

Article

Identification of Abnormal Processes with Spatial-Temporal Data Using Convolutional Neural Networks

Yumin Liu ¹, Zheyun Zhao ^{1,*} , Shuai Zhang ¹ and Uk Jung ² 

¹ Business School, Zhengzhou University, Zhengzhou 450001, China; yuminliu@zzu.edu.cn (Y.L.); zhang1227zzu@163.com (S.Z.)

² Department of Management, School of Business, Dongguk University-Seoul, Seoul 04620, Korea; ukjung@dongguk.edu

* Correspondence: zzu_zzy@126.com

Received: 2 December 2019; Accepted: 2 January 2020; Published: 6 January 2020



Abstract: Identifying abnormal process operation with spatial-temporal data remains an important and challenging work in many practical situations. Although spatial-temporal data identification has been extensively studied in some domains, such as public health, geological condition, and environment pollution, the challenge associated with designing accurate and convenient recognition schemes is very rarely addressed in modern manufacturing processes. This paper proposes a general recognition framework for identifying abnormal process with spatial-temporal data by employing a convolutional neural network (CNN) model. Firstly, motivated by the pasting case study, the spatial-temporal data are transformed into process images for capturing spatial and temporal interrelationship. Then, the CNN recognition model is presented for identifying different types of these process images, leading to the identification of abnormal process with spatial-temporal data. The specific architecture parameters of CNN are determined step by step. According to the performance comparison with alternative methods, the proposed method is able to accurately identify the abnormal process with spatial-temporal data.

Keywords: spatial-temporal data; pasting process; process image; convolutional neural network

1. Introduction

Advanced sensing technologies are being increasingly applied in data collection systems for the areas including public health, geological condition, environment pollution, and manufacturing process. If the output of sensors is represented by the data with space and time structure, it can be termed as spatial-temporal data [1]. A lot of research focuses on the abnormality identification of abnormal-spatial temporal data, such as identifying outliers of the hourly air quality [2], detecting abnormal ozone measurements caused by air pollution or correlation among neighbor sensors [3], and diagnosing whether a disease is randomly distributed over space and time [4]. With the development of manufacturing technology, many sensors have been installed in the production lines, and a large number of spatial-temporal data can be collected from such processes. In order to improve the quality of manufacturing process, the abnormality identification of such spatial-temporal data has attracted much attention. Wang et al. [5] proposed a spatial-temporal data modeling method to identify the abnormality of a wafer production process. The identification scheme developed by Megahed et al. [6] can quickly detect the emergence of a fault in the nonwoven textile production process. Yu et al. [1] presented a rapid spatial-temporal quality control procedure for detecting systematic and random outliers. Current research is being conducted on identifying whether the process with spatial-temporal data is normal or not. Their common objective is to accurately detect the time and location of changes in

the occurrence rate as soon as possible [7]. In other words, existing research only focuses on the process monitoring of spatial-temporal data, but it is very rare to consider the root cause of abnormal data. In a real process, however, both normal and abnormal data can be collected from the process, and process monitoring and fault diagnosis can be applied simultaneously with spatial-temporal data. Normal and abnormal data display different variation patterns, which can be observed in process. Hence, how to identify such patterns precisely is the key problem in modern manufacturing process control.

Generally speaking, spatial-temporal data collected from the production process has many observations over time and location, and the adjacent observations are highly correlated. Thus, the high volume and their correlation of spatial-temporal data provide a considerable challenge for the identification of abnormal process. Moreover, the curse of dimensionality and complex data structure makes it difficult to build an identification model. Therefore, dimension reduction techniques are required before using the identification model. To capture intrinsic spatial- and temporal- correlations in an abnormal process state, principal component analysis (PCA), a widely used technique in dimension reduction, can be applied to extract features from spatial-temporal data via unfolding the original data set [8]. However, PCA cannot be directly applied to two or higher dimensional tensor data, unless such data are reshaped into a vector. Because the vectorization operation breaks the spatial and temporal correlation structure, it would lose potentially useful information contained in original data [9]. It is known that analyzing spatial-temporal data is more challenging than analyzing one-dimensional data. To overcome this issue, multilinear PCA (MPCA) and uncorrelated multilinear PCA (UMPCA) are proposed as alternatives to PCA [10]. In these methods, the tensor structure of spatial-temporal data are considered and more effective representation can be extracted [11,12]. Although these methods have desirable performance in processing spatial-temporal data, they focus separately on how to extract useful features from raw data and construct an effective identification model of an abnormal process [13]. If extracted features cannot interpret abnormal processes sufficiently or the identification model does not understand the extracted features, the performance is not robust. Hence, how to propose an effective approach integrating features self-learning and the identification of abnormal process with spatial-temporal data are still a challenge to be overcome.

Convolutional neural networks (CNN), one of the most effective deep learning models for tensor data processing, has been widely applied in natural language processing [14], image recognition [15], electrocardiogram (ECG) analysis [16], and fault diagnosis [17]. Benefiting from the mechanism of CNN in tensor data processing, the correlation structure of spatial-temporal data can be well-preserved by CNN. Meanwhile, CNN does not need to extract abnormal features manually, as the features can be learned from spatial-temporal data hierarchically and automatically. Taking the advantages of CNN, a novel method for identifying abnormal production process with spatial-temporal data is proposed in this paper. The case study considered is a pasting process, which is a critical process in lead-acid battery production and the output of sensors constitute a typical example of spatial-temporal data. Motivated by this process, a CNN-based identification approach for abnormal process with spatial-temporal data are presented. In order to show the recognition accuracy and effect of this approach, UMPCA is used as a benchmark in our study.

This paper is organized as follows. In Section 2, the pasting process as a motivating example is introduced and its spatial-temporal data are acquired. Section 3 develops a general CNN framework for identifying abnormal process with spatial-temporal data. We investigate the validation of the CNN recognition model in Section 4. In Section 5, the CNN method is applied to online identify the abnormal pasting process, and then the performance of the proposed method is evaluated. Suggestions and directions for further research are discussed in conclusions.

2. Case Study: Pasting Process

2.1. Spatial-Temporal Data Acquisition

A lead-acid battery consists of basic cell blocks, and each cell block contains several plates. Plates are the basic components of lead-acid batteries, and unqualified plates will directly affect the initial capacity and cycle life of batteries. In general, plate production includes five processes: ball-milling, paste mixing, grid casting, pasting, and plate curing. Pasting is a critical process in plate production [18,19], and most components of poor quality batteries can be traced back to this process. The identification of abnormal pasting process is the key to ensure the quality of batteries. In the pasting process, the lead oxide pastes are squeezed into the gap between two sides of the grid, and then it is turned into a plate. The mechanism of the whole pasting process is shown in Figure 1. The uniformity of the lead oxide paste in the plate surface is a critical quality characteristic, which can be measured by plate thickness. Therefore, the change of uniformity will directly reflect the abnormal state of the pasting process.

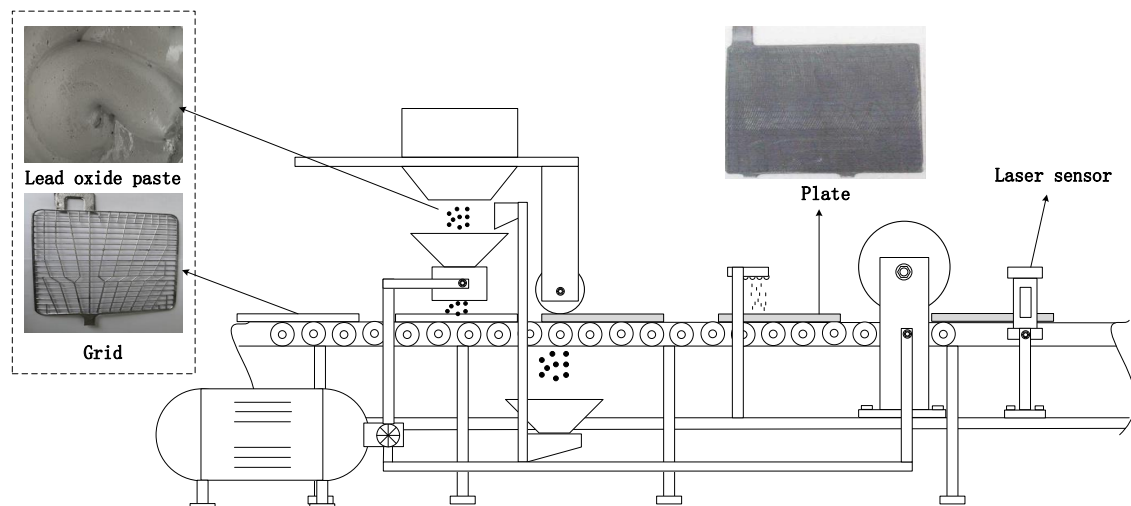


Figure 1. Pasting process.

To obtain plate thickness data, a laser sensor is installed at the end of pasting equipment to collect the observations of the uniformity as shown in Figure 1. When a plate moves in the pasting process, the laser sensor records its thickness values at different locations over time as shown in Figure 2. In the pasting example, there are m locations to measure the uniformity of the plate. When a plate moves through the sensor, the data observed over m locations can be collected at one time. In other words, the uniformity of the plate can be described by the observations measured in the m locations, and the uniformity of different plates can be observed over time to indicate the condition of the current process. Actually, the observation of uniformity is represented as a vector, which can be collected at time t . The vector will become a matrix over time, which can indicate the stability of the pasting process. The matrix collected from the pasting process is bi-dimensional, including space and time dimension. The matrix visualized by a surface is shown in Figure 2.

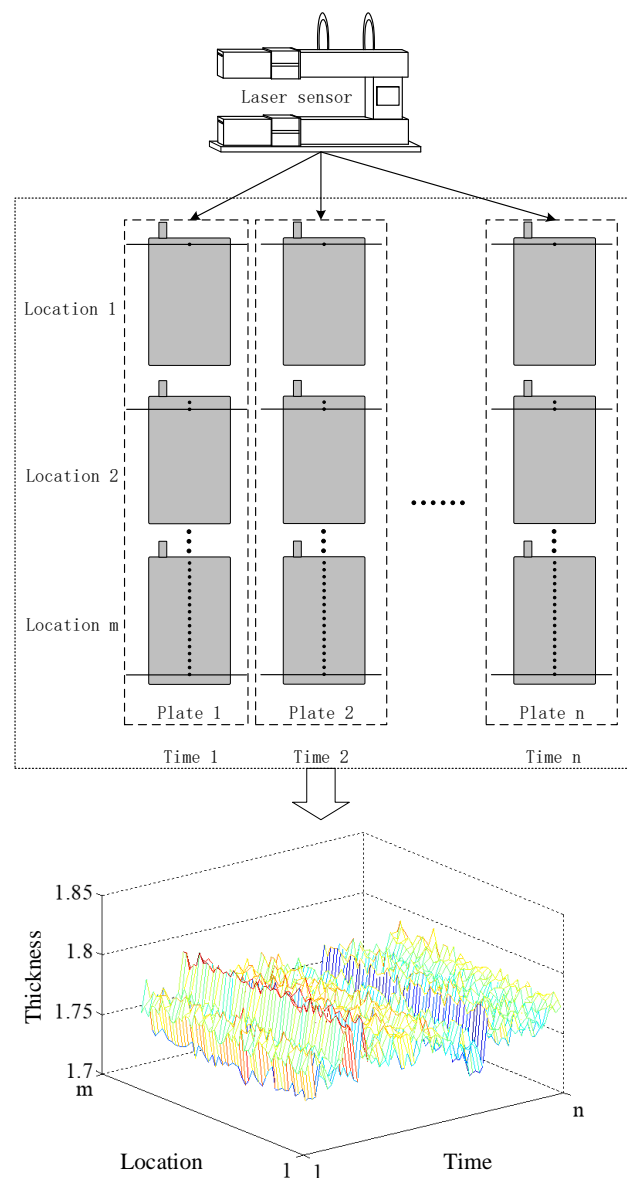


Figure 2. Spatial-temporal data collection in the pasting process.

The abnormal changes of plate thickness in the pasting process often result from unexpected causes, such as the failure of the pasting machine and unqualified grids, which are the root causes of the abnormal process. Once these root causes are identified and removed, the pasting process will return to normal. The plate thickness data change randomly in space and time when the pasting process is running normally, which is referred to as a normal process pattern, F_0 , seen in Figure 3a. In general, different causes will lead to various abnormal process patterns, which can be reflected by the changes of plate thickness in space and time domains. For example, when the plate thicknesses have an upward shift, it is usually caused by the wear of the parts in the pasting machine.

According to the changes of spatial-temporal data from pasting process and engineering experience, we find out seven common abnormal process patterns. When the uniformity of lead oxide paste becomes worse, the plate thickness data will change unevenly in time and space, which is denoted by abnormal process pattern F_1 , seen in Figure 3b. This abnormal process pattern results from the failure of the compression roller or acid spouting system, such as the blockage of the sprayer in the acid spouting system and aging spring in a compression roller. When the plate thickness is not uniform on both sides, that is, one side is thick and another is thin, the spatial-temporal data of

plate thickness at different locations will gradually become steeper with time. This abnormal process pattern is labeled as F_2 , and its corresponding causes are unusual clearance between the pasting machine and conveyor belt. When the plate thickness becomes thicker gradually, there is a steady rise over time in the spatial-temporal data, seen in Figure 3d. This pattern, denoted by abnormal process pattern F_3 , is caused by the insufficiency of conveyor belt tension under the pasting machine. When the thickness uniformity of plates becomes worse in a sudden way, the spatial-temporal data will become nonuniform suddenly, which is denoted by abnormal process pattern F_4 , as seen in Figure 3e. In general, this situation could be attributed to the low strength of the steel in a new batch of grids. When the thickness between the two sides of plates becomes nonuniform suddenly, the spatial-temporal data of plate thickness at one side will step up and become steep suddenly, which can be seen in Figure 3f. This situation is denoted by abnormal process pattern F_5 and its corresponding cause is usually that the roller in the pasting machine slants to one side. When the plate thickness becomes thicker suddenly, the overall upward step will be shown in the spatial-temporal data of plate thickness, seen in Figure 3d. The abnormal process pattern F_6 could be used as the label of this situation, which results from the electromagnetic fault in the pressure machine of compressed air. When the thickness of plates become thicker in a periodic manner, a periodical change will be observed in spatial-temporal data, seen in Figure 3h. This type of situation is labeled as abnormal process pattern F_7 , and the corrosion status of pasting conveyor belt is usually required to be checked by engineers. All types of abnormal process patterns are shown in Figure 3.

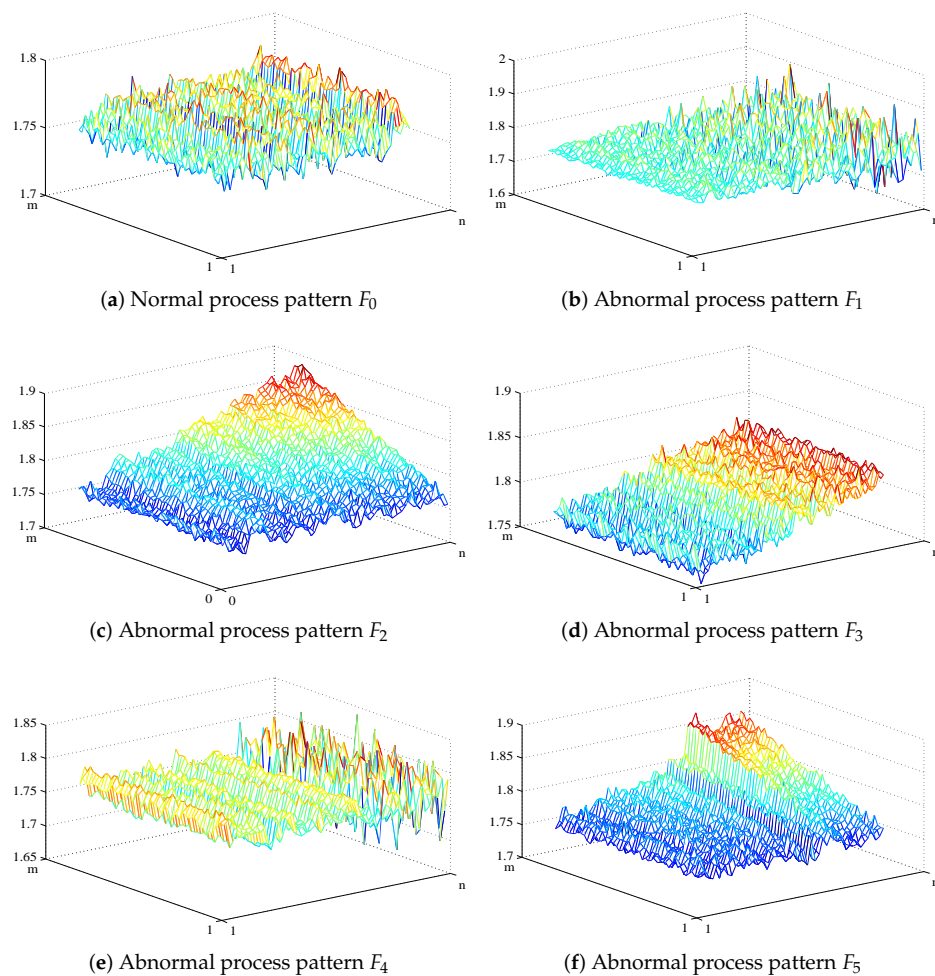


Figure 3. Cont.

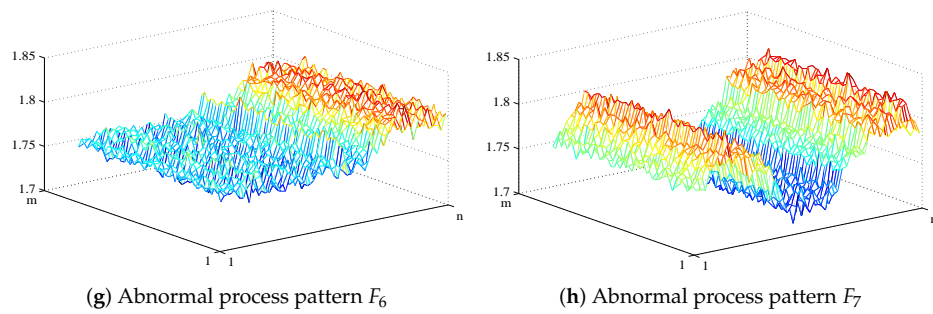


Figure 3. Normal and abnormal process patterns in the pasting process.

From Figure 3, it can be observed that the obvious shape differences among the spatial-temporal data of plate thicknesses show different abnormal process patterns. Therefore, the identification of abnormal process with spatial-temporal data are converted into the problem of identifying these abnormal process patterns. Once a certain abnormal process pattern is identified, its corresponding root causes could be found out simultaneously.

2.2. Abnormal Process Image Collection

Although the abnormal process patterns describe the abnormal states of the process with spatial-temporal data very well, high dimensionality, spatial-temporal correlation, and a high amount of noise make it difficult to identify them directly. Grayscale image can visually capture important abnormal patterns of observation data without any parameters predefined by users' experiences [20]. In order to effectively distinguish normal and seven abnormal process patterns, spatial-temporal data from the pasting process can be transformed into grayscale images, which is called process image. Given a spatial-temporal data matrix X , $x_{i,j}$ refers to a measured value, where i and j represent the location and time of the spatial-temporal data matrix, respectively. For generating the process image of the spatial-temporal data, each grayscale can be calculated from data matrix X by normalizing, multiplying 255 and taking integer. The transformation formula is given by:

$$y_{i,j}^{(0)} = \text{rounding}\left(\frac{x_{i,j} - \text{Min}(X)}{\text{Max}(X) - \text{Min}(X)} \times 255\right), \quad (1)$$

where $y_{i,j}^{(0)}$ is the grayscale value corresponding to $x_{i,j}$, *rounding* is the function for taking integer, and $\text{Min}(X)$ and $\text{Max}(X)$ are used for extracting maximum and minimal elements from matrix X . Thus, the spatial-temporal data have been represented as a process image to visualizing the operation status of the manufacturing process. Process images of normal and seven abnormal process patterns discussed above are shown in Figure 4. For the convenience, they are still denoted as F_0 , F_1 , F_2 , F_3 , F_4 , F_5 , F_6 , and F_7 .

In the grayscale image of a normal process pattern, the pixels are arranged randomly without obvious change. Because the spatial and temporal data of plate thickness varies unevenly, in the grayscale image F_1 , some pixels are white and some are dark. When the spatial-temporal data of plate thickness becomes steeper gradually in F_2 , the pixels of the steep data gradually appear to be white in the process image F_2 . The important features of other abnormal process patterns can be directly represented in their process images, as shown in Figure 4. From the above discussion, it can be observed that each process image can capture the trend and variation features of its corresponding pattern shown in Figure 3. Therefore, the problem of identifying abnormal process patterns is converted into detecting abnormal process images. How to use an effective recognition method of the abnormal process image is a challenge in this paper.

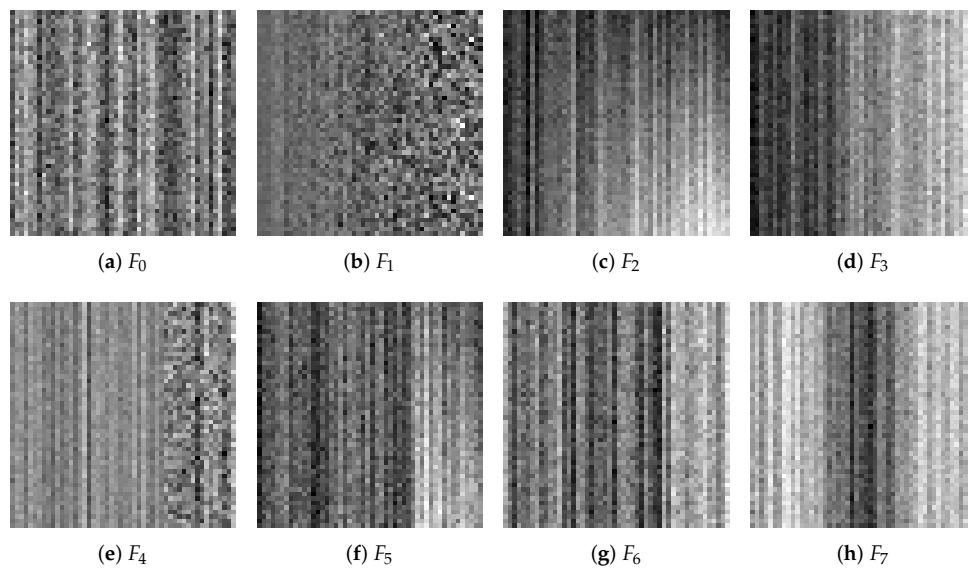


Figure 4. Normal and abnormal process images in the pasting process.

3. CNN Framework for Process Images Recognition

3.1. Architecture of the CNN Model

Convolutional neural network (CNN) is a kind of feed-forward artificial neural network (ANN), which is inspired by biological vision from pixel to abstract feature [21]. Unlike the traditional neural networks that need to concatenate raw data into a vector, CNN can directly deal with image data. Taking this advantage of CNN, the recognition model for process image will be constructed in this paper. A CNN recognition model is comprised of an input layer, several convolution layers and pooling layers, fully-connected (FC) layers, and an output layer, as illustrated in Figure 5. The input layer is to import process images for detection. The convolution layer and pooling layer are used to extract abnormal information from process images. The fully-connected layer serves as the integration of abnormal process information. The output layer is to provide the categories of abnormal process images.

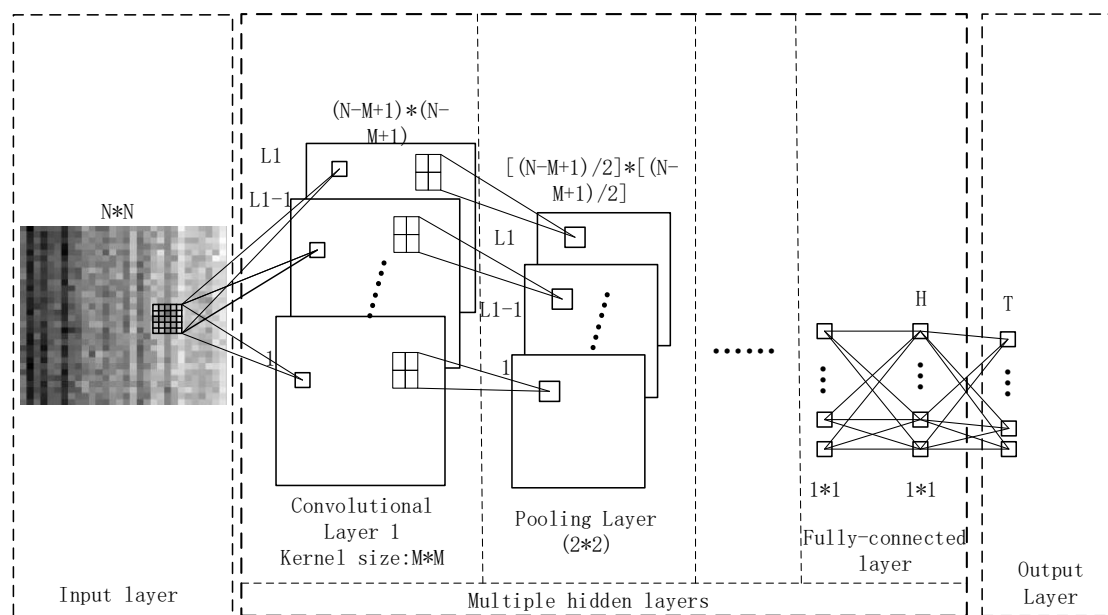


Figure 5. The general architecture of convolutional neural network (CNN) model for the process image.

A CNN model usually consists of several convolution and pooling layers. For convenience, suppose that there are R convolution and pool layers in CNN model and a process image is a square with $N \times N$ size, seen in Figure 5. Convolution layer is a most important component in the CNN model, which assigns weights to the grayscales of the input image by a convolution kernel, so as to extract process abnormal features.

3.2. Underlying Mechanism of the CNN Model

In the first convolution layer, suppose that there are L_1 convolution kernels. The $(w_{k,i,j}^{(1)})_{M \times M}$, ($k = 1, 2, \dots, L_1; i = 1, 2, \dots, M; j = 1, 2, \dots, M$) is used to represent the k th convolution kernel with $M \times M$ size ($M < N$), where $w_{k,i,j}^{(1)}$ means the weight value at row i and column j of the k th kernel. In order to get the convolution results, a moving window with $M \times M$ size should be set up. The window moves one stride at a time, where the area formed is called a receptive field. When all pixels in the process image are covered by the moving window, $(N - M + 1) \times (N - M + 1)$ receptive fields can be obtained. For all the receptive fields and convolution kernels in the first layer, the convolution result can be obtained by the following:

$$\begin{aligned} y_{k,i,j}^{(1)} &= f\left(\sum_{s=1}^M \sum_{t=1}^M w_{k,s,t}^{(1)} \cdot y_{i+s-1,j+t-1}^{(0)} + b_k^{(1)}\right), \\ i \text{ and } j &= 1, \dots, (N - M + 1), \\ s \text{ and } t &= 1, \dots, M, \\ k &= 1, \dots, L_1, \end{aligned} \quad (2)$$

where $y_{k,i,j}^{(1)}$ is the convolution output of the current receptive field for the k th convolution kernel, $w_{k,s,t}^{(1)}$ is the weight element at row s and column t of the k th convolution kernel, $y_{i+s-1,j+t-1}^{(0)}$ is the grayscale in current receptive field, and $b_k^{(1)}$ ($k = 1, 2, \dots, L_1$) represents the bias value. $f(\cdot)$ is ReLU activation function [22]. For the k th convolution kernel, all of the $y_{k,i,j}^{(1)}$ values from receptive fields formed a matrix, which is called a feature map corresponding to the k th convolution kernel ($k = 1, 2, \dots, L_1$). The feature maps of the first convolutional layer will be input to the pooling layer.

The pooling layer is mainly used to compress feature maps and obtain condensed feature maps via pooling function. In the first pooling layer, L_1 feature maps are entered respectively. Each feature map is partitioned into several non-overlapping square areas, which is known as pooling fields. In general, every pooling field with 2×2 size is preferred [23]. To get the pooling results, maximum pooling and average pooling are widely used pooling functions [24]. For the k th feature map, the pooling results, also known as condensed feature map, are obtained. The pooling value at row i and column j of the k th condensed feature map, $y_{k,i,j}^{(2)}$, can be computed by following equation:

$$\begin{aligned} y_{k,i,j}^{(2)} &= \text{pooling}(y_{k,2i-1,2j-1}^{(1)}, y_{k,2i-1,2j}^{(1)}, y_{k,2i,2j-1}^{(1)}, y_{k,2i,2j}^{(1)}), \\ i &= 1, \dots, \frac{N - M + 1}{2}, j = 1, \dots, \frac{N - M + 1}{2}, \end{aligned} \quad (3)$$

where $y_{k,2i-1,2j-1}^{(1)}, y_{k,2i-1,2j}^{(1)}, y_{k,2i,2j-1}^{(1)}, y_{k,2i,2j}^{(1)}$ are four values in the current pooling field connecting to $y_{k,i,j}^{(2)}$, and pooling refers to pooling function. After pooling operation, L_1 condensed feature maps are obtained, which will be input into the next convolution layer.

Similarly, the above work is repeated on several alternative convolution and pooling layers until the last pool layer. The condensed feature maps of the last pooling layer will be unfolded and imported to the Fully-connected (FC) layer. Before fully-connected operation, L_R condensed feature maps are unfolded to a vector, y_i^{FC} , $i = 1, \dots, L_R \times q \times q$. Suppose there are H nodes in the FC

layer, and, for the h th node, the weights connected to the i th input nodes can be denoted as $w_{i,h}^{(FC)}$ ($i = 1, 2, \dots, L_R \times q \times q, h = 1, 2, \dots, H$). The value of the h th node in the FC layer can be computed as follows:

$$y_h^F = f\left(\sum_{i=1}^{L_R \times q \times q} y_i^{(FC)} \times w_{i,h}^{(FC)} + b_h^{(F)}\right), h = 1, \dots, H, i = 1, \dots, L_R \times q \times q, \quad (4)$$

where y_h^F is the output the h th node in FC layer, $b_h^{(F)}$ ($h = 1, 2, \dots, H$) represents the bias of the h th node, and $f()$ is ReLU activation function. In the output layer, the connection between the h th FC node and the j th output node is represented by $w_{hj}^{(O)}$, ($h = 1, 2, \dots, H; j = 1, 2, \dots, T$). To classify the input data, the probability output of the j th output node is required to be computed as follows:

$$P_j = f\left(\sum_{h=1}^H y_h^{(F)} \times w_{hj}^{(O)} + b_j\right), h = 1, \dots, H, j = 1, \dots, T, \quad (5)$$

where T is the number of output nodes, P_j is the result of the j th node in the output layer, and b_j is the bias value of the j th output nodes. $f()$ is the normalized exponential function, through which the probability can be obtained:

$$f(y_j) = \frac{e^{y_j}}{\sum_{j=1}^T e^{y_j}}, j = 1, \dots, T. \quad (6)$$

As discussed above, the weights and biases in each layer of the CNN model can directly affect the final results of the output layer. If the difference between the actual output and the expected output is too large to be accepted, the weights and biases are required to be updated. The difference can be measured by the following loss function:

$$F = -\frac{1}{T} \sum_{j=1}^T [P_j^* \ln(P_j) + (1 - P_j^*) \ln(1 - P_j)], \quad (7)$$

where T refers to the total number of trained categories, P_j^* and P_j are the expected output and the actual output, respectively. A small loss value means an accurate probability output. To reduce the loss value, The back-propagation (BP) algorithm [25] is utilized. For convenience, the weights and biases in all layers are referred to as w and b , respectively. Thus, F is the loss function of w and b . The partial derivatives of F for w and b can be expressed as follows:

$$\begin{aligned} w_t &= w_{t-1} - \eta \frac{\partial F}{\partial w}, \\ b_t &= b_{t-1} - \eta \frac{\partial F}{\partial b}, \end{aligned} \quad (8)$$

where w_t and b_t represent the updated weights and biases after the t th iteration. η is the learning rate, and $\eta = 0.01$ is a common selection [26]. The CNN architecture parameters consist of the numbers of convolution layers and convolution kernels, the size of convolution kernels per layer, the selection of pooling function, and the number of output nodes in the FC layer. These parameters are required to be stepwise determined, which will be discussed in the next section.

3.3. CNN Identification Framework

A CNN framework for the identifying abnormal process with spatial-temporal data is presented in this paper. This framework can be divided into two phases: offline learning and online identifying. The offline learning phase aims to train a CNN model from offline collected process images and establish an appropriate CNN recognition model. In the online identifying phase, this CNN model trained in the offline phase can be applied to identify the abnormal process with real-time

spatial-temporal data. The CNN identification framework is shown in Figure 6, and the details are introduced in the following.

There are two main steps in offline learning: The first step is to obtain training samples, which consists of process images and their corresponding categories. Process images are converted by spatial-temporal data and their corresponding categories can be obtained according to engineering experience. The second step is to determine the architecture parameters of CNN, and update the weights and biases from the convolution, pooling, and FC layer of the CNN using the BP algorithm. After offline learning, the CNN recognition model will be obtained.

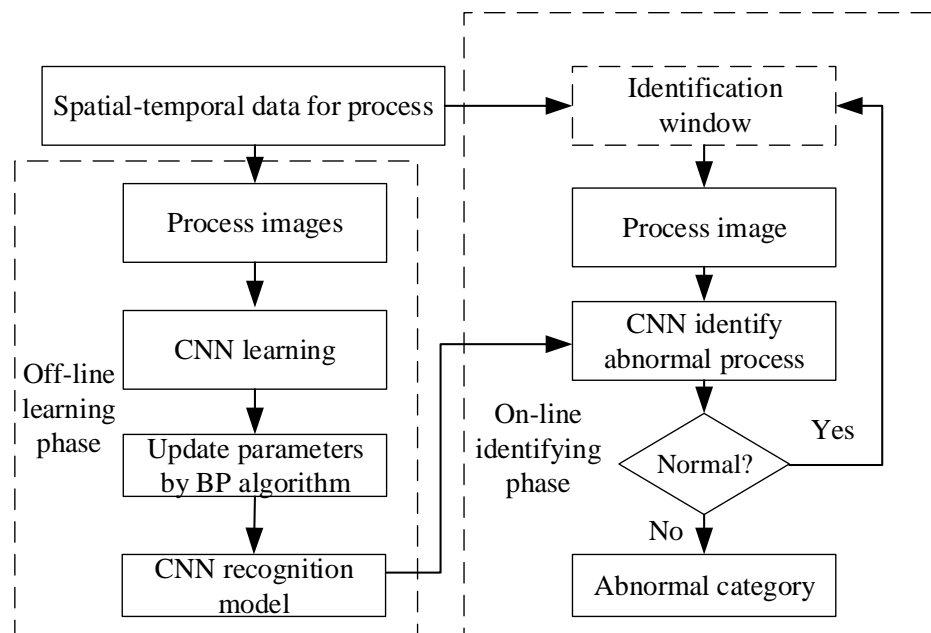


Figure 6. The CNN identification framework.

After the offline learning phase, the CNN recognition model is established and applied to online recognition of the abnormal process with spatial-temporal data. Two main steps in the online identifying phase are as follows: The first step is to collect the real-time spatial-temporal data from a process through the moving identification window. The size of the window should be determined by the product processing time, so that the spatial-temporal data in the current window can be mapped into a suitable process image. The second step is to identify the abnormal process images. The process images in the current identification window will be recognized by the CNN recognition model. If the decision result made by the CNN recognition model is a normal process image, the sliding window will move forward to collect new observations until an abnormal process image is identified.

According to the above two phases, the abnormal process with spatial-temporal data can be identified in a real-time way. It benefits from the powerful learning ability of the CNN recognition model.

4. Validation Results

The validation of the proposed approach depends on the architecture parameters of the CNN model, which will be investigated by the spatial-temporal data collected from the pasting process in this section. To evaluate the ability of detecting process images under various architectures, the recognition accuracy of test data (ATD) is used for the evaluation index, which is a ratio of the number of samples recognized correctly and the total number of samples. When an abnormal pattern happens in the process, the CNN model is expected to identify the abnormal type as precisely as possible. Meanwhile, when the process is normal, the CNN model is required to recognize the normal pattern more accurately,

so that the higher ATD indicates the better validity of the CNN model. The following validation analysis consists of the determination of layer and node number, and selection of the kernel size and pooling function. In addition, 700 process images with 50×50 size are collected from the pasting process under each process pattern separately, where 200 and 500 process images are selected randomly as the training and test samples, respectively. Training data are applied to learn the CNN recognition model with different architectures, and test data are used to compare their recognition accuracy to find the best one, which is conducted in Caffe, Ubuntu 16.04 with a Tesla K80 GPU.

4.1. Determination of Layer and Node Numbers

The number of layers and nodes discussed here contains the numbers of convolutional layers, convolution kernels per layer, and FC nodes. To achieve the optimal performance of CNN model, the number of convolution layers and kernels per layer are important parameters required to be specified in advance [27]. Adding more convolutional layers and kernels to the network could capture high-level features of input data at the price of making the model complex to train [28]. Thus, to obtain an optimal network architecture, the numbers of convolutional layers and convolution kernels will be determined layer by layer.

For the process image size of 50×50 size, the size of a condensed feature map will be 2×2 after four convolution and pooling operations, which cannot be further convolved. Thus, here we consider the CNN model with 4 convolution layers. Twenty scenarios for the numbers of convolution kernels from 5 to 100 are tested via recognition accuracy. The average accuracy and standard deviation are used to measure the performance of the proposed CNN model. To evaluate recognition accuracy, all the results of proposed method are replicated at least 100 times, and the results are shown in Table 1. By comparing different scenarios in the first convolution layer, the highest average recognition accuracy can reach 95.18% when the number of convolution kernels is 65. Considering the same scenarios in the second convolution layer, the number of convolution kernels is set to 70 because the highest average accuracy is 96.36%. In a similar way, the optimal numbers of convolution kernels can be determined layer by layer, shown in Table 1. It is expected that the recognition accuracy will increase until the optimal number of convolution layer is found out.

Generally, the recognition accuracy will be improved with the increase of convolution layers. However, we can see that the recognition accuracy begins to decrease from the third convolution layer in Table 1. In this situation, it can be inferred that the optimal number of convolutional layer is 3, and the corresponding number of convolution kernels is 65, 70, and 80 respectively, which is denoted as 65-70-80.

Under the 65-70-80 structure of the convolutional layer, the number of nodes in the FC layer needs to be determined. Ten scenarios for the numbers of FC nodes from 100 to 1000 are shown in Figure 7. By comparing the average recognition accuracy and standard deviation in different scenarios, we find that the accuracy increases to 98.08% at 800 nodes, and, after that, the accuracy does not exceed 98.08% by adding more nodes. Additionally, we observe that the variation of recognition accuracy becomes lower as the number of nodes in FC layer increases. From Figure 7, it is noted that the difference of recognition accuracy using various FC node numbers is not obvious enough, especially using 600 and 800. In this research, the node number corresponding to higher recognition accuracy and lower variation is preferred. Because the recognition accuracy using 800 nodes is slightly better than using 600, and the optimal number of FC nodes is set to 800. For the above discussion, the optimal architecture of CNN model is denoted as 65-70-80-800.

Table 1. The recognition accuracy comparison among different number of layers and convolution kernels. The boldface entries represent the highest average accuracy under the current layer. Numbers in parentheses are standard deviation of recognition accuracy.

The Number of Convolution Kernels	Accuracy of Test Data (%)			
	1st Convolution Layer	2nd Convolution Layer	3rd Convolution Layer	4th Convolution Layer
5	94.16 (0.232)	95.56 (0.253)	96.45 (0.225)	96.32 (0.336)
10	94.37 (0.145)	95.90 (0.238)	96.96 (0.169)	96.68 (0.278)
15	94.67 (0.117)	96.01 (0.146)	97.12 (0.187)	96.70 (0.168)
20	94.76 (0.156)	96.21 (0.154)	97.18 (0.119)	96.83 (0.165)
25	94.78 (0.110)	96.12 (0.162)	97.21 (0.172)	96.97 (0.157)
30	94.86 (0.118)	96.12 (0.102)	97.28 (0.123)	97.03 (0.146)
35	94.80 (0.114)	96.23 (0.162)	97.31 (0.168)	97.08 (0.137)
40	94.92 (0.147)	96.25 (0.166)	97.35 (0.158)	97.25 (0.148)
45	94.98 (0.120)	96.22 (0.116)	97.40 (0.162)	97.10 (0.151)
50	94.97 (0.116)	96.24 (0.168)	97.01 (0.128)	97.02 (0.139)
55	95.05 (0.100)	96.25 (0.081)	97.42 (0.144)	96.89 (0.155)
60	95.10 (0.115)	96.30 (0.104)	97.66 (0.201)	96.82 (0.187)
65	95.18 (0.112)	96.28 (0.336)	97.61 (0.159)	96.82 (0.103)
70	95.12 (0.126)	96.36 (0.120)	97.83 (0.173)	96.69 (0.098)
75	95.09 (0.098)	96.14 (0.064)	97.91 (0.139)	96.64 (0.111)
80	95.10 (0.102)	96.32 (0.149)	98.00 (0.106)	96.69 (0.132)
85	95.11 (0.121)	96.33 (0.105)	97.78 (0.092)	96.76 (0.139)
90	95.14 (0.133)	96.30 (0.071)	97.43 (0.094)	96.53 (0.167)
95	95.16 (0.075)	96.33 (0.090)	97.39 (0.122)	96.50 (0.143)
100	95.14 (0.116)	96.34 (0.110)	97.36 (0.118)	96.31 (0.121)

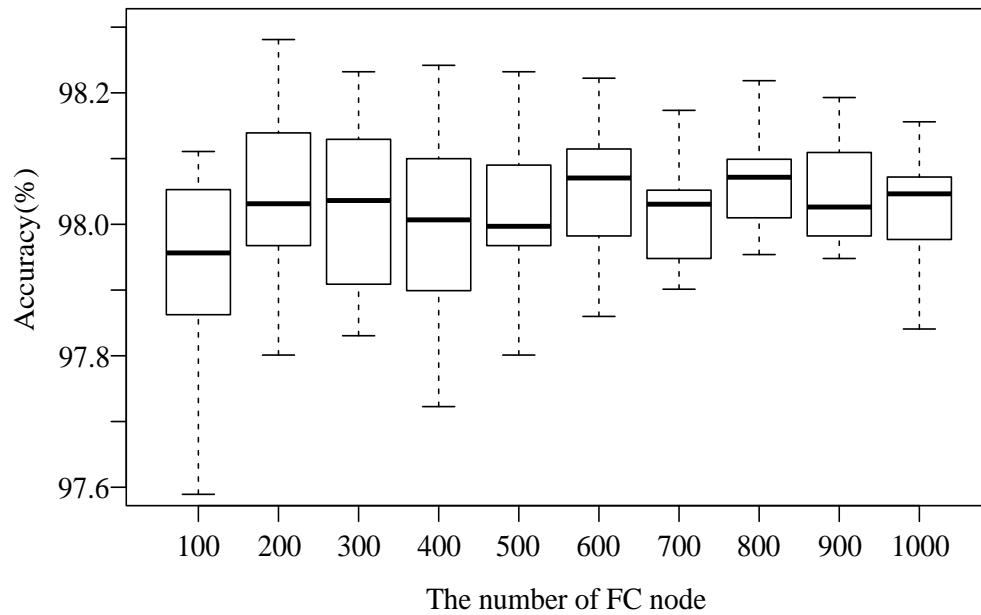


Figure 7. The accuracy comparison among fully-connected (FC) node numbers.

4.2. Selection of Kernel Size and Pooling Function

The sizes of convolution kernels and pooling function are also important architecture parameters of a CNN model. The size of convolution kernels will be first discussed. In general, a convolution kernel with a small size can capture details of abnormal information from process images. Thus, under the optimal architecture, 65–70–80–800, the CNN recognition models with kernel sizes 3×3 , 5×5 , and 7×7 are considered respectively, seen in Table 2. The results of the CNN model with three kernel sizes are obtained and the optimal kernel size can be determined as 3×3 . Then, the optimal pooling function will be selected in the following.

Table 2. The recognition accuracy comparison for kernel sizes. The boldface entries represent the highest accuracy.

Architecture	Kernel Size	Average Accuracy of Test Data (%)
65-70-80-800	3×3	98.08
65-70-80-800	5×5	97.27
65-70-80-800	7×7	96.58

Like the above discussion, the max-pooling and average-pooling shown as Equations (4) and (5) are widely used pooling functions. The comparison of recognition accuracy for two pooling functions is shown in Table 3. The max-pooling function is the best one, which can be used in the CNN recognition model.

Table 3. The recognition accuracy comparison for pooling functions. The boldface entries represent the highest accuracy.

Architecture	Kernel Size	Pooling Function	Accuracy of Test Data (%)
65-70-80-800	3×3	Max pooling function	98.08
65-70-80-800	3×3	Average pooling function	92.29

To summarize, the CNN recognition model for abnormal pasting process images shown in Table 4 has been constructed, and its validation has been proved.

Table 4. The architecture parameters of the proposed convolutional neural network (CNN) recognition model for the pasting process.

The Number of Convolution Layers	The Number of Kernels	Kernel Size	Pooling Function	The Number of Nodes in FC Layer
3	65-70-80	3×3	Max pooling function	800

5. Performance Comparison

In order to demonstrate the performance of the proposed method, the abnormal pasting process with spatial-temporal data will be identified in this section. The plate thickness data used for performance comparison are obtained from the pasting machine. After data cleaning and clustering analysis, the data of normal pattern and abnormal patterns is extracted and converted to process images. Each type of process patterns includes 200 and 500 process images with 50×50 size for training and testing respectively.

The abnormal pasting process images are online recognized using the constructed CNN recognition model. In the real pasting process, because the processing time of a plate is 0.5 s, the plate thickness observations at 50 locations in 25 s can be mapped to a process image with 50×50 . Thus, the size of the moving identification window is set to 50×50 . Taking 0.5 s as a stride length, once a new observation is obtained, the sliding window will move one step. As the window slides, the process images formed by the first 64 observations change smoothly and steadily, in which the output results of the CNN model are normal process images, shown in Figure 8. When the window moves to observation 115, the probability output is (0, 0, 0.005, 0.012, 0.003, 0.007, **0.973**, 0), which indicates that the abnormal process pattern F_6 happens in the pasting process. The corresponding cause is insufficient conveyor belt tension caused by the electromagnetic fault of the air pressure machine. When the root cause is removed, the plate thickness comes back to the target value nearby. When the identification window moves to the 199th observation, the probability output of CNN model is (0.011, 0.027, 0.02, 0, **0.934**, 0.008, 0, 0), which means that the abnormal pattern F_4 is detected. After checking the current process, we find that the plate flatness changes suddenly, which results from the low steel strength of new grids. After replacing this batch of grids with other qualified grids, the process returns to normal. When observation 281 enters the moving window, the probability output is (0.002, 0.008, 0, **0.9825**, 0, 0.003, 0.0045, 0). Adjusting the levelness of the pasting machine makes the process go back to normal.

From the above application, we can conclude that the proposed method has practicality to the online identification of abnormal process with spatial-temporal data. To further evaluate the performance of proposed CNN recognition model, the UMPCA based recognition method [11] as a benchmark method is considered for comparison. In this UMPCA based method, Bayes Classifier (BC) is utilized to achieve identification, which is denoted as UMPCA-BC. After 100 experiments, the comparison results are obtained and shown in Figure 9.

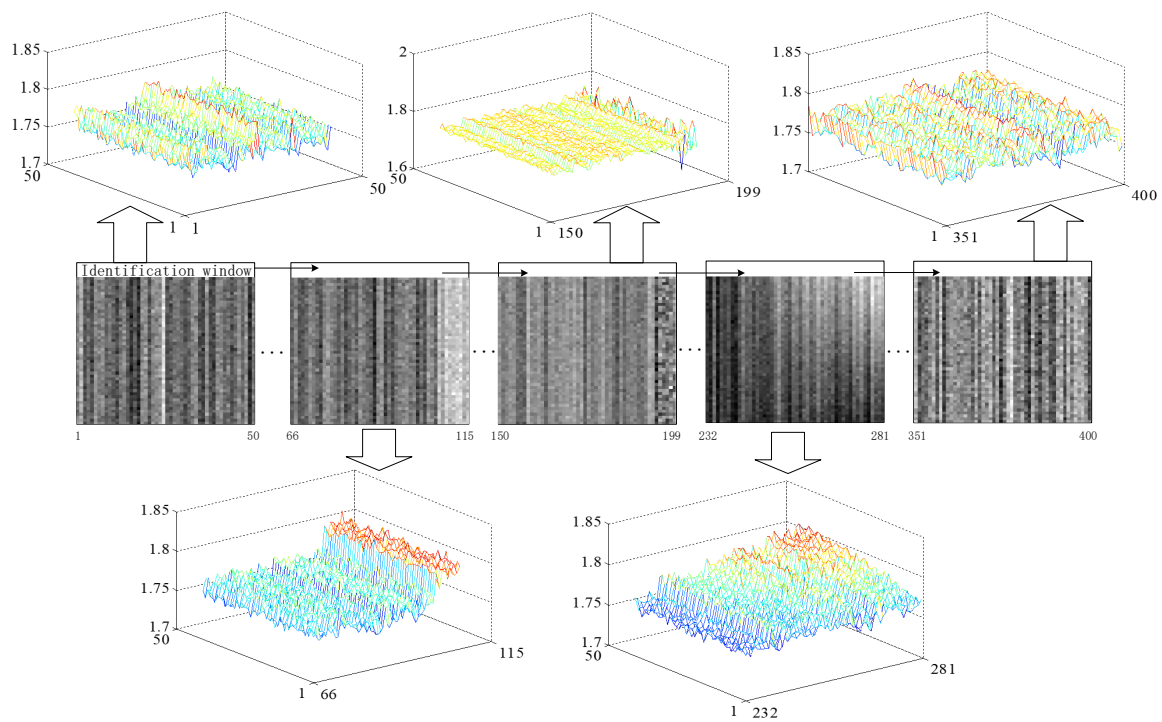


Figure 8. Online identification for the pasting process.

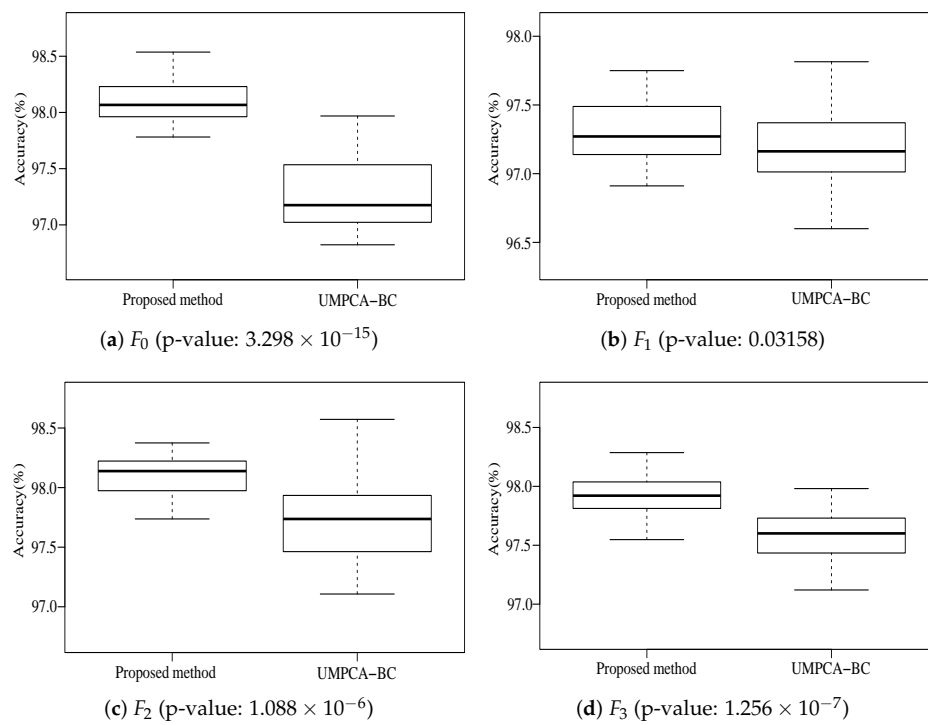


Figure 9. Cont.

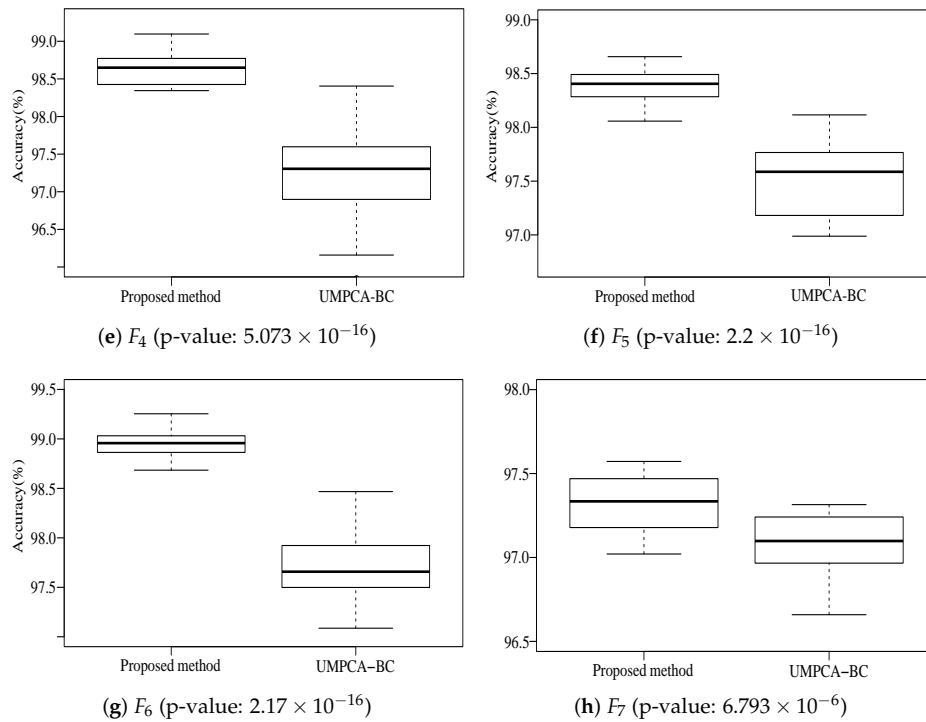


Figure 9. Performance comparison. Numbers in parentheses are the p -values of the Mann–Whitney U test.

From Figure 9, in all process patterns, the proposed method is better than UMPCA-BC, though only slightly better at F_1 . To further test whether the proposed method outperforms the alternative method significantly in all process patterns, Mann–Whitney U tests are conducted, and their p -values are also presented in Figure 9. If the 1% level of significance is taken, the superiority of the proposed model is significant except F_1 . There are two reasons for this situation. Firstly, the rounding operation of the grayscale image makes the F_1 process lose detail information, which further makes our model not sensitive to a small shift in abnormal process pattern F_1 . Secondly, in practical production, the failure of the acid spouting system does have a great impact on the thickness uniformity of plates in a short time. In other words, the variation of thickness uniformity in F_1 is usually small in a limited size of identification window, which increases the similarity between normal pattern and F_1 pattern. As a result, the performance of the proposed method has a limited improvement at F_1 . However, in general, the CNN method proposed in this paper can identify the abnormal process with spatial-temporal data more accurately than the UMPCA-BC method.

To further verify the reliability of the proposed model, sensitivity analysis is carried out and the effect of noise on performance is studied. The white Gaussian noise is generated and added to the original test data as follows:

$$x_{i,j}^{noise} = x_{i,j} + g, g \sim N(0, \sigma^2), \quad (9)$$

where $x_{i,j}$ is the value at row i and column j of original data matrix, and g refers to the noise that obeys normal distribution with a mean of zero and a standard deviation of σ . In this example, the specification limits of plate thickness are 1.75 ± 0.02 , and only the values of σ at 0 to 0.02 are considered. For convenience, five scenarios, including 0, 0.005, 0.01, 0.015, and 0.02 are implemented, shown in Table 5.

Table 5. The performance comparison for various noise levels.

Noise Level (σ)	Average Recognition of Test Data (%)	
	Proposed Method	Uncorrelated Multilinear Principal Component Analysis and Bayes Classifier (UMPCA-BC)
0	98.08	97.41
0.005	97.17	95.98
0.01	96.01	95.13
0.015	94.24	93.46
0.02	93.44	92.73

The recognition accuracy of the proposed method and UMPCA-BC is inevitably decreased with an increase of σ . However, from Table 5, the proposed method still outperforms UMPCA-BC with an increasing noise level. Therefore, the CNN method proposed in this paper can identify the abnormal process with spatial-temporal data with better results.

6. Limitations of the Proposed Methodology

Some aspects may limit the application and assessment of the proposed framework, such as the following ones:

- All the process images are generated by normalizing, multiplying by 255, and rounding, which may result in some information loss. How to measure the impact of missing information on CNN model performance is worthy of attention.
- In this study, the recognition accuracy as an evaluation criterion of performance is not very comprehensive. In practice, there are other metrics and representations used to evaluate the performance of CNN model, such as training time and test data loss. It is an interesting topic to evaluate and optimize CNN model based on other metrics.
- Generally, the operation state of manufacturing process is also affected by other unexpected factors, such as the new abnormal mode caused by an unknown fault. This situation will affect the performance of the current recognition model. Therefore, it is necessary to update the current data and build the recognition model again to consider the new pattern.

7. Conclusions and Further Work

This paper develops a general framework based on the CNN model to detect the abnormal pattern and diagnose the causes in the pasting process with spatial-temporal data. Different from traditional schemes, the main contribution of our proposed framework makes full use of both normal and abnormal information from historical data, and it overcomes the dilemma of multiple data types in real applications. The proposed model is tested on the example of the pasting process and achieved a better recognition performance than the alternative method. Experimental results demonstrated that better performance can be achieved at all abnormal process patterns in the pasting process. In addition, the sensitivity analysis of noise is also provided to verify the superiority of the proposed method. In addition, the procedure for constructing the recognition model is convenient. Our proposed CNN recognition model shows the good potential of online monitoring and tracing the root cause simultaneously. Benefiting from the CNN model, the spatial and temporal interrelationship of abnormal information can be captured and all the historical information can be utilized by the proposed CNN model.

However, there are two outstanding issues on this topic. First, although this paper focuses on the pasting process, the CNN recognition framework we proposed could be applied to any other abnormal process monitoring and diagnosis, where the observations are spatial-temporal data. In order to improve the performance of the CNN recognition model, the CNN model should be investigated further to make the proposed framework more suitable for other general situations.

Second, the parameter optimization of the CNN model is a challenging work in the deep learning domain. However, it is not our concern in current work, thus the parameters of the CNN model only for the pasting process are determined in our paper. In fact, the architecture parameters are related to the data type, shape of the abnormal patterns, and the number of the data sample, thus more suggestions for determining proper parameters are needed. Some other advanced parameter optimization techniques should be added to the CNN framework to improve the recognition accuracy further. Therefore, future improvements can be conducted in the following ways. First, this framework can be modified to identify the process of other general situations by using the transfer learning method. Second, other advanced techniques for hyper-parameters optimization can be studied further to replace the manual method, such as heuristic search algorithms and design of experiments' techniques.

Author Contributions: Conceptualization, Z.Z. and Y.L.; methodology, Z.Z.; software, S.Z.; validation, Y.L.; formal analysis, Y.L. and Z.Z.; investigate Z.Z.; resources, Y.L.; data curation, Z.Z.; writing—original draft preparation, Z.Z.; writing—review and editing, Y.L.; visualization, Z.Z.; supervision, U.J.; project administration, Y.L.; funding acquisition, Y.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China Grant No. 71672182, U1604262, U1904211, and 71672209.

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

References

1. Yu, Y.; Workman, A.; Grasmick, J.G.; Mooney, M.A.; Hering, A.S. Space-time outlier identification in a large ground deformation data set. *J. Qual. Technol.* **2018**, *50*, 431–445. [\[CrossRef\]](#)
2. Michel, B.; Michel, M.; Yves, M.; Jean-Michel, P.; Bruno, P. Spatial outlier detection in the PM10 monitoring network of Normandy (France). *Atmos. Pollut. Res.* **2015**, *6*, 476–483. [\[CrossRef\]](#)
3. Harrou, F.; Kadri, F.; Khadraoui, S.; Sun, Y. Ozone measurements monitoring using data-based approach. *Process Saf. Environ. Prot.* **2016**, *100*, 220–231. [\[CrossRef\]](#)
4. Kulldorff, M. Prospective time periodic geographical disease surveillance using a scan statistic. *J. R. Stat. Soc. Ser. A Statistics Soc.* **2001**, *164*, 61–72. [\[CrossRef\]](#)
5. Wang, A.; Wang, K.; Tsung, F. Statistical surface monitoring by spatial-structure modeling. *J. Qual. Technol.* **2014**, *46*, 359–376. [\[CrossRef\]](#)
6. Megahed, F.M.; Wells, L.J.; Camelio, J.A.; Woodall, W.H. A spatiotemporal method for the monitoring of image data. *Qual. Reliab. Eng. Int.* **2012**, *28*, 967–980. [\[CrossRef\]](#)
7. Tsui, K.L.; Wong, S.Y.; Jiang, W.; Lin, C.J. Recent research and developments in temporal and spatiotemporal surveillance for public health. *IEEE Trans. Reliab.* **2011**, *60*, 49–58. [\[CrossRef\]](#)
8. Colosimo, B.M.; Pacella, M. On the use of principal component analysis to identify systematic patterns in roundness profiles. *Qual. Reliab. Eng. Int.* **2007**, *23*, 707–725. [\[CrossRef\]](#)
9. Ye, J.; Janardan, R.; Li, Q. GPCA: An efficient dimension reduction scheme for image compression and retrieval. In Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Seattle, WA, USA, 22–25 August 2004; pp. 354–363.
10. Lu, H.; Plataniotis, K.N.; Venetsanopoulos, A.N. Uncorrelated multilinear principal component analysis for unsupervised multilinear subspace learning. *IEEE Trans. Neural Netw.* **2009**, *20*, 1820–1836.
11. Paynabar, K.; Jin, J.; Pacella, M. Monitoring and diagnosis of multichannel nonlinear profile variations using uncorrelated multilinear principal component analysis. *Iie Trans.* **2013**, *45*, 1235–1247. [\[CrossRef\]](#)
12. Pacella, M. Unsupervised classification of multichannel profile data using PCA: An application to an emission control system. *Comput. Ind. Eng.* **2018**, *122*, 161–169. [\[CrossRef\]](#)
13. Zhang, L.; Gao, H.; Wen, J.; Li, S.; Liu, Q. A deep learning-based recognition method for degradation monitoring of ball screw with multi-sensor data fusion. *Microelectron. Reliab.* **2017**, *75*, 215–222. [\[CrossRef\]](#)
14. Dahl, G.E.; Yu, D.; Deng, L.; Acero, A. Context-dependent pre-trained deep neural networks for large-vocabulary speech recognition. *IEEE Trans. Audio Speech Lang. Process.* **2011**, *20*, 30–42. [\[CrossRef\]](#)
15. He, K.; Zhang, X.; Ren, S.; Sun, J. Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *37*, 1904–1916. [\[CrossRef\]](#)

16. Kiranyaz, S.; Ince, T.; Gabbouj, M. Real-time patient-specific ECG classification by 1D convolutional neural networks. *IEEE Trans. Biomed. Eng.* **2015**, *63*, 664–675. [[CrossRef](#)]
17. Wen, L.; Li, X.; Gao, L.; Zhang, Y. A new convolutional neural network-based data-driven fault diagnosis method. *IEEE Trans. Ind. Electron.* **2017**, *65*, 5990–5998. [[CrossRef](#)]
18. Brik, K.; ben Ammar, F. Causal tree analysis of depth degradation of the lead acid battery. *J. Power Sources* **2013**, *228*, 39–46. [[CrossRef](#)]
19. Schiffer, J.; Sauer, D.U.; Bindner, H.; Cronin, T.; Lundsager, P.; Kaiser, R. Model prediction for ranking lead-acid batteries according to expected lifetime in renewable energy systems and autonomous power-supply systems. *J. Power Sources* **2007**, *168*, 66–78. [[CrossRef](#)]
20. Wang, Y.; Zhou, H.; Feng, H.; Ye, M. Network traffic classification method basing on CNN. *J. Commun.* **2018**, *1*, 14–23.
21. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. In Proceedings of the Advances in Neural Information Processing Systems, Stateline, NV, USA, 3–8 December 2012; pp. 1097–1105.
22. Zhao, H.; Liu, F.; Li, L.; Luo, C. A novel softplus linear unit for deep convolutional neural networks. *Appl. Intell.* **2018**, *48*, 1707–1720. [[CrossRef](#)]
23. Feiyan, Z.; Linpeng, J.; Jun, D. Review of convolutional neural network. *Chin. J. Comput.* **2017**, *40*, 1229–1251.
24. Boureau, Y.L.; Ponce, J.; LeCun, Y. A theoretical analysis of feature pooling in visual recognition. In Proceedings of the 27th International Conference on Machine Learning (ICML-10), Haifa, Israel, 21–24 June 2010; pp. 111–118.
25. Khaw, H.Y.; Soon, F.C.; Chuah, J.H.; Chow, C.O. Image noise types recognition using convolutional neural network with principal components analysis. *IET Image Process.* **2017**, *11*, 1238–1245. [[CrossRef](#)]
26. Zou, J.; Wu, Q.; Tan, Y.; Wu, F.; Wang, W. Analysis Range of Coefficients in Learning Rate Methods of Convolution Neural Network. In Proceedings of the 2015 14th International Symposium on Distributed Computing and Applications for Business Engineering and Science (DCABES), Guiyang, China, 18–24 August 2015; pp. 513–517.
27. Gu, J.; Wang, Z.; Kuen, J.; Ma, L.; Shahroudy, A.; Shuai, B.; Liu, T.; Wang, X.; Wang, G.; Cai, J.; et al. Recent advances in convolutional neural networks. *Pattern Recognit.* **2018**, *77*, 354–377. [[CrossRef](#)]
28. Zeng, H.; Edwards, M.D.; Liu, G.; Gifford, D.K. Convolutional neural network architectures for predicting DNA–protein binding. *Bioinformatics* **2016**, *32*, i121–i127. [[CrossRef](#)]



© 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/>).