Article

Improved Long Short-Term Memory Network with Multi-Attention for Human Action Flow Evaluation in Workshop

Yun Yang ¹, Jiacheng Wang ^{1,*}, Tianyuan Liu ¹, Xiaolei Lv ² and Jinsong Bao ¹

¹ College of Mechanical Engineering, Donghua University, Shanghai 201620, China; yun@dhu.edu.cn (Y.Y.); tylu@mail.dhu.edu.cn (T.L.); bao@dhu.edu.cn (J.B.)

² Shanghai Space Propulsion Technology Research Institute, Shanghai 201620, China; 2180703@mail.dhu.edu.cn

* Correspondence: jiage@mail.dhu.edu.cn

Received: 19 September 2020; Accepted: 2 November 2020; Published: 5 November 2020

Featured Application: Our method can be used for worker action recognition and action flow evaluation in workshops, which can improve production standardization.

Abstract: As an indispensable part of workshops, the normalization of workers' manufacturing processes is an important factor that affects product quality. How to effectively supervise the manufacturing process of workers has always been a difficult problem in intelligent manufacturing. This paper proposes a method for action detection and process evaluation of workers based on a deep learning model. In this method, the human skeleton and workpiece features are separately obtained by the monitoring frame and then input into an action detection network in chronological order. The model uses two inputs to predict frame-by-frame classification results, which are then merged into a continuous action flow, and finally, input into the action flow evaluation network. The network effectively improves the ability to evaluate action flow through the attention mechanism of key actions in the process. The experimental results show that our method can effectively recognize operation actions in workshops, and can evaluate the manufacturing process with 99% accuracy using the experimental verification dataset.

Keywords: intelligent monitoring; human factors; action recognition; long short-term memory network; attention mechanism

1. Introduction

In manufacturing, the key to quality control lies in the manufacturing process itself; thus, monitoring and control of the manufacturing process is the focus of manufacturing quality control [1]. The product quality is the result of multiple working steps which are the basic units of the manufacturing process [2,3]. The mapping of workers, that is, of the process flow of workers' processing actions of the workpiece, consists of multiple separate processing actions [4]. Identifying workers' processing actions and processing action flows can contribute to effectively judging whether the processing status of the workpiece and the processing flow are disordered, making it possible to control the quality of the final product [5]. Monitoring and controlling the quality of processes has become the most important task of quality control. Manual observation is the main method of traditional monitoring, which is time-consuming and labor-intensive. It is increasingly difficult to meet production needs using this method. The emergence and continuous development of intelligent monitoring based on deep learning provides an effective way to monitor processing actions, and is also an important enabling

technology for the future transformation and upgrading of manufacturing and management to digital and intelligent modes [6,7].

As the key technology of intelligent monitoring, action recognition has received significant attention in recent years [8–10]. Its research routes can be divided into two categories: image-based and human skeleton-based recognition methods.

- Image-based recognition methods

Wang et al., analyzed the changes of pixels in video sequences and used the dense trajectories (DT) [11] and improved dense trajectories (iDT) methods [12] to identify actions. Wang et al. proposed a method using CNN (Convolutional Neural Networks) to classify trajectories [13]. Song et al., used 3D convolution to process temporal and spatial features at the same time [14]. Tran used two-stream CNN, with one stream extracting spatial information from single-frame optical flow and another extracting temporal information from multiframe optical flow [15]. Image-based methods can extract more features from the whole frame, but due to the excessive attention to the background, light and other information of the image, much information is redundant and the method is less efficient and accurate.

- Human skeleton-based recognition method

Ke et al., used the distance between human joints to generate grayscale images and then employed CNN for classification [16]. Gaglio used 3D human posture data and three different machine learning techniques to recognize human activity [17]. Wei et al. considered high order features like relative motion to different joints and proposed a novel high order joint relative motion feature and a novel human skeleton tree RNN network [18]; Yan et al. took the skeleton as a graph and used a graph neural network for classification [19]. The accuracy of the methods based on skeleton joint information is generally greater than methods based on image recognition, but these methods only consider the movement of the human body, which is too simple to further improve the accuracy of human action recognition.

Current research is still overly focused on the recognition of a single action rather than of continuous action processes. However, in a complex workshop production scenario, an analysis of a single action by a worker cannot meet current quality control needs. In many production processes, the continuous flow of workers' actions also needs to follow strict yet flexible process standards [20–22]. Therefore, it is equally important to analyze and judge the normalization of the action flow. On the other hand, the task of action recognition in a workshop is different from the same task in a natural scenario, because workers' actions mostly involve the use of tools to operate the workpiece or the equipment. Information such as that relating to the workpiece has a vital influence on action recognition. Taking account of this situation, this paper proposes a workpiece attention-based, long short-term memory (WA-LSTM) network for action detect and a key action attention-based, long short-term memory (KAA-LSTM) network for action flow evaluation. The overall framework of the method is shown in Figure 1. The experiment results show that our methods can recognize workers' actions online with high accuracy and can effectively judge whether workers' action flows follow the requirements.

Figure 1. The overall framework of our method. An action detect network is used to acquire segmented actions and insert them into the process evaluation network for standardized discrimination.

2. Methods

There are two requirements for the recognition of workers' manufacturing actions. Firstly, it is necessary to detect the individual manufacturing action of a worker in the monitoring video. We detect the action classification results frame-by-frame using the action detection network, and segment continuous actions by filtering and combining. Secondly, it is also necessary to evaluate the process level of the identified action flow to determine whether the worker's action flow meets the specifications. The recognized actions are sequentially input in chronological order and the action flow is evaluated in combination with all actions.

2.1. Manufacture Action Recognition Based on WA-LSTM

For the recognition of a single action, methods such as [10–12,17,18] need to divide the video segment in advance to obtain action fragments, which cannot meet the online recognition requirements of surveillance video. Temporal neural networks such as recurrent neural networks (RNNs) [23] and long short-term memory networks (LSTMs) [24] can effectively retain temporal information and make frame-level action classifications to meet the online needs of action recognition [25,26]. However, traditional human skeleton-based recognition methods ignore the presence of the workpiece and other factors which are critical in the recognition of manufacturing actions in workshop. Thus, we propose a WA-LSTM method for action detection and recognition. The overall framework of WA-LSTM is shown in Figure 2.

Figure 2. The overall framework of WA-LSTM. The method takes continuous frames of monitoring video as input and consists of two branches. One branch predicts the skeleton position of the worker and inserts it into the LSTM, while the other extracts the workpiece feature to be used as an attention mechanism to boost the LSTM.

2.1.1. Encoding of Worker Skeleton Sequence

A worker's action is composed of a series of static skeletons in temporal sequence. In this paper, the static skeletons of workers in each frame are defined as the action features of workers. The skeleton features obtained by different methods or formats are different. For example, the Kinect depth sensor captures 3D coordinates of 25 joints of the human body [27], while the OpenPose framework captures the 2D coordinates of 18 joints [28]. We use OpenPose for the experiment. The skeleton feature is defined as

$$P = \{(x_1, y_1), (x_2, y_2), \dots, (x_{18}, y_{18})\} \quad (1)$$

where (x_i, y_i) represents the cartesian coordinates of the i th joint. Thus, the skeleton feature sequence in the manufacturing process can be defined as

$$S = [P^1, P^2, \dots, P^m] \quad (2)$$

where m represents the m th frame in the manufacturing process.

2.1.2. Feature Extraction of Workpiece and Fusion Method

Workers' actions are different from actions in a natural scenario, i.e., they are more closely related to the workshop environment, such as the workpiece and tools. These environmental factors often determine the actions of workers. Therefore, it is better to recognize workers' actions using these factors.

In this paper, we use a pretrained convolutional neural network to extract workpiece features. As shown in Figure 3, we use a simple fully convolutional network (FCN) [29], a common framework in semantic segmentation, to train a model with the capacity to segment frequently-used workpieces in one image. Our dataset consists of 500 frames extracted from the video data of the dataset which will be introduced in Section 3. We labeled the frames for every workpiece or tool at a pixel level for the image segmentation task.

Figure 3. An image segmentation network was trained on our workpiece datasets in advance. We then used the downsampled part of the network and added an additional fully connected layer to extract workpiece features, which is a one-dimensional vector. The additional fully connected layer was trained jointly with LSTM.

When extracting features from images, only the downsampled parts with their parameters frozen are used. After that, a new, fully connected (FC) neural network is added to acquire high-level semantic features, which can be defined as

$$W = [w_1, w_2, \dots, w_n] \quad (3)$$

where n is the dimension of the features vector that can be adjusted dynamically by the number of output layer cells. The additional neural network is jointly trained with the LSTM.

2.1.3. Overall Workflow of WA-LSTM

For each frame in the video stream, the human skeleton features P^i , representing the human skeleton features at frame i , are extracted through the OpenPose framework.

After smooth processing with a Kalman filter, an algorithm for optimal time sequence estimation [30,31], the skeleton features are input into the LSTM network which maps the worker skeleton feature information to another high dimension for a high-level feature. Meanwhile, the LSTM can effectively preserve the temporal information for short-term temporal association learning. The output of LSTM is denoted as

$$C^i = [c_1^i, c_2^i, \dots, c_n^i] \quad (4)$$

representing the classification information of skeleton features at moment i , where n is the number of categories of action.

Meanwhile, the workpiece feature is extracted from the pretrained FCN with an additional FC layer marked as W_i to represent the workpiece semantic.

The output C^i of the LSTM network is fused with the workpiece semantic W_i by weighted summation. Softmax is used as the final activation function to generate the probability y_j^i of each action category j at time i . The formula is expressed as

$$y_j^i = \text{softmax}(W^i \cdot C^i) \quad (5)$$

The generated frame-level classification needs to be further filtered and combined to obtain the temporal action.

2.2. Manufacturing Process Evaluation Based on KAA-LSTM

The manufacturing process is composed of several manufacturing actions, a typical sequence model with obvious context information between actions. There is a different degree of importance between every action in the action sequence. Referring to the research on semantic focus [32,33] in the natural language process domain, we propose a KAA-LSTM which uses an attention mechanism to express different degrees of importance. The overall framework is shown in Figure 4.

Figure 4. The KAA-LSTM takes actions step-by-step and uses key action attention to merge all the hidden states to get a final scale which represents the normative score.

2.2.1. Encoding of Action Sequences

We use LSTM for process evaluation; this requires the input to have a certain dimension. Thus, we use the one-hot encoding method [34] to solve this problem. For the single action A_i , which means the i th action classification category, we use the vector

$$A_i = \{0, 0, \dots, 1, \dots, 0\} \quad (6)$$

to represent the action. The length of the vector is determined by the number of action categories. Its elements are all set to zero with only the corresponding position i being set to one. The action sequence can be represented by a one-hot encoding sequence as follows:

$$\begin{pmatrix} A_1^1 \\ A_2^1 \\ \vdots \\ A_2^t \end{pmatrix} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 1 & \cdots & 0 \end{pmatrix} \quad (7)$$

where t is the t th action in chronological order. The LSTM takes every one-hot vector as input in sequence. After the last input, the hidden layer states c_i of each input are weight-summed and the sigmoid function is used to generate the probability of whether the process is in conformity with the specification. The output of the function is between 0 and 1. The higher the output, the more standardized it is.

2.2.2. Key Action Attentional Mechanisms

In text classification, different words and sentences contain different amounts of information and the same words have different importance in different semantic contexts. The normative discrimination of the manufacturing action process has similar characteristics. The importance of each kind of manufacturing action in a particular process may be different, and the importance of the same kind of production action occurring many times may be different.

In view of the above characteristics, this paper proposes the key action attention mechanism [35] to extract the action information that is most critical to normative judgment in the process of production

action, and to assign the attention weight according to the importance of obtaining the most identifiable feature vector. The formula is expressed as follows:

$$u_i = \tanh(W_s c_i + b_s) \quad (8)$$

$$\alpha_i = \frac{\exp(u_i^\top u_s)}{\sum_i^n \exp(u_i^\top u_s)} \quad (9)$$

$$v = \sum_i \alpha_i u_i \quad (10)$$

where u_i is an implicit representation of the state information obtained through a simple FC layer, α_i is the importance of each action in the process calculated by the softmax function, and W_s , b_s and u_s are training parameters obtained by jointly training. Finally, u_i is weight-summed and put through the sigmoid function. The key action attention mechanism quantifies the importance of each action in the action process, which can assist the further identification of information for normative discrimination.

3. Results

3.1. Case Description and Dataset

At present, there is no dataset for manufacturing actions. This paper takes the precleaning procedure of the combustion chamber of a rocket engine as a test case. The propellant of rocket engines is a typical firework product with strict process specification requirements for every manufacturing action. The standard process is as follows:

- (a) Polish the surface, shaping it;
- (b) Blow on the surface and cool it;
- (c) Tap the internal thread to ensure that the thread form meets the requirements;
- (d) Clean with a brush to remove surface debris;
- (e) Move the propellant to the next working procedure.

However, in the actual manufacturing process, there will be reworking and repetition. The present study is based on an actual workers' precleaning processes gathered from interviews and simulated in a laboratory environment. We simulated 520 processes, including 300 standard processes and 220 nonstandard ones. Each process consists of several types of manufacturing actions, as shown in Figure 5. Our dataset was recorded using a camera with a resolution of 640×480 pixels and a frame rate of 15 FPS. Each video sample ranged in time from 10 to 18 s and included between four and ten actions. Standard samples were executed in accordance with the above standard procedure while nonstandard samples were not, for example those in which cleaning was executed before cooling.

The main difference between our dataset and other common datasets is that each action recorded in our dataset is an operation performed on a workpiece. Datasets like MSR [36] are simply records of the body's own behavior. Furthermore, each sample of our dataset consists of several continuous actions. Samples in datasets like UCF101 [37], on the other hand, are snippets of single actions. Finally, our dataset provides a label of evaluation for every sample to produce a "good or not" classification of sequential actions. To the best of our knowledge, no other such dataset exists which meets our requirements.

Figure 5. Sample of our dataset, including the original RGB frame and the skeleton sequence of every action.

3.2. Experiment and Result for WA-LSTM

The most notable feature of WA-LSTM is that it integrates the features of the workpiece and the skeleton of the worker, and analyzes their temporal information through the LSTM network, so that each feature can be used to the maximum advantage. In order to verify its superiority, the following models are used for comparative experiments:

- DNN: The worker skeleton features and the workpiece features are directly input into the FC network to obtain the classification results without considering the temporal information.
- LSTM: Only the skeleton features of the worker and the temporal information are considered, but not the workpiece information.

The cross-entropy is used as the loss function during training. For predicted result y^i with the real label $\hat{y}^i$, loss is defined as:

$$loss = \sum_{j=0}^n -y_j^i \log(\hat{y}_j^i) \quad (11)$$

where n indicates the number of categories. The iterative process of loss and accuracy during the training process is shown in Figure 6.

Figure 6. Comparison of experimental loss and accuracy between three groups.

3.3. Experiment and Result for KAA-LSTM

The manufacturing action process recognized by WA-LSTM is input into the KAA-LSTM model to discriminate standardization. To prove the performance of KAA-LSTM, we also set up two comparison experiments using the following models:

- a simple DNN model;
- an LSTM model without attentional mechanisms.

We set the true label of the standard process to $\hat{y} = 1$, and that of the nonstandard process to $\hat{y} = 0$. After activation by the sigmoid function, the model outputs a single value, y , as a normative evaluation score. Taking cross-entropy as the loss function, the loss value of a single sample is expressed as:

$$loss = -y \log(\hat{y}) - (1 - y) \log(1 - \hat{y}) \quad (12)$$

During the prediction process, an output result greater than 0.5 is set as the standard sample and the accuracy calculated based on this. The experimental results are shown in Table 1 [38].

Table 1. The results of three comparative experiments.

Index	DNN	LSTM	KAA-LSTM
Cross Entropy	0.1991	0.1843	0.1274
Accuracy	0.9478	0.9775	0.9936
Recall	0.8832	0.9023	0.9525
Precision	0.9227	0.9343	0.9710
F1-score	0.9025	0.9180	0.9617

Note. DNN = Deep Neural Network; LSTM = Long Short-Term Memory; KAA-LSTM = Key Action Attentional LSTM.

4. Discussion

4.1. Discussion on the Results of WA-LSTM

The convergence rate of the LSTM model is faster than that of the DNN model, and the convergence value is smaller, which indicates that temporal information has an important influence on action

recognition and can better promote convergence. From the iterative process of accuracy, it can be seen that the LSTM model can achieve 98% accuracy in the training set and more than 95% accuracy in the validation set, while the DNN model can only achieve roughly 87% accuracy in both sets. This is because the DNN model does not take temporal information into account, making it difficult to distinguish between ambiguously transitional actions. In addition, after adding the workpiece feature on the basis of LSTM, the converged value is further reduced and the accuracy is further improved, indicating that the workpiece feature as attention information can effectively improve the performance of the original model.

Stacking the truth label and the predicted label of each frame along the time line, we get the result shown in Figure 7. We can see that due to the lack of temporal information, the results of the DNN model are unstable, and there are many small segments, which will greatly affect the segment of action. As it takes account of the context information of the action, the LSTM model has better continuity in its results, and its performance is further improved after the addition of the workpiece attention mechanism. Its predicted results can better reflect the real values but have a small delay compared with the truth label. At the same time, it is easy to make recognition mistakes in the process of action cohesion when tools are changed. We consider the reason for this to be because all kinds of action have the same probability when tools are changed; the breakpoint of the action division when marking the truth value also causes interference in the experimental results. In conclusion, the proposed WA-LSTM model can significantly improve accuracy and make the prediction of results smoother.

Figure 7. The predicted results of each model and the truth label are stacked along the time axis, which can reflect the detection results of each model.

4.2. Discussion on the Results of KAA-LSTM

It can be seen from Table 1 that DNN and LSTM have lower accuracy, while KAA-LSTM can achieve almost 100% accuracy in the verification set. This is due to the information mining of important actions by the attention mechanism, which enhances the discrimination ability of the model. Furthermore, the recall and precision levels of KAA-LSTM are much higher than those of the other two models. This indicates that our model is capable of a high level of discernment.

4.3. Prospect

Our research is an attempt at standard analysis of worker action flow, which can be used for the normalization of work flow such as the assembly process in the workshop. Future research could be developed from the following points. First, the workpiece feature could be obtained in different ways, for example, object detection or knowledge graphs. Second, our dataset is a simulation and simplification of real manufacturing processes, which include more situations. A new dataset could be collected and analyzed in a real environment.

5. Conclusions

This paper discusses the application of monitoring video in the normative recognition of worker action flow. A sequence model based on workpiece attention is proposed to detect workers' manufacturing actions in the video. The experiment proves that the fusion of workpiece attention and human skeleton detection can effectively improve the accuracy and stability of action recognition. Another sequence model based on key action attention is proposed to evaluate the process of manufacturing action. Experiments show that this method can achieve 99.36% accuracy on the validation set, and can accurately identify some incorrect manufacturing processes. The method developed in this paper can be applied to monitor key stations in workshop processes and undertake the automatic supervision of workers' processing actions.

Author Contributions: Y.Y. conceived the idea; T.L. performed the static analyses; Y.Y. and J.W. designed the methodology; J.B. was responsible for project administration; X.L. provided resources; J.W. prepared the manuscript and checked the writing. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Fundamental Research Funds for the Central Universities (NO. 2232018D3-26).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Rude, D.J.; Adams, S.; Beling, P.A. Task Recognition from Joint Tracking Data in an Operational Manufacturing Cell. *J. Intell. Manuf.* **2015**, *29*, 1203–1217. [\[CrossRef\]](#)
2. Goecks, L.S.; Dos Santos, A.A.; Korzenowski, A.L. Decision-Making Trends in Quality Management: A Literature Review about Industry 4.0. *Producao* **2020**, *30*, 30. [\[CrossRef\]](#)
3. Tsao, L.; Li, L.; Ma, L. Human Work and Status Evaluation Based on Wearable Sensors in Human Factors and Ergonomics: A Review. *IEEE Trans. Hum. Mach. Syst.* **2018**, *49*, 72–84. [\[CrossRef\]](#)
4. Wang, D.; Kotake, Y.; Nakajima, H.; Mori, K.; Hata, Y. A Relationship between Product Quality and Body Information of Worker and Its Application to Improvement of Productivity. In Proceedings of the 2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Miyazaki, Japan, 7–10 October 2018; pp. 1433–1438.
5. Song, K.; Lee, S.; Shin, S.; Lee, H.J.; Han, C. Simulation-Based Optimization Methodology for Offshore Natural Gas Liquefaction Process Design. *Ind. Eng. Chem. Res.* **2014**, *53*, 5539–5544. [\[CrossRef\]](#)
6. Moustafa, N.; Adi, E.; Turnbull, B.; Hu, J. A New Threat Intelligence Scheme for Safeguarding Industry 4.0 Systems. *IEEE Access* **2018**, *6*, 32910–32924. [\[CrossRef\]](#)
7. Fernandez-Carames, T.M.; Fraga-Lamas, P. A Review on Human-Centered IoT-Connected Smart Labels for the Industry 4.0. *IEEE Access* **2018**, *6*, 25939–25957. [\[CrossRef\]](#)
8. Jobanputra, C.; Bavishi, J.; Doshi, N. Human Activity Recognition: A Survey. *Procedia Comput. Sci.* **2019**, *155*, 698–703. [\[CrossRef\]](#)
9. Kong, Y.; Fu, Y. Human Action Recognition and Prediction: A Survey. *arXiv* **2018**, arXiv:1806.11230.
10. Lan, Z.; Zhu, Y.; Hauptmann, A.G.; Newsam, S. Deep Local Video Feature for Action Recognition. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Honolulu, HI, USA, 21–26 July 2017; 2017, pp. 1219–1225.
11. Wang, H.; Kläser, A.; Schmid, C.; Liu, C.-L. Dense Trajectories and Motion Boundary Descriptors for Action Recognition. *Int. J. Comput. Vis.* **2013**, *103*, 60–79. [\[CrossRef\]](#)
12. Wang, H.; Schmid, C. Action Recognition with Improved Trajectories. In Proceedings of the 2013 IEEE International Conference on Computer Vision, Sydney, Australia, 1–8 December 2013; pp. 3551–3558.
13. Wang, L.; Qiao, Y.; Tang, X. Action Recognition with Trajectory-Pooled Deep-Convolutional Descriptors. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 4305–4314.
14. Song, J.; Yang, Z.; Zhang, Q.; Fang, T.; Hu, G.; Han, J.; Chen, C. Human Action Recognition with 3D Convolution Skip-Connections and RNNs. *Lect. Notes Comput. Sci.* **2018**, *11301*, 319–331. [\[CrossRef\]](#)

15. Tran, A.; Cheong, L.-F. Two-Stream Flow-Guided Convolutional Attention Networks for Action Recognition. In Proceedings of the 2017 IEEE International Conference on Computer Vision Workshops (ICCVW), Venice, Italy, 22–29 October 2017; pp. 3110–3119.
16. Ke, Q.; Bennamoun, M.; An, S.; Sohel, F.; Boussaid, F. A New Representation of Skeleton Sequences for 3D Action Recognition. In Proceedings of the 30th IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 4570–4579.
17. Gaglio, S.; Re, G.L.; Morana, M. Human Activity Recognition Process Using 3-D Posture Data. *IEEE Trans. Hum. Mach. Syst.* **2014**, *45*, 586–597. [[CrossRef](#)]
18. Wei, S.; Song, Y.; Zhang, Y. Human Skeleton Tree Recurrent Neural Network with Joint Relative Motion Feature for Skeleton Based Action Recognition. In Proceedings of the 2017 IEEE International Conference on Image Processing (ICIP), Beijing, China, 17–20 September 2017; pp. 91–95.
19. Li, Y.; He, Z.; Ye, X.; He, Z.; Han, K. Spatial Temporal Graph Convolutional Networks for Skeleton-Based Dynamic Hand Gesture Recognition. *EURASIP J. Image Video Process.* **2019**, *2019*, 1–7. [[CrossRef](#)]
20. Klochkov, Y.; Gazizulina, A.; Golovin, N.; Glushkova, A.; Zh, S. Information Model-Based Forecasting of Technological Process State. In Proceedings of the 2017 International Conference on Infocom Technologies and Unmanned Systems (Trends and Future Directions) (ICTUS), Dubai, UAE, 18–20 December 2017; pp. 709–712.
21. Cimini, C.; Pirola, F.; Pinto, R.; Cavalieri, S. A Human-in-the-Loop Manufacturing Control Architecture for the Next Generation of Production Systems. *J. Manuf. Syst.* **2020**, *54*, 258–271. [[CrossRef](#)]
22. Du, S.; Wu, P.; Wu, G.; Yao, C.; Zhang, L. The Collaborative System Workflow Management of Industrial Design Based on Hierarchical Colored Petri-Net. *IEEE Access* **2018**, *6*, 27383–27391. [[CrossRef](#)]
23. Zaremba, W.; Sutskever, I.; Vinyals, O. Recurrent Neural Network Regularization. *arXiv* **2014**, arXiv:1409.2329.
24. Gers, F.; Schmidhuber, E. LSTM Recurrent Networks Learn Simple Context-Free and Context-Sensitive Languages. *IEEE Trans. Neural Netw.* **2001**, *12*, 1333–1340. [[CrossRef](#)]
25. Li, Y.; Lan, C.; Xing, J.; Zeng, W.; Yuan, C.; Liu, J. Online Human Action Detection Using Joint Classification-Regression Recurrent Neural Networks. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 8–16 October 2016; pp. 203–220.
26. Liu, J.; Li, Y.; Song, S.; Xing, J.; Lan, C.; Zeng, W. Multi-Modality Multi-Task Recurrent Neural Network for Online Action Detection. *IEEE Trans. Circuits Syst. Video Technol.* **2018**, *29*, 2667–2682. [[CrossRef](#)]
27. Shotton, J.; FitzGibbon, A.; Cook, M.; Sharp, T.; Finocchio, M.; Moore, R.; Kipman, A.; Blake, A. Real-Time Human Pose Recognition in Parts from Single Depth Images. In Proceedings of the CVPR 2011, Colorado Springs, CO, USA, 20–25 June 2011; Volume 56, pp. 1297–1304.
28. Cao, Z.; Simon, T.; Wei, S.-E.; Sheikh, Y. Realtime Multi-person 2D Pose Estimation Using Part Affinity Fields. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 June 2017; pp. 1302–1310.
29. Shelhamer, E.; Long, J.; Darrell, T. Fully Convolutional Networks for Semantic Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *39*, 640–651. [[CrossRef](#)] [[PubMed](#)]
30. Welch, G.; Bishop, G. *An Introduction to the Kalman Filter*; University of North Carolina at Chapel Hill: Chapel Hill, NC, USA, 1995.
31. Zhao, S.; Shmaliy, Y.S.; Liu, F. Fast Kalman-Like Optimal Unbiased FIR Filtering with Applications. *IEEE Trans. Signal Process.* **2016**, *64*, 2284–2297. [[CrossRef](#)]
32. Sharma, S.; Kiros, R.; Salakhutdinov, R. Action Recognition using Visual Attention. *arXiv* **2015**, arXiv:1511.04119.
33. Bahdanau, D.; Cho, K.H.; Bengio, Y. Neural Machine Translation by Jointly Learning to Align and Translate. In Proceedings of the 3rd International Conference on Learning Representations, ICLR, San Diego, CA, USA, 7–9 May 2015.
34. Uriarte-Arcia, A.V.; López-Yáñez, I.; Yáñez-Márquez, C. One-Hot Vector Hybrid Associative Classifier for Medical Data Classification. *PLoS ONE* **2014**, *9*, e95715. [[CrossRef](#)] [[PubMed](#)]
35. Song, S.; Lan, C.; Xing, J.; Zeng, W.; Liu, J. An End-to-End Spatio-Temporal Attention Model for Human Action Recognition from Skeleton Data. In Proceedings of the 31st AAAI Conference on Artificial Intelligence, Marina del Rey, CA, USA, 1–3 June 2017; pp. 4263–4270.

36. Yuan, J.; Liu, Z.; Wu, Y. Discriminative Subvolume Search for Efficient Action Detection. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2009), Miami, FL, USA, 20–25 June 2009; pp. 22–24.
37. Soomro, K.; Zamir, A.R.; Shah, M. UCF101: A Dataset of 101 Human Actions Classes from Videos in The Wild. *arXiv* **2012**, arXiv:1212.0402.
38. Sáiz-Manzanares, M.C.; Escolar-Llamazares, M.-C.; Arnaiz-González, Á. Effectiveness of Blended Learning in Nursing Education. *Int. J. Environ. Res. Public Health* **2020**, *17*, 1589. [[CrossRef](#)]

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

© 2020 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 (<http://creativecommons.org/licenses/by/4.0/>).