and an example from industry is presented. The proposed approach can prove useful in the assembly process analysis.

1. INTRODUCTION

Deep learning (DNN) is one of the fields of artificial intelligence and artificial neural networks ANN, which is a widely used method in many engineering fields. Authors discuss different approaches to intelligent data analysis in the manufacturing process. Putnik et al. proposed an intelligent machine architecture with multiple learning meta-levels [1], and noticed that artificial intelligence is useful as a method for manufacturing control and supporting the decision making process based on simulation [2]. Uhlmann et al. [3] proposed integration of ANN models in the process of accuracy improvement of industrial robots. Encoder-decoder network for assembly feature processing was presented by Qibing et al. [4], who adopted the historical assembly experience to guide the design of human-robot action sequence and planning industrial robot paths towards new tasks. Wang et al. [5] proposed deep learning as a method for supporting visual observation of human workers' movements. Ceglarek et al. [6] propose a novel Object Shape Error Response (OSER) approach to estimate the dimensional and geometric variation of the assembled products and [7] use deep learning for 3D object shape error modelling and estimate dimensional and geometric quality defects in multi-station assembly systems. Image based robot manipulator control system was developed by Copot et al. [8], and particularly visual control robot manipulator using image moments. Chaumette [9] presents the experimental results that show that a correct behaviour

[1] Faculty of Mechanical Engineering and Computer Science, University of Bielsko-Biala, Bielsko-Biala Poland
[*] E-mail: ipraszkiewicz@ath.bielsko.pl
https://doi.org/10.36897/jme/163318

of the system is obtained if it considers either a simple symmetrical object or a planar object with complex and unknown shape. According to literature analysis, DNN bring good results in the area of image recognition.

Among the various methods used to classify images, a neural network is commonly used. A structure of neural network which effectively recognized begin-end points was proposed and an example from industry was presented. The proposed approach can be useful for automatic analysis of the assembly process.

The proposed application of ANN in assembly process analysis is presented in Fig. 1. Control and prediction of the assembly process can be supported by ANN. The aim of the paper is to develop an ANN structure which can effectively recognize begin-end points in the assembly process.

According to literature analysis, the demonstration data used in the training process is one of the key factors affecting the effectiveness of deep learning applications [10].

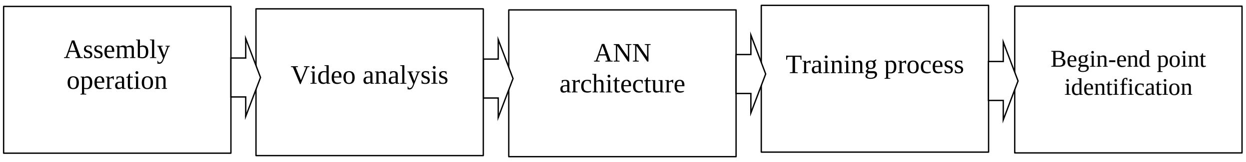

Fig. 1. ANN application in assembly process analysis

The begin-end point identification is necessary for setting the time standard. The period between the begin-end points is the basis for the cycle time calculation of the assembly process.

The problem addressed in the paper is the development of a DNN structure for the visual identification of begin-end points in the assembly process in a manual work system in which assembly tasks are repeated with a high degree of similarity.

The current approach related to assembly analysis in manual work system is focused on video analysis of the process without any AI support.

The proposed approach uses DNN for identification of begin-end points in manual assembly. The novelty of the proposed approach is the automation of pointing out the begin-end points in the video recording of a manual assembly process. There is a gap in the literature – no application of AI was found for identification of begin-end points in the assembly process. The presented approach resulted in the creation of an IT tool for controlling the assembly process.

2. ASSEMBLY ANALYSIS

Among various types of work systems such as [11]: manual work system, worker-machine system, and automated system, manual work system can cause failures in the production process, so design and control in this area are very important. The presented approach is focused on the manual work system associated with the analysis of assembly tasks, where manual tasks consist of a work cycle that is repeated with high degree of similarity, and each cycle corresponds to processing one work unit.

It is important to design the work cycle so as to minimize the time required to perform it [11]. Work measurement techniques are useful in comparing different possible methods of work.

Work measurement includes the time study, which is the most common method of deter--mining standard assembly time. Performing a time study involves analysing and measuring the time of the assembly process. To analyse the assembly process (Fig. 2), it is necessary to observe the process in order to break it down into tasks (elements) using begin-end points. Begin-end points are easily discernible moments at which individual elements of the work process begin or end. As a rule, the same begin-end point is simultaneously the moment when the previous element is completed and the moment when the next element starts [12].

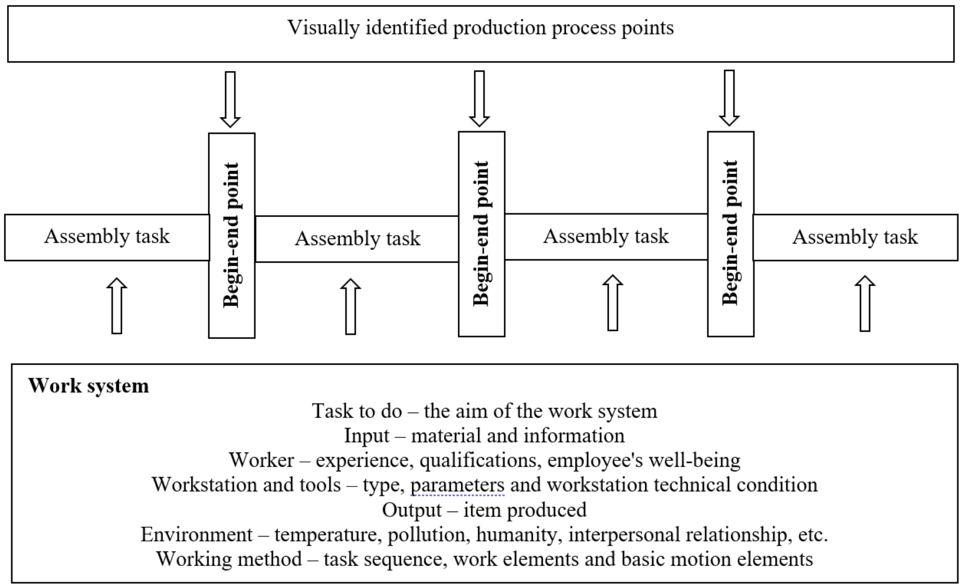

Fig. 2. Begin-end points in the assembly process

Conventional work measurement methods like time study require human involvement in analysing the production process. Such an approach is time-consuming and errors may occur in the measurement process. It is therefore necessary to analyse the production process using artificial intelligence methods, such as neural networks, especially deep learning, which is one of the most successful methods using machine learning.

The proposed approach applies a deep learning neural network to identify the begin-end points in the assembly process.

3. IMAGE CLASSIFICATION

One of the main problems in computer vision is image classification (Fig. 3), which analyses the numerical properties of image features and organizes them into categories [13].

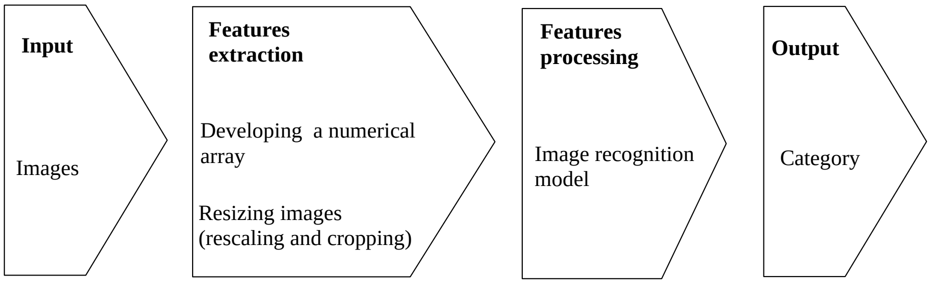

Fig. 3. Image classification process

In assembly process analysis, human detection and object detection methods can be useful. Image analysis of components in the production process can be performed using different types of visual features, which are often divided into: point features (centroids or corners), line or ellipse features, and generic descriptors (image moments) [14, 15]. Image moments represent generic visual features that can be computed easily from a binary or a segmented image, or from a set of point features [14] [16].

Human detection techniques based on feature extraction and classification through a machine-learning algorithm [17]. Feature extraction based on human detection methods are categorized into two significant approaches, namely: holistic human detection and part-based human detection [17]. Numerous holistic human detection methods were proposed based on Haar wavelet features, Haar features with motion information, edge templates, shape contexts, implicit shape models, adaptive contour features, scale-invariant feature transform (SIFT) descriptors, Gabor filters, covariance descriptor, local binary pattern (LBP) and histograms of oriented gradients (HOG) [17]. To deal with the problem of high intraclass variability in humans, classification methods such as cluster boosted tree, intersection kernel for support vector machine (SVM), multiple instances learning, "seed-and-grow" scheme and convolutional neural network (CNN) can be used [17].

In recent years, many advanced classification approaches have been widely applied for image classification, such as Support Vector Machines SVMs, which work by making histograms of images, Bag of Features Models like Scale Invariant Feature Transformation (SIFT) and Maximally stable extremal regions (MSER) [19]. Deep learning is part of a broader family of machine learning methods based on artificial neural networks (Fig. 4) [18, 19], which includes a wide and diverse number of architectures and algorithms and is understood as a neural network with three or more layers [20]. The structure of deep learning networks involves multiple hidden layers to enable the extraction of features that are deeply embedded in the data, forming abstract concepts in a hierarchical manner [5, 21]. Deep learning architectures have been applied to fields including computer vision, speech recognition, natural language processing, machine translation, bioinformatics, medical image analysis, climate science, etc. [22, 23]. Deep architectures include many variations of several basic approaches. Each architecture has been successful in specific application areas [23].

CNN is made up of several layers that implement feature extraction and then classification (Fig. 5). The input image is divided into receptive fields that feed into a convolutional layer, which then extracts features from the input image. The next step is pooling, which reduces the dimensionality of the extracted features (through down-sampling) while retaining the most important information (typically, through max pooling, which is a pooling operation that selects the maximum element from the region of the feature map covered by the filter). Another convolution and pooling step is then performed that feeds into a fully connected multilayer perceptron. The final output layer of this network is a set of nodes that identify features of the image. CNN can be trained by using back-propagation [23].

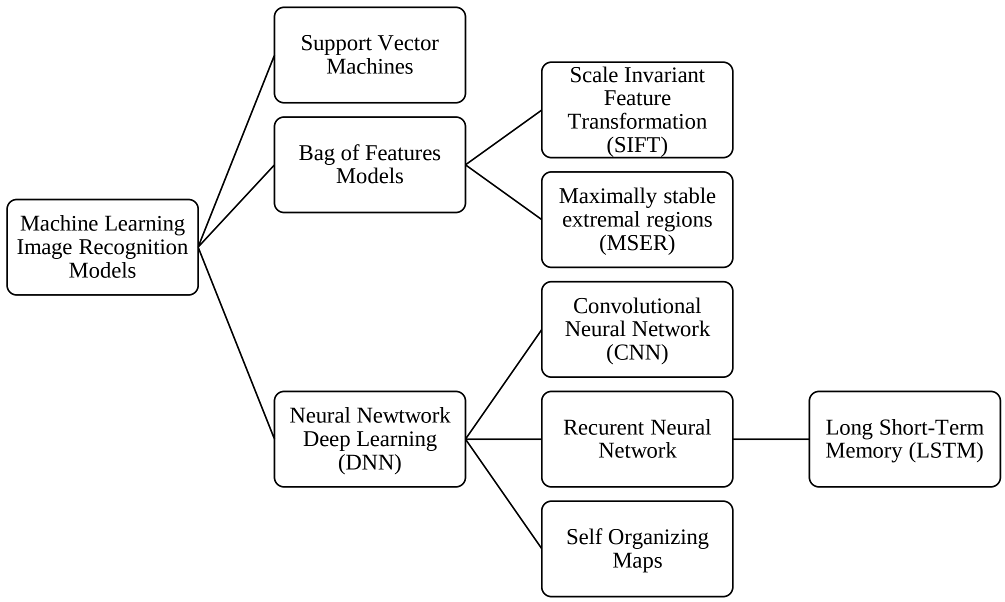

Fig. 4. Chosen Machine Learning Image Recognition Models (based on [19, 23])

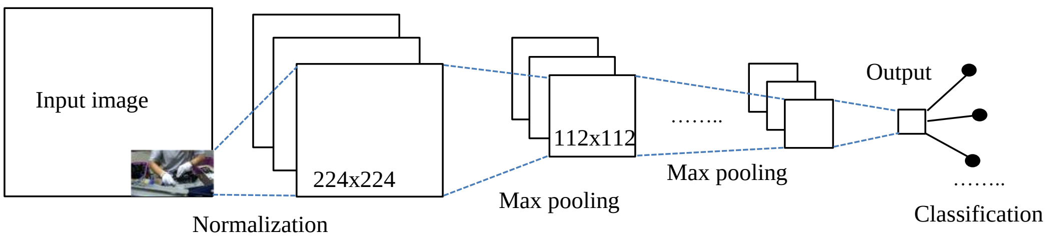

Fig. 5. Convolutional neural network

The Long Short-Term Memory (LSTM) departed from typical neuron-based neural network architectures and instead introduced the concept of a memory cell. The memory cell can retain its value for a short or long time as a function of its inputs, which allows the cell to remember what's important and not just its last computed value. The LSTM memory cell contains three gates that control how information flows into or out of the cell. The input gate controls when new information can flow into the memory. The forget gate controls when an existing piece of information is forgotten, allowing the cell to remember new data. Finally, the output gate controls when the information that is contained in the cell is used in the output from the cell (Fig. 6). The cell also contains weights, which control each gate. The training algorithm, commonly backpropagation through time (BPTT), optimizes these weights based on the resulting network output error [18].

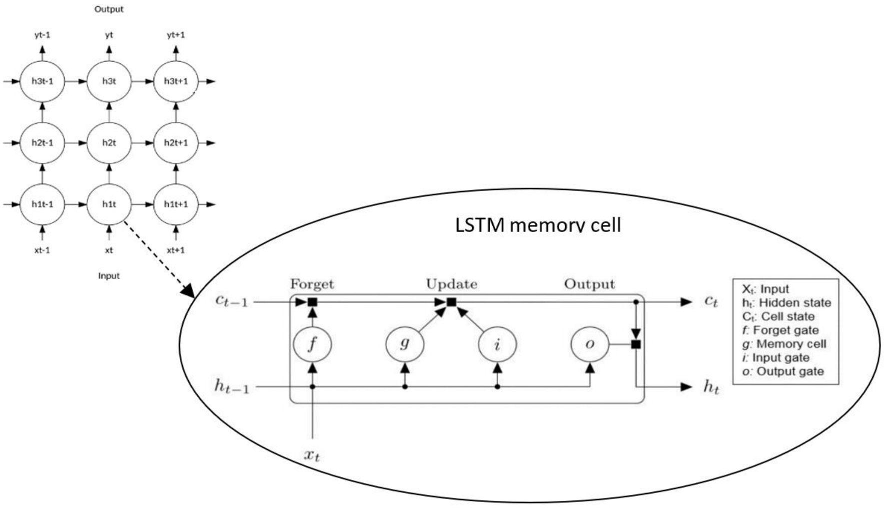

Fig. 6. The architecture for an LSTM [18, 24]

Self-organized map (SOM) was popularly known as the Kohonen map. SOM is an unsupervised neural network that creates clusters of the input data set by reducing the dimensionality of the input [18].

The points in input space have a correspondent points in output space. In the Kohonen Networks, a kind of SOM, there is a single layer with two dimensions and the input points are fully connected with neurons on this layer [25].

SOMs vary from traditional artificial neural networks in quite a few ways. The first significant variation is that weights serve as characteristics of the node. After the inputs are normalized, a random input is first chosen. Random weights close to zero are initialized to each feature of the input record. These weights now represent the input node. Several combinations of these random weights represent variations of the input node. The Euclidean distance between each of these output nodes with the input node is calculated. The node with the least distance is declared as the most accurate representation of the input and is marked as the best matching unit (BMU). With these BMUs as central points, other units are similarly calculated and assigned to the cluster in a way so as to maximise the distance from one cluster to another. The radius of points around BMU weights is updated based on proximity by minimising the distance. Next, in a SOM, no activation function is applied, and because there are no target labels to compare against, there is no concept of calculating error and back propagation [18].

Effective feature selection at successive layers that distinguish between different image categories is critical to achieve accurate image recognition and classification [5].

4. DEEP NEURAL NETWORK IN ASSEMBLY ANALYSIS – THE PROPOSED APPROACH

Identifying the begin-end points is a key issue in measuring and designing work. Methods such as time study use begin-end points to accurately analyse work tasks. Work task improvement influences, among others, the flow of material, production costs, efficiency, and delivery time.

The purpose of this paper is to develop a method that helps automate the analysis of the assembly process and is easily applicable. The proposed approach consists of the following steps:

– Assembly process analysis, identification of begin-end points,
– Training / testing set development,
– ANN structure development, training and validation.

4.1. ASSEMBLY PROCESS ANALYSIS

The research on the process of identifying begin-end points was carried out using the example of assembling a subassembly consisting of four components using a hand tool. The components were taken from containers located at the workstation. One begin-end point was identified in the single-cycle assembly process. The components were deposited in a container located at the workstation. The begin-end point was indicated as the release of the assembled component. A video and series of images on the production process were prepared. The assembly process was analysed for more than a dozen work cycles.

4.2. TRAINING / TESTING SET DEVELOPMENT

Using the video from the working area of the operator, a set of images can be created. The corresponding class is assigned in the following stages: image processing, accumulation and classification [26].

In the example presented here (Fig. 7), the training sets contained 2452 images, of which 2377 images formed the first class associated with the assembly task, and 75 images were used to create the second class associated with component release.

The validation sets contained 1051 images, wherein 1019 images formed class one, and 32 images formed class two.

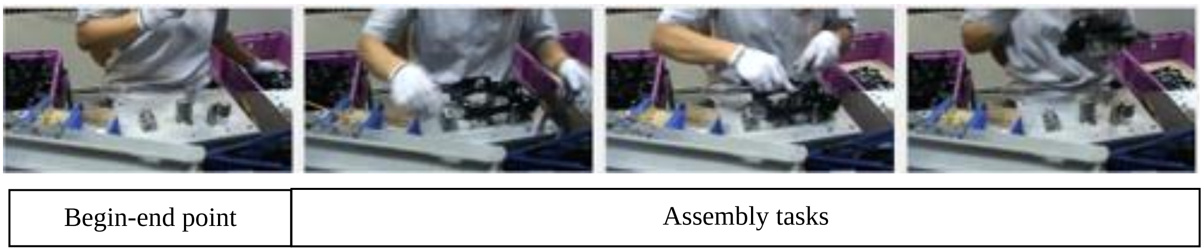

Fig. 7. Training set examples

4.3. DNN STRUCTURE DEVELOPMENT, TRAINING AND VALIDATION

For Deep Learning of 2D image recognition, different network architectures have been developed, such as AlexNet, GoogLeNet, ResNet or Inception [27].

Deep neural networks DNN contain multiple non-linear hidden layers and this makes them very expressive models that can learn very complicated relationships between their inputs and outputs [28]. One of successful techniques used in DNN that prevent overfitting is dropout, which temporarily removes units in ANN along with all its incoming and outgoing connections. The choice of which units to drop is random. In the simplest case, each unit is retained with probability p, independent of other units, where p can be chosen using a validation set or can simply be set to 0.5, which appears to be close to optimal for a wide range of networks and tasks. For input units, however, the optimal stopping probability is usually closer to 1 than 0.5 [28]. Another useful DNN technique is the SoftMax function, also known as normalized exponential function [29], which converts a vector of K real numbers into a probability distribution of K possible outcomes.

In the study, three ANN structures were developed, namely: CNN, which consists of 17 layers, LSTM of the first variant, which consists of 7 layers, and LSTM of the second variant, which consists of 5 layers. A comparison of the ANN structures is shown in Table 1.

Table 1. ANN structure comparison

NN type	No of layers	Training time [min]	Obtained validation accuracy
LSTM_1	7	30	96.96
LSTM_2	5	21	99.33
CNN	17	53	96.96

The best network structure used in the study consists of the following layers (Fig. 8):
- input layer which applies data normalization,
- a dropout layer randomly sets input elements to zero with a given probability,
- a fully connected layer multiplies the input by a weight matrix and then adds a bias vector,
- a SoftMax layer applies a SoftMax function to the input,
- a classification layer computes the cross entropy loss for multi-class classification problems with mutually exclusive classes.

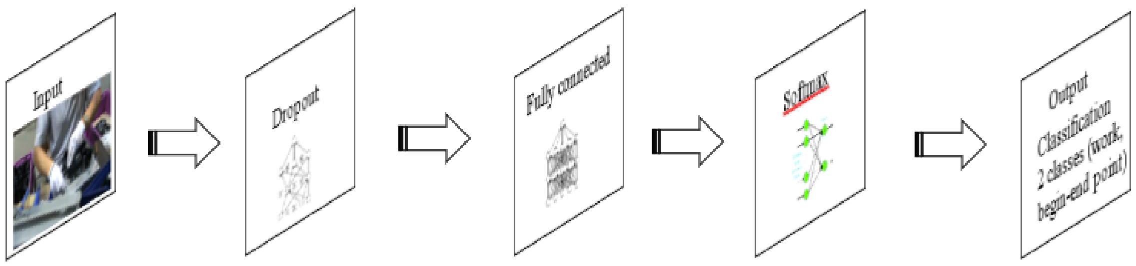

Fig. 8. The best network structures

The training process included 30 epochs and 570 iterations with a learning rate of 0.01. The obtained validation accuracy is 99.33%.

The results are promising, but there is still area for further research. The designed NN structure gives good results for the standard assembly process, but the analysed process is performed by a human, so it may happen that the assembly task is not performed according to the standard method. In such a case, the ANN will not recognize the assembly operations correctly.

Automatic analysis of the assembly process can be used as a tool for process control as well as a tool for process improvement.

To use ANN, it is necessary to create a training set which is, among others, a method for developing assembly process documentation. Developing a training set creates a base of best practices. Work standards can be improved in accordance with the principle of the employee being the best at his job.

5. CONCLUSIONS

ANN is a powerful tool for data analysis. Among many ANN architectures, DNN is useful for image classification. The proposed approach used deep learning ANN to identify the begin-end points of assembly task. In the study three DNN architectures were compared and the best structure was chosen. LSTM consisting of 5 layers give the best results in begin-end point classification in a manual assembly process. The research was conducted on an assembly workstation where four components were assembled using a hand tool.

The assembly process is a crucial part of the manufacturing process, so analysis and measurement of the work should be performed. Each assembly task should be described with the identification of begin-end points.

Conventional assembly task analysis is based on human-observed tasks and is time consuming. Automating assembly analysis provides an opportunity to identify errors in the manufacturing process and improve them.

Future research will focus on developing image features that can be used to identify the begin-end points. In the presented approach, the begin-end point was identified based on the experience of an industrial engineer, but future research will focus on developing an algorithm that will be useful in identifying the begin-end points based on image features.

The presented study was limited to 2D images recognition, so future research can focus on 3D analysis.

REFERENCES

[1] PUTNIK G., SHAH V., PUTNIK Z., FERREIRA L., 2020, *Machine Learning in Cyber-Physical Systems and Manufacturing Singularity – It Does not Mean Total Automation, Human is Still in the Centre: Part I – Manufacturing Singularity and an Intelligent Machine Architecture*, Journal of Machine Engineering, 20/4, 161–184, https://doi.org/10.36897/jme/131000.

[2] PUTNIK G., SHAH V., PUTNIK Z., FERREIRA L., 2021, *Machine Learning in Cyber-Physical Systems and Manufacturing Singularity – It Does not Mean Total Automation, Human is Still in the Centre: Part II – In-CPS and a View from Community on Industry 4.0 Impact on Society*, Journal of Machine Engineering, 21/1, 133–153, https://doi.org/10.36897/jme/134245.

[3] UHLMANN E., POLTE M., BLUMBERG J., KRAFT Z., 2021, *Hyperparameter Optimization of Artificial Neural Networks to Improve the Positional Accuracy of Industrial Robots*, Journal of Machine Engineering, 21/2, 47–59, https://doi.org/10.36897/jme/134275.

[4] LÜ Q., LIU T., ZHANG R. et al., 2022, *Generation Approach of Human-Robot Cooperative Assembly Strategy Based on Transfer Learning*, J. Shanghai Jiaotong Univ. (Sci.), 27, 602–613.

[5] WANG P., LIU H., WANG L., GAO R., 2018, *Deep Learning-Based Human Motion Recognition for Predictive Context-Aware Human-Robot Collaboration*, CIRP Annals – Manufacturing Technology, 67, 17–20.
[6] SINHA S., FRANCIOSA P., CEGLAREK D., 2020, *Object Shape Error Response Using Bayesian 3D Convolutional Neural Networks for Assembly Systems with Compliant Parts*, IEEE 18th International Conference on Industrial Informatics (INDIN), Warwick, United Kingdom, 104–109, https://doi.org/10.1109/INDIN45582.-2020.9442245.
[7] SINHA S., FRANCIOSA P., CEGLAREK D., 2021, *Building a Scalable and Interpretable Bayesian Deep Learning Framework for Quality Control of Free Form Surfaces*, IEEE Access, 9, 50188–50208, https://doi.org/10.1109/ACCESS.2021.3068867.
8] COPOT C., BURLACU A., LAZAR C., 2009, *Image Moments Based Visual Control Algorithm for Servoing Systems*, IEEE 5th International Conference on Intelligent Computer Communication and Processing, Cluj-Napoca, Romania, 157–160, https://doi.org/10.1109/ICCP.2009.5284768.
[9] CHAUMETTE F.,2004, *Image Moments: a General and Useful set of Features for Visual Servoing*, in IEEE Transactions on Robotics, 20/4, 713–723, https://doi.org/10.1109/TRO.2004.829463.
[10] WANG F., CUI B., LIU Y., REN B., 2022, *Deep Reinforcement Learning for Peg-in-hole Assembly Task Via Information Utilization Method*, Journal of Intelligent & Robotic Systems, 106, 16.
[11] GROOVER M., 2016, *Work Systems, the Methods,Measurement and Management of Work*, Pearson.
[12] KUTSCHENREITER-PRASZKIEWICZ I., 2015, *Time Study*, Wydawnictwo Naukowe Akademii Techniczno-Humanistycznej, (in Polish).
[13] GAVALI P., BANU S., 2019, C*hapter 6 – Deep Convolutional Neural Network for Image Classification on CUDA Platform*, Editor(s): Arun Kumar Sangaiah, Deep Learning and Parallel Computing Environment for Bioengineering Systems, Academic Press, 99–122.
[14] BURLACU A., COPOT C.,·LAZAR C., 2014, *Predictive Control Architecture for Real-Time Image Moments Based Servoing of Robot Manipulators*, J. Intell. Manuf., 25, 1125–1134.
[15] MARCHAND E., CHAUMETTE F., 2005, *Feature Tracking for Visual Servoing Purposes*, Robotics and Autonomous Systems, 52/1, 53–70.
[16] CHAUMETTE, F.,1998, *Potential Problems of Stability and Convergence in Image-Based and Position-Based Visual Servoing*, The Confluence of Cision and Control, 66–78.
[17] SANGEETHA D., DEEPA P.,2017, *A Low-Cost and High-Performance Architecture for Robust Human Detection Using Histogram of Edge Oriented Gradients*, Microprocessors and Microsystems, 53, 106–119.
[18] MADHAVAN S., JONES M., 2017, *Deep Learning Architectures. The Rise of Artificial Intelligence*, https://developer.ibm.com.
[19] BOESCH G., 2023, *Image Recognition: The Basics and Use Cases*, https://viso.ai/computer-vision/image-recognition.
[20] *What is deep learning*? https://www.ibm.com/topics/deep-learning.
[21] LECUN Y, BENGIO Y, HINTON G., 2015, *Deep Learning*, Nature, 521(7553), 436–444, https://doi.org/10.1038/nature14539.
[22] CIRESAN D., MEIER U., SCHMIDHUBER J., 2012, *Multi-Column Deep Neural Networks for Image Classification*, IEEE Conference on Computer Vision and Pattern Recognition, 3642–3649.
[23] KRIZHEVSKY A., SUTSKEVER I., HINTON G., 2012, *ImageNet Classification with Deep Convolutional Neural Networks*, Advances in Neural Information Processing Systems, 25, 1097–1105.
[24] Long Short-Term Memory (LSTM) https://uk.mathworks.com/discovery/lstm.html.
[25] ALVES G., 2018, *Discovering SOM*, An Unsupervised Neural Network, https://Medium.Com/Neuronio/-Discovering-Som-An-Unsupervised-Neural-Network-12e787f38f9.
[26] ZAMORA-HERNANDEZ M., CASTRO-VARGAS J., AZORIN-LOPEZ J., GARCIA-RODRIGUEZ J., 2021, *Deep Learning-Based Visual Control Assistant for Assembly in Industry 4.0*, Computers in Industry 131,103485.
[27] KRÜGER J., LEHR J., SCHLÜTER M., BISCHOFF N., 2019, *Deep Learning for Part Identification Based on Inherent Features*, CIRP Annals – Manufacturing Technology, 68, 9–12.
[28] SRIVASTAVA N., HINTON G., KRIZHEVSKY A., SUTSKEVER I., SALAKHUTDINOV R., 2014, *Dropout: A Simple Way to Prevent Neural Networks from Overfitting*, Journal of Machine Learning Research, 15, 1929–1958.
[29] BISHOP C., 2006, *Pattern Recognition and Machine Learning*, Springer.