

Article

An Automated Recognition of Work Activity in Industrial Manufacturing Using Convolutional Neural Networks

Justyna Patalas-Maliszewska ¹, Daniel Halikowski ² and Robertas Damaševičius ^{3,*}

¹ Institute of Mechanical Engineering, University of Zielona Góra, 65-417 Zielona Góra, Poland; J.Patalas-Maliszewska@iim.uz.zgora.pl

² Faculty of Technical Science, University of Applied Science in Nysa, 48-300 Nysa, Poland; daniel.halikowski@pwsz.nysa.pl

³ Department of Applied Informatics, Vytautas Magnus University, 44404 Kaunas, Lithuania

* Correspondence: robertas.damasevicius@vdu.lt

Abstract: The automated assessment and analysis of employee activity in a manufacturing enterprise, operating in accordance with the concept of Industry 4.0, is essential for a quick and precise diagnosis of work quality, especially in the process of training a new employee. In the case of industrial solutions, many approaches involving the recognition and detection of work activity are based on Convolutional Neural Networks (CNNs). Despite the wide use of CNNs, it is difficult to find solutions supporting the automated checking of work activities performed by trained employees. We propose a novel framework for the automatic generation of workplace instructions and real-time recognition of worker activities. The proposed method integrates CNN, CNN Support Vector Machine (SVM), CNN Region-Based CNN (Yolov3 Tiny) for recognizing and checking the completed work tasks. First, video recordings of the work process are analyzed and reference video frames corresponding to work activity stages are determined. Next, work-related features and objects are determined using CNN with SVM (achieving 94% accuracy) and Yolov3 Tiny network based on the characteristics of the reference frames. Additionally, matching matrix between the reference frames and the test frames using mean absolute error (MAE) as a measure of errors between paired observations was built. Finally, the practical usefulness of the proposed approach by applying the method for supporting the automatic training of new employees and checking the correctness of their work done on solid fuel boiler equipment in a manufacturing company was demonstrated. The developed information system can be integrated with other Industry 4.0 technologies introduced within an enterprise.

Keywords: employee work recognition; employee training support; human resources management; business intelligence; deep learning; Industry 4.0



Citation: Patalas-Maliszewska, J.; Halikowski, D.; Damaševičius, R. An Automated Recognition of Work Activity in Industrial Manufacturing Using Convolutional Neural Networks. *Electronics* **2021**, *10*, 2946. <https://doi.org/10.3390/electronics10232946>

Academic Editors: Agapito Ledezma Espino and Araceli Sanchis de Miguel

Received: 5 October 2021

Accepted: 23 November 2021

Published: 26 November 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The Fourth Industrial Revolution, often known as Industry 4.0, is based on the Industrial Internet of Things (IIoT) and other technology enablers such as Artificial Intelligence (AI), digitization and automation [1]. Its goal is to establish direct communication between industrial machinery, people, and processes. IIoT can more critically bring considerable gains in productivity, product quality, and safety through proactive detection of problems by tapping and analyzing such data. While most IIoT research is currently focused on predictive maintenance of industrial machines (unplanned production stoppages result in significant increases in costs and lost plant productivity), monitoring, assessing, and improving worker productivity and performance is a future challenge of the Industry 4.0 system [2]. In fact, human workers are the most dynamic factor in any advanced intelligent manufacturing system; so, any development in this area must account for the concept of human-centered intelligent manufacturing (HCIM) [3]. To develop such human-centered systems, the main task is to understand human behavior that leads to achieving the optimal

work performance. This field of research is especially significant for large manufacturing plants, where most production tasks are carried out by several workers in a production line that follows complex work routines. The factory workers' assembly labor is still at the heart of the industrial manufacturing system, and improving assembly work is one of the most critical tasks for increasing productivity [4]. The awareness of worker activities in spatially-extant and distributed manufacturing facilities is a pre-requisite for efficient workflow organization [5]. For example, a worker in a conveyor line-based production system conducts predetermined operation procedures repeatedly, each of which consists of a sequence of operations. Automatic and accurate worker activity detection is important for work performance management [6], evaluating work efficiency [7], assessing workloads and reducing the risk of injuries [8], and preventing safety accidents [9]. In general, it contributes to the sustainability of work practices [10]. However, observing and analyzing the worker activity workflow together with machines and tools in the middle of the industrial production process in the real-world manufacturing environment is difficult.

The growing need for IT innovation in manufacturing enterprises, according to Industry 4.0, has brought forward a great challenge to adopt advanced computer vision methods and Convolutional Neural Networks (CNNs) to recognize work activity and automatically generate work instructions in the workplace [11]. The use of deep CNNs to solve the problem of recognizing working practices, is promising [12–14] due to the integration of three elements: (1) object detection, (2) human pose, and (3) the recognition of work activity [15,16]. The applications developed in this manner—and based on such manufacturing technologies—are capable of learning-by-doing and, more importantly, are capable of self-improvement [17].

Nowadays, managers are looking for innovative solutions that will be helpful when deciding on the employment of new employees, by training them individually without involving other employees, to increase innovative working behaviour at the work place [18], and to keep the employees interested in boring day-to-day activities [19]. Current studies of manual manufacturing work often use monitoring of security cameras, checklists, and (often imprecise) work logs. The automated assessment and analysis of employee activity is essential for a quick and precise diagnosis of work quality, especially in the process of training a new employee. Overall productivity assessment, progress review, labor training programs, and safety and health management all require effective and timely analysis and tracking of personnel operations. Currently, the employee training process is costly, time-consuming, and requires the involvement of experienced employees who are required to supervise and control this process. The implementation of a dedicated information system, based on deep learning (DL) techniques, can smoothen the course of the training in checking the employee progress and correcting any mistakes.

Human action recognition has been used as a method of automatically analyzing and comprehending worker activities to provide real-time help and facilitate worker-machine collaboration [20]. The integration of ambient sensing technologies such as wearable sensors or surveillance cameras, and artificial intelligence-based analysis using deep learning models leads to the rise of the Worker 4.0 [21] that implements the main principles and behavior of workers in an Industry 4.0 scenario.

The digitization of the industrial workplace through the ubiquity of sensors, combined with digital information systems and intelligent monitoring, generates huge amounts of data every day, which capture the factual manual workflows [22]. The creation of workflow models to control, analyze, and optimize such industrial workflows, if done manually, is time consuming and costly [23]. To make the most of available monitoring data, we need to create a workflow recognition framework that can automatically extract features from uncut videos to recognize human and machine behavior. In this work, we have developed an information system supporting the automatic training of new employees and checking their work. We have built a training set for workers' service procedure when checking a solid fuel boiler in a manufacturing company, by collecting data from videoed work activity sequences, using the DL approach. The camera is the most popular type of

technology when it comes to recognizing activity and can be used for video or image-based recognition, using trained models to capture various activities [24]. According to Ijjina and Chalavadi [25] and Chen et al. [26], using a camera with deep learning techniques would be a good solution for detecting activity.

In our approach, the stages of the procedure and their duration time were extracted and determined. Next, the features obtained from the video are given as input to CNN to learn the discriminating features. We use the objects of the workplace to assist the detection of each stage of the procedure. To identify objects in an image, CNN is used with a support vector machine (SVM) plus CNN with Softmax and R-CNN respectively, in a series of experiments. The main contributions of this paper are as follows.

(1) For the workflow recognition problem, we propose the integrated CNN, SVM, CNN, Softmax, and R-CNN approach, thus improving the accuracy of recognition.

(2) We design a framework for the automatic generation of work instructions, based on key objects in each stage of the procedure, which is the first attempt to combine real-time action recognition tasks, with specific practical application scenarios.

(3) We demonstrate the practical usefulness of our proposed approach by applying the method to practical systems, supporting the automatic training of new employees, and checking their work on a solid fuel boiler, in a manufacturing company.

2. Research Literature

The domain of worker activity recognition and work process discovery in industrial environments has multiple dimensions such as the type of recognition (supervised, unsupervised recognition, and semi-supervised), the type of sensor used for recognition (motion, vision, sound and radio signal -based) of worker movement, and sensor location (wearable, ambient and attached to objects) [27]. In the computer vision literature, the task of recognizing an image is defined as recognizing that an object belongs to certain classes [28]. The object detection algorithms ‘recognize’ based on the training sequence data containing the objects, together with information about their classification [29,30]. In the deep learning (DL) approach, algorithms are designed to model complex levels of data abstraction using multiple layers of non-linear transformations [31] with artificial, neural networks assuming the role both of generator and classifier of diagnostic features. Computer image recognition systems have found an application in industry with the solutions implemented using both machine learning methods and DL techniques [32]. Image recognition systems use the classification and reduction techniques of features based on Kernel Principal Component Analysis (KPCA), Linear Discriminant Analysis (LDA), Support Vector Machine (SVM), Hidden Markov Model (HMM) [33], Decision Trees (DT), K-Nearest Neighbours (KNN), Random Forest (RF) and Discriminant Analysis (DA) [34].

The limitations related to the machine learning approach, consisting of processing natural data in its raw form, resulted in the introduction of solutions based on the DL techniques. Thanks to this approach, the data analyzed was transformed into an appropriate, internal representation (feature vector) from which the learning subsystem can detect or classify patterns [35]. The DL techniques based on Artificial Neural Networks (ANNs) are used in human activity recognition systems [36], systems with Myo bands [33,37], or with sensors such as Microsoft Kinect [38,39] and Intel RealSense [40], accelerometer, gyroscope, or sensors of mobile devices (such as smartphone) [41] or wearable devices [42] as well as in systems based on video analysis [11]. To recognize a worker in action, smart manufacturing systems may use the ultrasonic sensors, Inertial Measurement Unit (IMU) [43], and the surface electromyography (sEMG) signals obtained from a Myo armband with Discrete Fourier Transformation (DFT) and Convolutional Neural Network (CNN) [37].

Activity recognition in the manufacturing area with ultrasonic and IMU sensors has been used to recognize worker activity in bicycle maintenance scenarios, in car manufacturing and to capture arm movements for classification into five activities using signals from a smartwatch for industrial assembly lines, so that factory working time can be estimated [37]. One interesting approach, when analyzing human activity, is based on

the Wi-Fi signal analysis method [44]. The principle of wireless detection is that human activities affect the propagation of the RF signal; different activities can cause different patterns of signal change. These techniques are used to process signals collected from ambient and wearable sensors (motion, proximity, microphone, video sensors, accelerometers, Body Sensor Networks, gyroscopes). The technique has been proposed to detect human activities [45], detect falls [46], recognize person's gait [47], predict body pose [48,49], classify gestures [50] and extract information on movement during interactive games and exercises [38]. Such systems may find applications in intelligent home healthcare systems and assisted living environments [51] and can monitor patients by diagnosing their health and controlling their drug intake. They can be used for the automatic surveillance of public places and to detect criminal activity [52], as well as to observe the worker's behaviour at workplace for the signs of fatigue or stress [53]. The activity signature patterns observed in the signals received from wearable motion sensors of ubiquitous smartphones are analyzed to recognize activities performed by different construction workers [54]. Previous studies have either employed remote-monitoring sensors such as RGB+D cameras [23] and RFIDs; or worker-attached wearable sensors such as accelerometers, gyroscopes and magnetometers [55].

Thanks to advancements in computer vision algorithms, practitioners and researchers may now use camera sensors to give semi-real-time information on worker actions at a cheap cost [56]. Object detection, object tracking, and hence activity recognition were all part of the development of camera-based systems [57]. In the case of industrial solutions, the DL approach is used in autonomous mobile robots, vision systems for cars, speech recognition, smartphones, cameras, and digital cameras, and robot control. CNN networks currently dominate almost all tasks related to recognition and detection. Despite the wide use of the CNN network, it is difficult to find solutions supporting the automated control and evaluation of activities performed by an employee in enterprises. A frequent element of activity control systems, in the context of the performance and evaluation of work, are CNN models based on data from various types of sensors. Al-Amin et al. [20] presented a method in which workers wear two devices on their hands to acquire IMU data while constructing things. Two CNN action recognition models, a left-hand model and a right-hand model, were developed independently and fused together to produce a better activity recognition result. AR models are refined through transfer learning, which allows them to adapt to new personnel. The approach was validated in the assembly of Bukito 3D printers. Angah and Chen [58] used Mask R-CNN for multiple worker tracking on the construction job site. Multi-Object Tracking Accuracy (MOTA) that considers the mismatch problem, was used to evaluate the tracking performance. Hu et al. [59] used structured two-stream convolutional neural networks (CNNs) to recognize the behavior of workers and machines. CNNs extracted the spatial-temporal activity features and included the attention mechanism to detect important behavior.

The literature also includes models for recognizing employees' dangerous activities, using the convolutional neural network and the LSTM network [60]. Zhao and Obonyo [61] proposed using Deep Neural network models through integrating CNN with Long Short-Term Memory (LSTM) for recognizing construction workers' postures from motion data, captured by wearable Inertial Measurement Units (IMUs) sensors. Similarly, Yang et al. [62] used on-body inertial sensors and deep learning to analyze workloads of worker lower body during a physical load carrying task. The LSTM-based technique is also used to analyze emotions based on signal changes from EEG [63]. Gong et al. [64] proposed a deep learning model for recognizing the activities of workers on the offshore drilling platform, using the characteristics of the human body's key points which remain unaffected by complex background noise, to assist the detection of the human target. HMM and Naive Bayes classifier were used to recognizing employee activity in production processes using the Kinect sensor. Zeng et al. [65] proposed isolating discriminant features to recognize the activity, based on CNN and sensors of mobile devices. Jaouedi et al. [66] presented an approach, based on an analysis of video content in which functions are based on all

visual characteristics of each frame of a video sequence, using the Gated Recursive Neural Network (RNN) unit (GRU) model. Pohlt et al. [67] suggested the Inflated 3D Convolutional Network (I3D ConvNet) as image encoder and then the Graph Convolutional Networks (GCN) as keypoint extractor for worker activity recognition. Tao et al. [68] performed assembly operation recognition in HCIM environment using image frames obtained from a visual camera. The recognition is performed in real time using a deep learning model trained by adopting a transfer learning approach. Son et al. [69] adopted very deep residual networks (ResNet-152) for feature map extraction and Faster regions with CNN features (R-CNN) for labeling to detect construction workers with different poses and against varying backgrounds in industrial image sequences. Sun et al. [70] employed the generative adversarial networks (GANs) for estimation of human body joints and performed worker efficiency analysis as a temporal action localization problem. First the teacher performed exemplar work activities recorded in a reference video. Then the video of activities performed by the worker is matched (using dynamic time warping) against a reference video using invariant spatio-temporal features extracted from the worker body posture sequences and perform cross-video matching. A detailed analysis of the in the form of a comparison study was made in our previous works: [11,71]. In this work, the main contribution is to build Algorithm 1; Algorithm 2 and to apply the proposed framework in the form of an integrated system to support the automatic training of new employees.

Summarizing, the DL approach is used in many fields, as the authors have presented in [11]. However, in the case of work quality and performance control systems on the market, there is a research niche of appropriate solutions using the DL approach in the process of training and verifying employee skills in a manufacturing company that would employ the benefits of smart learning and artificial intelligence [72]. There are attempts to implement solutions checking the activity of people in the workplace (such the correct sitting posture of office workers [40] and postures of construction workers [73]), but no studies were done regarding effective solutions for the automatic evaluation of the work of a newly hired employee.

3. Materials and Methods

3.1. The Service Used as a Case Study

The solid fuel boiler service procedure analyzed, consisting of five service activities as follows:

1. stopping the furnace operating,
2. checking the solid fuel tank,
3. checking the gear motor and auger,
4. assembling the auger and the gear motor, and
5. tightening the mounting screws of the gear motor and mounting the cleanout.

Each service step consists of stages, based on the relevant reference frames of the video material referenced. The total number of steps in the service procedure was 38. The image with the general framework was presented in Figure 1.

The set of graphic files were created, which contained class objects related to the process of servicing the solid fuel boiler. The training set contained 3440 graphic files to include a total selection of objects from 11 classes, namely, a person—925, a hand—1065, a solid fuel boiler—355, a controller—407, a shovel—32, a bucket—758, a plastic storage box—370 an auger—317, a solid fuel tank—716, a gear motor—1253, and a spanner—421. In total, the training set contained 6619 marked objects for 11 classes. Eight video sequences were tested showing the correctly performed service procedure for checking a solid fuel boiler. This set was created automatically from entries in a configuration file, created with Data Transformation Language. The configuration settings assumed that 80% of the dataset would be allocated to training data and 20% to validation data. Ultimately, 2787 (81.02%) images were included in the training set and 653 (18.98%) in the validation set.



Figure 1. Four steps of sample activity (assembly of the auger and gear motor).

3.2. New Approach to the Automatic Generation of Work Instructions

The new approach to the automatic generation of work instructions has been designed according to a flowchart (Figure 2).

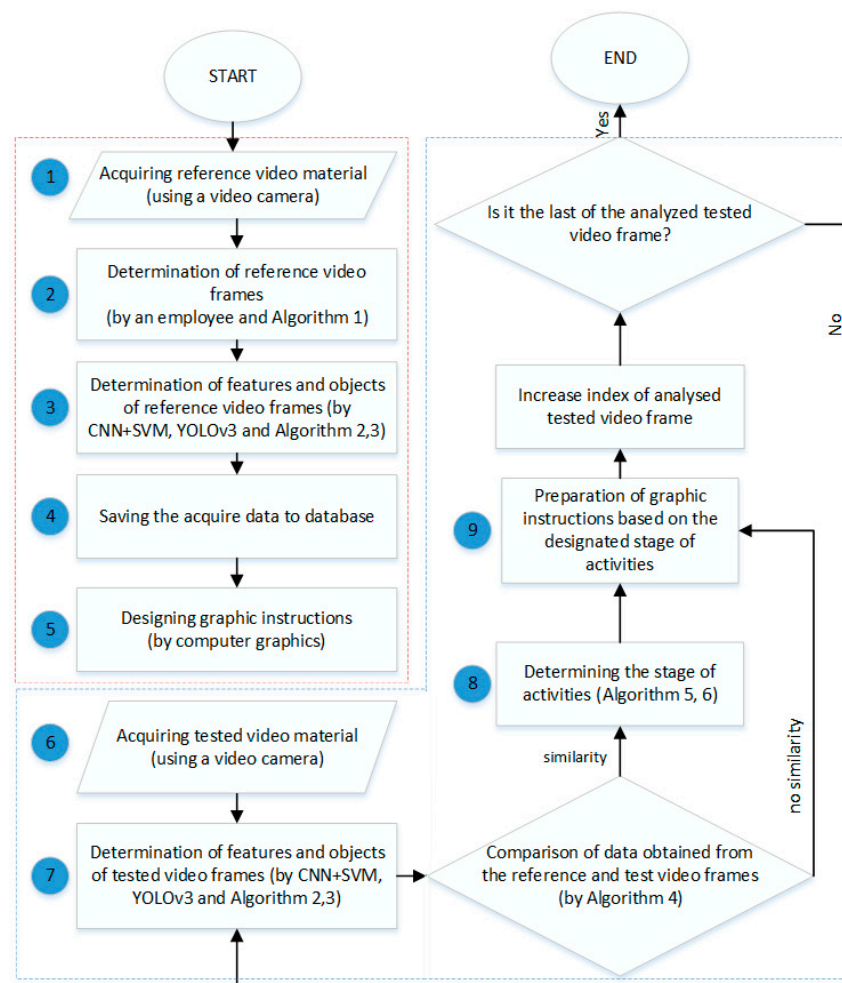


Figure 2. Flowchart of the approach to the automatic generation of work instructions.

The system's preparatory stages for supporting the automatic training of new employees and checking their work for clearance are marked in red, with a dashed line. Correct working practices, in the system, are marked in blue with a dashed line. In the so-called preparatory stages (Figure 2), the video material referenced, showing a service procedure being correctly performed, should be recorded. After determining the reference frames corresponding to the individual stages of the service procedure, the next step was to prepare the CNN to detect objects on the video sequence. For this purpose, artificial neural network training was carried out with the use of a dataset containing classes of objects appropriate for the service process.

The proposed CNN architecture uses SVM and Softmax to extract video frame features and information regarding the object, from the images, that is, the establishment of the referenced features and objects and the video frames tested. To train the CNN, an original dataset was used created by the authors. The dataset contains images of training materials used in the process of training employees to operate a solid fuel boiler, with selected object classes. The implementation of tasks in the system, relating to the analysis and interpretation of the image, requires the correct identification of the elements in the image. This is connected, not only with choosing the right network architecture but also with proper training thereof. Publicly available datasets made it possible to train the network to detect a person-class object, but they did not allow other objects to be detected, such as the gear motor, the controller, or the auger. Therefore, the training set had to be created from scratch, and the elements that the system was to identify had to be determined manually. This process turned out to be very time-consuming, as the effective operation of the CNN network as a detector of objects requires training on a large dataset.

In the new method proposed, 6 algorithms have been created and implemented, the use of which allows video sequences to be processed, to check the service activities performed.

The Algorithms 3–6 were strictly described in the form of pseudo-code in the previous publication of the authors [11]. The task of the Algorithm No. 3 is to analyze the set of reference frames and divide this set into component activities of the service procedure along with the determination of the stages of each activity. The START/STOP rule has been defined, according to which the set of reference frames is divided into subsets (each subset is a separate activity). The rule is a video frame where the technician and the tool board will be identified. The effect of the algorithm is a set of service activities containing the steps corresponding to this activity. The task performed by Algorithm 3 is described in [11]. The task of the algorithm No. 4 is to extract the features of the previously determined reference frames and test frames using CNN and SVM. The acquired features make it possible to identify the activity currently performed by the technical employee, and the feature extraction algorithm is an element of the two-stage process of determining the stage of service activities [11]. The task of the Algorithm No. 5 is to determine the data set containing the labels of the classes of objects located in the previously determined reference frames and test frames [11]. Technologies used in Algorithm 5 are based on CNN, R-CNN and YOLOv3 networks. The set of labels generated for each frame, corresponding to the names of classes of identified objects, will be used in the two-stage process of identifying the currently performed activity. The task of Algorithm 6, described in [11], is to identify the step of the activity by comparing the labels of the object classes and the characteristics of the frames of the test material, and the characteristics and labels of the object classes for the reference frames. The features of the test video material were extracted using the CNN network.

The pseudo-code of Algorithm 1, designed to generate a set of graphic instructions under the stage of the service operation, is presented as follows:

Algorithm 1:

```

initialization;
Input  $\leftarrow (k, l)$ ; // data from algorithm 6
 $z \leftarrow 0$ ;
for  $i = 1$  to 3 do
  begin
    inc( $z$ );
    if  $(l + z - 1 == \text{size}(I[k][l]))$  then
      begin
        inc( $k$ );
         $z = 0$ ;
         $l = 1$ ;
      end;
    send  $\rightarrow (I[k, l + z])$ ;
  end;
end

```

where: k, l —table indices, corresponding to the identified stage of service activities, i —loop counter, z —auxiliary variable, $I [] []$ —jagged array storing graphic instructions for each step of the activity, $k, l, i, z \in N$.

The pseudo-code for Algorithm No. 2, which analyses the times of the stage of activities, based on which it will be possible to identify irregularities in the service works performed, is as follows:

Algorithm 2:

```

initialization;
activity = get(activity)
stage = get(stage)
time = find(act_time, next(activity, stage))
if time == 0 then time = find(act_time, next(activity, stage))
a = time_now
while ( $a < \text{time}$ )
  begin
    send  $\rightarrow$  activity, stage
  end;
  send  $\rightarrow$  next(activity, stage)
end

```

where: *activity*—a variable that stores information about the current activity, *stage*—a variable that stores information about the current stage of the activity, *time*—a variable that stores information about the time allocated to a given stage of activity, *act_time*—a one-dimensional array that stores times specified for each stage, *next*—a function that determines the next stage of the service procedure, *a*—a variable that stores current time.

The process of training the artificial neural network, testing the activity detection system, and generating the scenario of conduct was carried out on a unit with the following configuration: processor AMD Ryzen 5 3600, motherboard MSI B450 TOMAHAWK MAX, RAM: PC CRUCIAL SPORT DDR4, 16GB, 3200MH, graphics card: Gigabyte GeForce RTX 2060 SUPER Gaming OC 8GB GDDR6, SSD: 2xPatriot Burst 240GB—Raid 0.

Network training was performed on the Ubuntu 19.04 operating system with CUDA 10.2 and cuDNN v7.6.4 software and the Docker Engine installed. The operation of the employee activity detection system was tested on the Windows 10 operating system, with the CUDA 10.2 and cuDNN v7.6.4 software as well as the PyTorch and OpenCV frameworks installed. The implementation of tasks related to the information system, used for image analysis, assumed the use of the Yolov3 network architecture for the purpose. The Yolov3 network has been used before for monitoring workers activity in construction site video sequences [74].

The research used a convolutional neural network based on the Yolov3 and Yolov3 Tiny architecture. The choice of architectures based on Yolov3 was due to the fact that they ensured high efficiency in recognizing objects contained in the tested video material using a PC class machine. The YOLOv3 architecture is based on the Darknet53 model and has 53 convolutional layers. A more efficient version of YOLOv3 is the YOLOv3-Tiny version which is based on the Yolov3 architecture, but has a less complex architecture and consists of 24 layers. The ANN, based on this model made it possible to detect objects of all classes contained in the dataset, but the speed of operation of the IT system was limited by the technical capabilities of the machine on which the application was launched. The speed of operation of the detection system for the Yolov3 model, after implementing the proprietary algorithms was about 5 frames per second. In connection with the above, the network architecture based on the Yolov3 Tiny model was adopted, which allowed the processing speed of about 26–27 frames per second to be obtained. The process of training the network to use the created dataset for the Yolov3 Tiny model took about 5 h, while for the Yolov3 model, it took about 12 h.

The efficiency of establishing the appropriate frames using CNN + SVM was about 94% (Resnet18 94%, Alexnet 87.51%) and with YOLOv3 73.15% (Cifar10Net: 61.66%, Alexnet: 48.33%).

Table 1 presents a comparison with state-of-the-art alternatives in order to highlight the new contribution of our work.

Table 1. Comparing the research results with state-of-the-art alternatives.

Paper	Method	Dataset	Type	Accuracy
[60]	CNN + LSTM	original dataset created by the authors	safe actions/unsafe action	97/92
[34]	CNN + SVM	bearing vibration dataset, Case Western Reserve University	ISAX based features	88
[33]	CNN, RNN	original dataset created by the authors		97
[75]	CNN + LSTM	DEAP dataset	2 (Hv/Lv))/SEED-dataset 3	84.16/90.81
[66]	RNN + Gated Recurrent Unit	UCF Sports, UCF101, KTH	KTH (GMM + KF)	71.1
[39]	Markov models, Naive Bayes, K-means	original dataset created by the authors	point-by-point/point-by-point with FSHMM	70/65.8
This paper	CNN, CNN + SVM/YOLOv3	original dataset created by the authors	Steps and objects of the service procedure	94.01/73.15

4. Research Results—An Integrated System to Support the Automatic Training of New Employees

Based on the proposed methodology (Figure 2), an information system to support the automatic training of new employees and verify their work was built. In the first stage, the system is prepared to work with test material while in the second stage, the test material is analyzed. The test material is a video recording of service activity correctly done.

4.1. Preparing the System to Work

The implementation of the IT system will facilitate the following: the extraction of reference frames (including START/STOP frames); the division of reference frames into sets of frames, corresponding to individual activities and their stages; the extraction of reference frame features and their transformation, where the three-dimensional tensor is transformed into a two-dimensional form, the extraction of objects in reference frames, the preparation of sets of graphic instructions, the storing of information about objects,

reference frame characteristics and file names for graphic instructions, the designation of activities and steps to be controlled by the system.

From the reference material obtained, reference frames corresponding to the individual stages of the activity were separated manually. For each reference frame, features were determined, via the extraction of features from an artificial neural network, in the form of a three-dimensional tensor, which was processed into a two-dimensional form. Objects in reference frames were also identified. Information about the characteristics and objects of the reference frames was saved in the system. Obtaining information about objects for each frame can be done automatically or manually with the decision being made by the system's user, based on the value of the program variable. The automatic detection of objects is performed, using the built-in code and the Python functions implemented additionally. Designated features and object class labels were used to identify the activity trained to work in a given employee's position, by comparing the characteristics and labels of the reference cage with the test cages. An example of an array of objects for activity 1 solely, appears thus:

```
ref_objects = ([ "",
    ["Gear motor"],
    ["Gear motor", "Hand"],
    ["Gear motor", "Hand", "Spanner"],
    ["Gear motor"],
    [ "" ],
    ["Spanner"],
    ["Hand", "Spanner"],
    ["Gear motor", "Spanner"],
    ["Gear motor", "Bucket"],
    [ "" ],
    ["Gear motor"],
    ["Controller", "Hand", "Bucket"],
    ["Gear motor", "Bucket"],
    ["Gear motor", "Hand", "Bucket"],
    [ "" ] )
```

A START/STOP frame was designated based on which activities and stages of the service procedure were distinguished. The appearance of this type of frame in the film means the end of service. The video frame features were extracted using an ANN, processed from a three-dimensional tensor to a two-dimensional form, and saved to a file. Next, graphic instructions were prepared corresponding to each stage of the service procedure. For testing purposes, the instructions contained only information about the activity and the stage of the activity performed.

The production version of the system should contain graphic instructions, suggesting the actions to be performed. The user of the system, while configuring its operation, has the option to select the activities to be controlled. The work activities and steps are determined through the appropriate configuration of the elements of the activities_to_proc array. An additional element of the system was the time control functionality for the completion of the service phase. In the module wherein the time of the service procedure is not controlled, the system will require each step to be implemented. In the time control module of the service activity, in the event of exceeding the time allotted for the implementation of a stage, the system will display information that the stage has not been completed and the system will go on to control the next stage. The stage is verified based on information about the time allocated to each stage, stored in a previously defined table.

4.2. The Correct Operation of the System

Implementation of the IT system will facilitate the following: START/STOP frame detection; analysis of the frames of the test video, based on the extracted features of the frame and information about the objects on the frame of the test material; detection of user-

defined activities and steps of the service procedure carried out on the test video material; generation of a set of graphical instructions corresponding to the conduct scenario, created based on the activities designated and the stages of activities; the displaying of graphic instructions; the displaying of information about the activity and step of the activity, the results of comparing features -for the reference frame and test frames features; the results of the comparison of objects, the common part of the set of labels for the object classes of the reference frame and the set of labels for the object classes of the test frames.

In the first stage of the operation of the system, in detection mode, the START/STOP frame features are loaded, based on which, the beginning and end of the service action will be determined, and the test frames will be compared to the feature values for this frame. The actual stage of the detection of activity begins with the analysis of the video sequence. Test material in the form of a video, or material recorded by a camera, is captured and processed in a loop; analysis, thereof, concerns the characteristics of each frame of the data processed. It was assumed that the start of the service procedure -and each activity that is part thereof- begins and ends with the appearance of the START/STOP frame; therefore, the system first compares the features of the test material frame with the features of the START/STOP frame. As a result, the system will search for this type of cage until it finds it. The feature analysis process for test frames begins with changing the tensor size (3D to 2D). The system then compares the features of the test frame, currently being analyzed, with the features of the START/STOP frame. If the features are similar, the system will identify the analyzed frame as a START/STOP frame, and the variable that counts the activities and stages of activities is incremented. Setting activity to 1 and stage to 1 means that the system has started searching for stage 1 of activity 1. In time control mode (mode = 1), it will measure the time from the commencement of the START/STOP frame and if the time for a stage is exceeded, it will go on to the next stage, taking with it information that the stage has not been completed. In the case of the control mode of each stage, from the moment the system identifies the next START/STOP frame, the value of the activity variable will be incremented, and the stage variable will be set to 1. This will enable the search for the next service steps. The system will analyze the image frames until it finds a frame corresponding to the last index of the table, storing information about the reference frames.

When the process of detecting activities and steps has been initiated, that is, when the system has identified the test frame as the first frame, START/STOP, the system acquires the features of the next reference frame. This will be the frame responsible for stage 1 of activity 1. For this purpose, the features of the reference frame that were previously stored there, are read from the file responsible for the stage of activity being searched for. After loading the features of the cage searched, they are compared with the features of the cage of the test material. If the features are similar, labels for the classes of objects in the test frame are set. The identified objects (labels) are compared with the objects assigned to the reference frame. For matching, we use mean absolute error (MAE) as an error metric. If the characteristics and labels of objects on both compared cages match, the stage of the service procedure operation is identified. A window is displayed with information about the stage of service activity found.

4.3. The Results of the Experiments

Analysis of the system efficiency consisted in assessing the detection of activities, the stages of service activities, and the scenario of conduct, along with graphic instructions, generated at a given workstation (see Figure 3).



Figure 3. Sample of the system interfaces demonstrating various functionalities.

Once the system was operating properly, the recorded test video sequences, showing the service procedure being performed by the trained employee, were analyzed. Based on the stage of service activities identified, the system generates a scenario of conduct at a given workstation and displays a set of workstation instructions for four consecutive stages of the service procedure (Figure 1). If the features and classes of objects of the reference frame and test frame currently being analyzed match (see an example of video frame matching results presented in Figure 4), the system will start to detect the next reference frame.

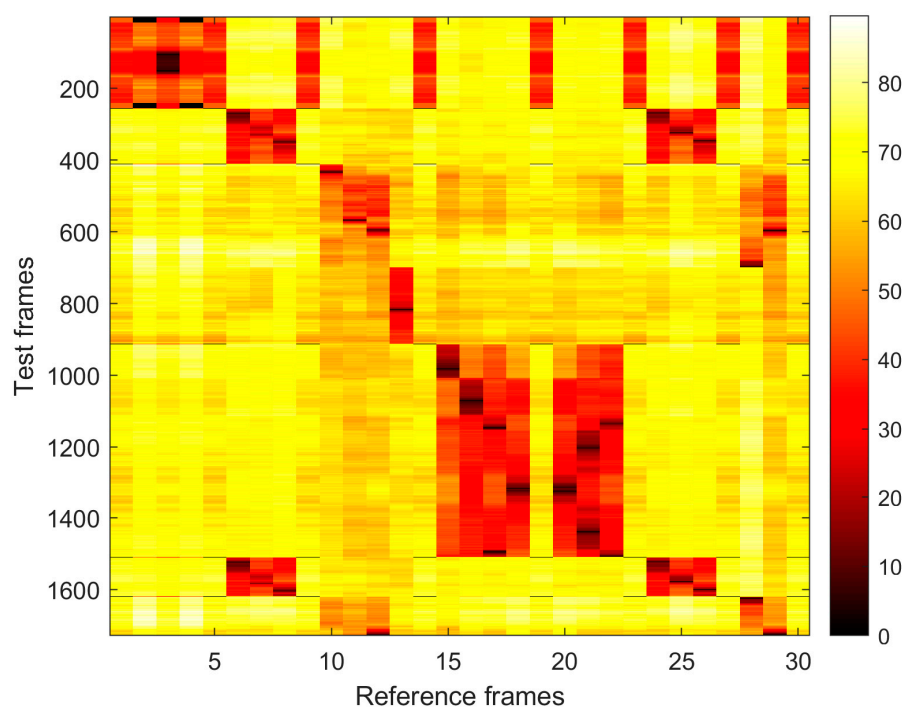


Figure 4. Matching matrix between the reference frames and the test frames using mean absolute error (MAE) as a measure of errors between paired observations. The matrix is used to detect the most similar test frames to the reference frames.

As a result of the analysis of the test video sequences by the system, the following results were obtained (for each of the tested sequences), which confirm its effectiveness: five service activities have been detected, a total of 38 steps have been detected, 11 classes of objects appearing in image frames were detected. So, the proposed system allowed for full control of the service procedure implementation.

Moreover, the proposed system is a solution to the problem of high turnover among employees with specialist knowledge. A frequent element of activity control systems in the context of work performance and evaluation are CNN network implementations based on data from various types of sensors [39,75,76]. So, despite the wide use of CNN, it is difficult to find solutions that allow automation of the training process for new employees without the involvement of an experienced employee in the process of training new employees.

The proposed approach can be extended to cover a knowledge management area in the context of the automation of the process of specialist knowledge transfer to a new employee. The implementation of an integrated system supports the automatic training of new employees in a company where training courses are conducted for employees taking up employment at a given position. During the activities, the new employee is recorded and the captured video frames are analyzed by CNN for the similarity of features and the presence of objects. The information system based on the proposed model includes two modes of monitoring the implementation of the service phase. In the first mode, the system will require the implementation of each stage of activities and will await the performance of the appropriate stage of service activities. In the second mode, in the event of exceeding the time allotted for the implementation of a stage, the system will display information that the stage has not been completed and the system proceeds to control the next stage.

5. Conclusions

Despite the increasing automation levels in emerging Industry 4.0 manufacturing, acquiring and transferring the explicit knowledge of highly skilled manufacturing workers remains a strategic challenge. In this paper we addressed this problem by employing the deep learning techniques for capturing human worker activities in industrial setting.

The research results enabled to build of the integrated system to supporting the automatic training of new employees and verifying their work using based on the deep learning approach based on the data collected on a properly conducted service procedure. The originality of the proposed new framework for the automatic generation of workplace instructions and real-time recognition of worker activities is demonstrated by integrating CNN, R-CNN, YOLOv3 (Yolo Tiny) in the new approach and therefore it is possible to generate the right scenario during the service procedure of solid fuel boiler. The original results of the developed solution are based on a two-stage model of identification of the currently performed activities in the service procedure. The two-stage process of identifying the service activity stage is carried out by comparing the designated reference and test characteristics of video frames and information about the classes of objects located on the analyzed frames. The limitations of our work are lack of simultaneous analysis of data obtained from a larger number of cameras, and lack of the analysis of activities carried out for a different service procedure. In our further works, it is planned to build algorithms, the use of which will increase the effectiveness of verifying the correctness of the activities performed by a new employee thanks to the development of the mechanism for controlling the location of objects and optimization the applied techniques and algorithms of image analysis in order to accelerate the extraction and comparison of image features in real time.

Verification of the performed service activities by an employee is a responsible task, therefore in our further work, it is planned to expand the proposed system with the possibility of simultaneous analysis of data obtained from a larger number of cameras in the workplace. This will allow for more precise control of the service work performed.

Currently, technology maturity, data and knowledge acquiring and sharing, new methods in information systems design, perception, and human–robot interaction are the important challenges in the traditional manufacturing companies. Moreover, in the

context of introducing changes in the manufacturing company operating in accordance with the assumptions of Industry 4.0 mobile platforms can accomplish tasks in workspaces. Therefore, the proposed information system should be integrated with other Industry 4.0 technologies introduced within an enterprise, but this requires further work in this area.

Author Contributions: Conceptualization, J.P.-M., D.H. and R.D.; methodology, J.P.-M. and D.H.; software, J.P.-M. and D.H.; validation, J.P.-M., D.H. and R.D.; formal analysis, J.P.-M., D.H. and R.D.; investigation, J.P.-M. and D.H.; writing—original draft preparation, J.P.-M. and D.H.; writing—review and editing, R.D.; visualization, J.P.-M., D.H. and R.D.; supervision, J.P.-M.; funding acquisition, R.D. and J.P.-M. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the programme of the Minister of Science and Higher Education under the name: “Regional Initiative of Excellence” in 2019–2022 project number 003/RID/2018/19; funding amount 11.936.596.10 PLN.

Data Availability Statement: Data is available from the corresponding author upon reasonable request.

Conflicts of Interest: There are no conflict of interest.

References

- Jimeno-Morenilla, A.; Azariadis, P.; Molina-Carmona, R.; Kyratzi, S.; Moulitanitis, V. Technology enablers for the implementation of Industry 4.0 to traditional manufacturing sectors: A review. *Comput. Ind.* **2021**, *125*, 103390. [\[CrossRef\]](#)
- Forkan, A.R.M.; Montori, F.; Georgakopoulos, D.; Jayaraman, P.P.; Yavari, A.; Morshed, A. An industrial IoT solution for evaluating workers' performance via activity recognition. In Proceedings of the International Conference on Distributed Computing Systems 2019, Richardson, TX, USA, 7–9 July 2019; pp. 1393–1403. [\[CrossRef\]](#)
- Wang, B.; Xue, Y.; Yan, J.; Yang, X.; Zhou, Y. Human-Centered Intelligent Manufacturing: Overview and Perspectives. *Chin. J. Eng. Sci.* **2020**, *22*, 139. [\[CrossRef\]](#)
- Maekawa, T.; Nakai, D.; Ohara, K.; Namioka, Y. Toward practical factory activity recognition: Unsupervised understanding of repetitive assembly work in a factory. In Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing, UbiComp 2016, Heidelberg Germany, 12–16 September 2016; pp. 1088–1099. [\[CrossRef\]](#)
- Stisen, A.; Mathisen, A.; Sorensen, S.K.; Blunck, H.; Kjargaard, M.B.; Prentow, T.S. Task phase recognition for highly mobile workers in large building complexes. In Proceedings of the 2016 IEEE International Conference on Pervasive Computing and Communications, PerCom 2016, Sydney, Australia, 14–18 March 2016. [\[CrossRef\]](#)
- Yin, S.; Li, X.; Gao, H.; Kaynak, O. Data-Based Techniques Focused on Modern Industry: An Overview. *IEEE Trans. Ind. Electron.* **2015**, *62*, 657–667. [\[CrossRef\]](#)
- Luo, X.; Li, H.; Yang, X.; Yu, Y.; Cao, D. Capturing and Understanding Workers' Activities in Far-Field Surveillance Videos with Deep Action Recognition and Bayesian Nonparametric Learning. *Comput. Civ. Infrastruct. Eng.* **2019**, *34*, 333–351. [\[CrossRef\]](#)
- Yu, Y.; Yang, X.; Li, H.; Luo, X.; Guo, H.; Fang, Q. Joint-Level Vision-Based Ergonomic Assessment Tool for Construction Workers. *J. Constr. Eng. Manag.* **2019**, *145*, 04019025. [\[CrossRef\]](#)
- Yu, Y.; Guo, H.; Ding, Q.; Li, H.; Skitmore, M. An experimental study of real-time identification of construction workers' unsafe behaviors. *Autom. Constr.* **2017**, *82*, 193–206. [\[CrossRef\]](#)
- Felsberger, A.; Reiner, G. Sustainable Industry 4.0 in Production and Operations Management: A Systematic Literature Review. *Sustainability* **2020**, *12*, 7982. [\[CrossRef\]](#)
- Patalas-Maliszewska, J.; Halikowski, D. A Deep Learning-Based Model for the Automated Assessment of the Activity of a Single Worker. *Sensors* **2020**, *20*, 2571. [\[CrossRef\]](#)
- Simonyan, K.; Zisserman, A. Two-stream convolutional networks for action recognition in videos. *Adv. Neural Inf. Process Syst.* **2014**, *1*, 568–576.
- Limin, Y.; Yue, L.I.; Bin, D.U.; Hao, P. Dynamic gesture recognition based on key feature points trajectory. *Optoelectron. Technol.* **2015**, *35*, 187–190.
- Kaiming, H.; Georgia, G.; Piotr, D.; Ros, G. Mask r-cnn. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *99*, 1.
- Toshev, A.; Szegedy, C. DeepPose: Human pose estimation via deep neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 1653–1660.
- Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, realtime object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.
- Mourtzis, D.; Angelopoulos, J.; Panopoulos, N. A Framework for Automatic Generation of Augmented Reality Maintenance Repair Instructions based on Convolutional Neural Networks. *Procedia* **2020**, *93*, 977–982. [\[CrossRef\]](#)
- Pham, Q.T.; Pham-Nguyen, A.; Misra, S.; Damaševičius, R. Increasing innovative working behaviour of information technology employees in vietnam by knowledge management approach. *Computers* **2020**, *9*, 61. [\[CrossRef\]](#)
- Ašeriškis, D.; Damaševičius, R. Gamification of a project management system. In Proceedings of the 7th International Conference on Advances in Computer-Human Interactions, ACHI 2014, Barcelona, Spain, 23–27 March 2014; pp. 200–207.

20. Al-Amin, M.; Qin, R.; Tao, W.; Doell, D.; Lingard, R.; Yin, Z.; Leu, M.C. Fusing and refining convolutional neural network models for assembly action recognition in smart manufacturing. *Proc. Inst. Mech. Eng. Part C J. Mech. Eng. Sci.* **2020**. [\[CrossRef\]](#)
21. Calvetti, D.; Mèda, P.; Gonçalves, M.C.; Sousa, H. Worker 4.0: The future of sensed construction sites. *Buildings* **2020**, *10*, 169. [\[CrossRef\]](#)
22. Rebmann, A.; Emrich, A.; Fettke, P. Enabling the discovery of manual processes using a multi-modal activity recognition approach. In *Business Process Management Workshops*; Lecture Notes in Business Information, Processing; Di Francescomarino, C., Dijkman, R., Zdun, U., Eds.; Springer: Cham, Switzerland, 2019; Volume 362. [\[CrossRef\]](#)
23. Knoch, S.; Ponpathirkootam, S.; Fettke, P.; Loos, P. Technology-enhanced process elicitation of worker activities in manufacturing. In *Business Process Management Workshops*; BPM 2017; Lecture Notes in Business Information Processing; Teniente, E., Weidlich, M., Eds.; Springer: Cham, Switzerland, 2017; Volume 308. [\[CrossRef\]](#)
24. Zou, H.; Zhou, Y.; Yang, J.; Spanos, C.J. Towards occupant activity driven smart buildings via WiFi-enabled IoT devices and deep learning. *Energy Build.* **2018**, *177*, 12–22. [\[CrossRef\]](#)
25. Ijjina, E.P.; Chalavadi, K.M. Human action recognition in RGB-D videos using motion sequence information and deep learning. *Pattern Recogn.* **2017**, *72*, 504–516. [\[CrossRef\]](#)
26. Chen, Z.; Jiang, C.; Xie, L. Building occupancy estimation and detection: A review. *Energy Build.* **2018**, *169*, 260–270. [\[CrossRef\]](#)
27. Mannhardt, F.; Bovo, R.; Oliveira, M.F.; Julier, S. A taxonomy for combining activity recognition and process discovery in industrial environments. In *Intelligent Data Engineering and Automated Learning—IDEAL 2018*; IDEAL 2018; Lecture Notes in Computer, Science; Yin, H., Camacho, D., Novais, P., Tallón-Ballesteros, A., Eds.; Springer: Cham, Switzerland, 2018; Volume 11315. [\[CrossRef\]](#)
28. Tadeusiewicz, R. *Image Recognition*; PWN: Warszawa, Poland, 1991.
29. Ge, H.; Zhu, Z.; Lou, K.; Wei, W.; Liu, R.; Damaševičius, R.; Woźniak, M. Classification of infrared objects in manifold space using kullback-leibler divergence of gaussian distributions of image points. *Symmetry* **2020**, *12*, 434. [\[CrossRef\]](#)
30. Zhou, B.; Duan, X.; Ye, D.; Wei, W.; Woźniak, M.; Połap, D.; Damaševičius, R. Multi-level features extraction for discontinuous target tracking in remote sensing image monitoring. *Sensors* **2019**, *19*, 4855. [\[CrossRef\]](#) [\[PubMed\]](#)
31. Goodfellow, I.; Bengio, Y.; Courville, A.; Bengio, Y. *Deep Learning*; MIT Press: Cambridge, UK, 2016; Volume 1, No. 2.
32. Hassan, M.M.; Uddin, M.Z.; Mohamed, A.; Almogren, A. A robust human activity recognition system using smartphone sensors and deep learning. *Future Gener. Comput. Syst.* **2018**, *81*, 307–313. [\[CrossRef\]](#)
33. Tao, W.; Lai, Z.H.; Leu, M.C.; Yin, Z. Worker activity recognition in smart manufacturing using IMU and sEMG signals with convolutional neural networks. *Procedia Manuf.* **2018**, *26*, 1159–1166. [\[CrossRef\]](#)
34. Zhang, F.; Niu, K.; Xiong, J.; Jin, B.; Gu, T.; Jiang, Y.; Zhang, D. Towards a diffraction-based sensing approach on human activity recognition. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* **2019**, *3*, 1–25. [\[CrossRef\]](#)
35. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [\[CrossRef\]](#) [\[PubMed\]](#)
36. Zheng, X.; Wang, M.; Ordieres-Meré, J. Comparison of data preprocessing approaches for applying deep learning to human activity recognition in the context of industry 4.0. *Sensors* **2018**, *18*, 2146. [\[CrossRef\]](#) [\[PubMed\]](#)
37. Tao, W.; Leu, M.C.; Yin, Z. Multi-modal recognition of worker activity for human-centered intelligent manufacturing. *Eng. Appl. Artif. Intell.* **2020**, *95*, 103868. [\[CrossRef\]](#)
38. Ryselis, K.; Petkus, T.; Blažauskas, T.; Maskeliūnas, R.; Damaševičius, R. Multiple kinect based system to monitor and analyze key performance indicators of physical training. *Hum.-Cent. Comput. Inf. Sci.* **2020**, *10*, 51. [\[CrossRef\]](#)
39. Rude, D.J.; Adams, S.; Beling, P.A. Task recognition from joint tracking data in an operational manufacturing cell. *J. Intell. Manuf.* **2018**, *29*, 1203–1217. [\[CrossRef\]](#)
40. Kulikajėvas, A.; Maskeliūnas, R.; Damaševičius, R. Detection of sitting posture using hierarchical image composition and deep learning. *PeerJ Comput. Sci.* **2021**, *7*, e442. [\[CrossRef\]](#)
41. Zhang, M.; Chen, S.; Zhao, X.; Yang, Z. Research on construction workers' activity recognition based on smartphone. *Sensors* **2018**, *18*, 2667. [\[CrossRef\]](#)
42. Xia, Q.; Korpela, J.; Namioka, Y.; Maekawa, T. Robust unsupervised factory activity recognition with body-worn accelerometer using temporal structure of multiple sensor data motifs. *Proc. ACM Interact. Mobile Wearable Ubiquitous Technol.* **2020**, *4*, 1–30. [\[CrossRef\]](#)
43. Menolotto, M.; Komaris, D.; Tedesco, S.; O'flynn, B.; Walsh, M. Motion capture technology in industrial applications: A systematic review. *Sensors* **2020**, *20*, 5687. [\[CrossRef\]](#)
44. Yang, J.; Liu, Y.; Liu, Z.; Wu, Y.; Li, T.; Yang, Y. A Framework for Human Activity Recognition Based on WiFi CSI Signal Enhancement. *Int. J. Antennas Propag.* **2021**, *6654752*, 1–18. [\[CrossRef\]](#)
45. Afza, F.; Khan, M.A.; Sharif, M.; Kadry, S.; Manogaran, G.; Saba, T.; Ashraf, I.; Damaševičius, R. A framework of human action recognition using length control features fusion and weighted entropy-variances based feature selection. *Image Vis. Comput.* **2021**, *106*, 104090. [\[CrossRef\]](#)
46. Helmi, A.M.; Al-qaness, M.A.A.; Dahou, A.; Damaševičius, R.; Krilavičius, T.; Elaziz, M.A. A Novel Hybrid Gradient-Based Optimizer and Grey Wolf Optimizer Feature Selection Method for Human Activity Recognition Using Smartphone Sensors. *Entropy* **2021**, *23*, 1065. [\[CrossRef\]](#)
47. Priya, S.J.; Rani, A.J.; Subathra, M.S.P.; Mohammed, M.A.; Damaševičius, R.; Ubendran, N. Local Pattern Transformation Based Feature Extraction for Recognition of Parkinson's Disease Based on Gait Signals. *Diagnostics* **2021**, *11*, 1395. [\[CrossRef\]](#)

48. Wozniak, M.; Wieczorek, M.; Silka, J.; Polap, D. Body pose prediction based on motion sensor data and recurrent neural network. *IEEE Trans. Ind. Inform.* **2021**, *17*, 2101–2111. [\[CrossRef\]](#)
49. Li, M.; Jiang, Z.; Liu, Y.; Chen, S.; Wozniak, M.; Scherer, R.; Damasevicius, R.; Wei, W.; Li, Z.; Li, Z. Sitsen: Passive sitting posture sensing based on wireless devices. *Int. J. Distrib. Sens. Netw.* **2021**, *17*, 15501477211024846. [\[CrossRef\]](#)
50. Mujahid, A.; Awan, M.J.; Yasin, A.; Mohammed, M.A.; Damaševičius, R.; Maskeliūnas, R.; Abdulkareem, K.H. Real-Time Hand Gesture Recognition Based on Deep Learning YOLOv3 Model. *Appl. Sci.* **2021**, *11*, 4164. [\[CrossRef\]](#)
51. Maskeliūnas, R.; Damaševičius, R.; Segal, S. A review of internet of things technologies for ambient assisted living environments. *Future Internet* **2019**, *11*, 259. [\[CrossRef\]](#)
52. Dammalapati, H.; Swamy Das, M. An efficient criminal segregation technique using computer vision. In Proceedings of the IEEE 2021 International Conference on Computing, Communication, and Intelligent Systems, ICCIS 2021, Greater Noida, India, 19–20 February 2021; pp. 636–641. [\[CrossRef\]](#)
53. Raudonis, V.; Maskeliūnas, R.; Stankevičius, K.; Damaševičius, R. Gender, Age, Colour, Position and Stress: How They Influence Attention at Workplace. In *Computational Science and Its Applications—ICCSA 2017*; ICCSA 2017; Lecture Notes in Computer Science; Gervasi, O., Murgante, B., Misra, S., Borruso, G., Torre, C.M., Rocha, A.M.A.C., Apduhan, B.O., Stankova, E., Cuzzocrea, A., Eds.; Springer: Cham, Switzerland, 2017; Volume 10408. [\[CrossRef\]](#)
54. Akhavian, R.; Behzadan, A.H. Smartphone-based construction workers' activity recognition and classification. *Autom. Constr.* **2016**, *71 Pt 2*, 198–209. [\[CrossRef\]](#)
55. Al Jassmi, H.; Al Ahmad, M.; Ahmed, S. Automatic recognition of labor activity: A machine learning approach to capture activity physiological patterns using wearable sensors. *Constr. Innov.* **2021**, *2*, 555–575. [\[CrossRef\]](#)
56. Yu, Y.; Li, H.; Cao, J.; Luo, X. Three-dimensional working pose estimation in industrial scenarios with monocular camera. *IEEE Internet Things J.* **2021**, *8*, 1740–1748. [\[CrossRef\]](#)
57. Sherafat, B.; Ahn, C.R.; Akhavian, R.; Behzadan, A.H.; Golparvar-Fard, M.; Kim, H.; Azar, E.R. Automated methods for activity recognition of construction workers and equipment: State-of-the-art review. *J. Constr. Eng. Manag.* **2020**, *146*, 03120002. [\[CrossRef\]](#)
58. Angah, O.; Chen, A.Y. Tracking multiple construction workers through deep learning and the gradient based method with re-matching based on multi-object tracking accuracy. *Autom. Constr.* **2020**, *119*, 103308. [\[CrossRef\]](#)
59. Hu, H.; Cheng, K.; Li, Z.; Chen, J.; Hu, H. Workflow recognition with structured two-stream convolutional networks. *Pattern Recognit. Lett.* **2020**, *130*, 267–274. [\[CrossRef\]](#)
60. Ding, L.; Fang, W.; Luo, H.; Love, P.E.; Zhong, B.; Ouyang, X. A deep hybrid learning model to detect unsafe behavior: Integrating convolution neural networks and long short-term memory. *Autom. Constr.* **2018**, *86*, 118–124. [\[CrossRef\]](#)
61. Zhao, J.; Obonyo, E. Convolutional long short-term memory model for recognizing construction workers' postures from wearable inertial measurement units. *Adv. Eng. Inform.* **2020**, *46*, 101177. [\[CrossRef\]](#)
62. Yang, K.; Ahn, C.R.; Kim, H. Deep learning-based classification of work-related physical load levels in construction. *Adv. Eng. Inform.* **2020**, *45*, 101104. [\[CrossRef\]](#)
63. Sakalle, A.; Tomar, P.; Bhardwaj, H.; Acharya, D.; Bhardwaj, A. A LSTM based deep learning network for recognizing emotions using wireless brainwave driven system. *Expert Syst. Appl.* **2021**, *173*, 114516. [\[CrossRef\]](#)
64. Gong, F.; Ma, Y.; Zheng, P.; Song, T. A deep model method for recognizing activities of workers on offshore drilling platform by multistage convolutional pose machine. *J. Loss Prev. Process Ind.* **2020**, *64*, 104043. [\[CrossRef\]](#)
65. Zeng, M.; Nguyen, L.T.; Yu, B.; Mengshoel, O.J.; Zhu, J.; Wu, P.; Zhang, J.Y. Convolutional Neural Networks for Human Activity Recognition using Mobile Sensors. In Proceedings of the 6th International Conference on Mobile Computing, Applications and Services, Austin, TX, USA, 6–7 November 2014; pp. 197–205.
66. Jaouedi, N.; Boujnah, N.; Bouhlel, M.S. Deep Learning Approach for Human Action Recognition Using Gated Recurrent Unit Neural Networks and Motion Analysis. *J. Comput. Sci.* **2019**, *15*, 1040–1049. [\[CrossRef\]](#)
67. Pohlt, C.; Schlegl, T.; Wachsmuth, S. Human work activity recognition for working cells in industrial production contexts. In Proceedings of the IEEE International Conference on Systems, Man and Cybernetics, Bari, Italy, 6–9 October 2019; pp. 4225–4230. [\[CrossRef\]](#)
68. Tao, W.; Al-Amin, M.; Chen, H.; Leu, M.C.; Yin, Z.; Qin, R. Real-time assembly operation recognition with fog computing and transfer learning for human-centered intelligent manufacturing. *Procedia Manuf.* **2020**, *48*, 926–931. [\[CrossRef\]](#)
69. Son, H.; Choi, H.; Seong, H.; Kim, C. Detection of construction workers under varying poses and changing background in image sequences via very deep residual networks. *Autom. Constr.* **2019**, *99*, 27–38. [\[CrossRef\]](#)
70. Sun, H.; Ning, G.; Zhao, Z.; Huang, Z.; He, Z. Automated work efficiency analysis for smart manufacturing using human pose tracking and temporal action localization. *J. Vis. Commun. Image Represent.* **2020**, *73*, 102948. [\[CrossRef\]](#)
71. Maliszewska, P.; Halikowski, D.; Patalas-Maliszewska, J. A Model for Generating Workplace Procedures Using a CNN-SVM Architecture. *Symmetry* **2019**, *11*, 1151. [\[CrossRef\]](#)
72. Wogu, I.A.P.; Misra, S.; Assibong, P.A.; Olu-Owolabi, E.F.; Maskeliūnas, R.; Damasevicius, R. Artificial intelligence, smart classrooms and online education in the 21st century: Implications for human development. *J. Cases Inf. Technol.* **2019**, *21*, 66–79. [\[CrossRef\]](#)
73. Roberts, D.; Torres Calderon, W.; Tang, S.; Golparvar-Fard, M. Vision-based construction worker activity analysis informed by body posture. *J. Comput. Civ. Eng.* **2020**, *34*, 04020017. [\[CrossRef\]](#)

-
74. Neuhausen, M.; Pawlowski, D.; König, M. Comparing classical and modern machine learning techniques for monitoring pedestrian workers in top-view construction site video sequences. *Appl. Sci.* **2020**, *10*, 8466. [[CrossRef](#)]
 75. Sharma, R.; Pachori, R.B.; Sircar, P. Automated emotion recognition based on higher order statistics and deep learning algorithm. *Biomed. Signal Process. Control* **2020**, *58*, 101867. [[CrossRef](#)]
 76. Sathyanarayana, A.; Joty, S.; Fernandez-Luque, L.; Ofli, F.; Srivastava, J.; Elmagarmid, A.; Taheri, S.; Arora, T. *Impact of Physical Activity on Sleep: A Deep Learning Based Exploration*; Cornell University: Ithaca, NY, USA, 2016.