



Article Multi-Criteria Spare Parts Classification Using the Deep Convolutional Neural Network Method

Ke Yang ^{1,2}, Yongjian Wang ³, Shidong Fan ^{1,*} and Ali Mosleh ^{2,*}

- ¹ School of Energy and Power Engineering, Wuhan University of Technology, Wuhan 430063, China; k_yang@whut.edu.cn
- ² B. John Garrick Institute for the Risk Science, University of California, Los Angeles, CA 90095, USA
- ³ School of Data Science, City University of Hong Kong, Hong Kong 999077, China; zhinengjuece@outlook.com
- * Correspondence: sdfan@whut.edu.cn (S.F.); mosleh@g.ucla.edu (A.M.)

Abstract: Spare parts management is a critical issue in the industrial field, alongside planning maintenance and logistics activities. For accurate classification in particular, the decision-makers can determine the optimal inventory management strategy. However, problems such as criteria selection, rules explanatory, and learning ability arise when managing thousands of spare parts for modern industry. This paper presents a deep convolutional neural network based on graph (G-DCNN) which will realize multi-criteria classification through image identification based on an explainable hierarchical structure. In the first phase, a hierarchical classification structure is established according to the causal relationship of multiple criteria; in the second phase, nodes are colored according to their criteria level status so that the traditional numerical information can be visible through graph style; in the third phase, the colored structures are transferred into images and processed by structure-modified convolutional neural network, to complete the classification. Finally, the proposed method is applied in a real-world case study to validate its effectiveness, feasibility, and generality. This classification study supplies a good decision support to improve the monitor-focus on critical component and control inventory which will benefit the collaborative maintenance.

Keywords: spare parts; classification; deep learning; collaborative maintenance

1. Introduction

The transportation industry plays an essential role in national economics, while vehicles may suffer downtime from spare parts being out of stock. For spare parts classification management, decision-makers should select optimal inventory strategies in terms of cost and reliability. One third of the maintenance support cost of an industrial system may be spent on stocking ample spare parts for timely maintenance [1]. It is especially challenging to manage tens of thousands of spare parts necessary for modern industry. Thus, classification has been an important technique in the spare parts management that can support decision-makers in strategizing to allocate inventory with optimal cost and high reliability [2,3].

Spare parts classification has been extensively investigated by scholars [4–6]. Different service backgrounds bring about significant differences in the criteria selection and algorithms adopted by decision-makers. Among the traditional spare parts classification methods, both ABC and VED analyses are simple and basic methods in which spare parts are arranged according to only one criterion. The ABC method was first proposed by Gelder and Van Looy [7] and Tanwari et al. [8] to divide spare parts into categories (A, B, and C) using total annual cost and inventory criteria. VED analysis is a simple qualitative approach to decision-making based on spare parts "criticality" as defined by experienced experts. In practice, however, they do not entirely reflect the management requirements for spare parts because a single criterion is too subjective. As proposed in 1988 [9], a good-structured spare parts classification should be a value assessment system



Citation: Yang, K.; Wang, Y.; Fan, S.; Mosleh, A. Multi-Criteria Spare Parts Classification Using the Deep Convolutional Neural Network Method. *Appl. Sci.* 2021, *11*, 7088. https://doi.org/10.3390/app11157088

Academic Editors: Cristiana Delprete, Nicolò Zampieri and Eugenio Brusa

Received: 5 June 2021 Accepted: 28 July 2021 Published: 31 July 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/). that requires the combination of maintenance engineers, material managers, quality control personnel, and other relevant experts. Therefore, Kraljic [10] first proposed a second-order matrix of supply risk and criticality in view of multi-criteria, and other scholars have also applied this principle to develop new multi-criteria classification methods such as AHP (Analytic Hierarchy Process) and similar hierarchical-based analysis. Braglia [11] proposed the Multi-Attribute Spare parts Tree Analysis (MASTA) method, which uses a decision tree and AHP to assign calculations to various criteria. Cakir and Canbolat [12] combined AHP with fuzzy logic to capture the uncertainty in an evaluation indicator and export priority values from a judging set via fuzzy optimization. Zeng et al. [13] proposed an algorithm that integrates AHP, fuzzy evaluation, and gray association analysis to translate qualitative descriptions into quantitative data for spare parts classification in a limited and uncertain environment. Golam Kabir [14] also employed fuzzy AHP to determine the relative weights of attributes/criteria in a given inventory. Ng [15] used an optimized approach to solve the weights of each criterion in a classification system—assuming that each criterion for all spare parts falls into a descending order—and finally translated the score of each spare part into a uniform scale score. Based on the Ng model, Hadi-Vencheh et al. [16] calculated the weight of each spare parts in a system using a nonlinear optimization algorithm as the basis for the subsequent calculation of the total score. Though effective in some regards, AHP-based methods are still not well-suited to real-world spare parts classification due to the subjective matrix assignments and limited learning ability.

In order to overcome the defects in multi-criteria models regarding to weight assignment and reclassification learning ability, scholars have attempted to optimize heuristic intelligent algorithms such as Data Envelopment Analysis (DEA), Support Vector Machine (SVM), Artificial Neural Network (ANN), Genetic Algorithm (GA), Particle Swarm Optimization (PSO), gray theory, fuzzy logic, and combinations thereof. Yu [17] compared SVM, backpropagation (BP), K-Nearest Neighbors (KNN) and other methods to find that artificial intelligence-based methods are relatively more accurate. Partovi [18] also utilized BP and GA learning methods in ANNs for reliable, accurate inventory classification. Cai [19] adopted an SVM based on a Kraljic three-dimensional (3D) model for enhanced qualityoriented material classification accuracy. Liu [20] integrated a fuzzy neural network and decision tree for spare parts classification in terms of actual demand in a steel enterprise with multiple attributes. Guo [21] constructed a novel kNN model wherein the k value can be optimized automatically to enhance the efficiency and accuracy of classification. Models based on intelligent or heuristic algorithms can process high-latitude and nonlinear sample data, but lack the rules explanatory in a degree.

Thus, for the existing spare parts classification methods, there may be some persistent problems summarized from the following three aspects:

- Visualization: Neither multi-criteria nor traditional ANN models can provide decisionmakers with intuitive, visually interpretable data.
- (2) Rules explanatory: Data-driven artificial intelligent algorithms proceed in a "black box", where rules are not explained and results are relatively uninterpretable, despite their certain learning ability.
- (3) Learning ability: While there are understandable variable-selection and structureconstruction processes in rule-based methods, their learning ability and generalization are relatively poor.

Therefore, a novel approach based on hierarchical classification and the Convolutional Neural Network (CNN) will be developed in this study to resolve the above problems. The proposed model combines the advantages of both hierarchical structure and CNN.

Hierarchical classification serves to construct a criteria structure based on the causal relationship of multi-criteria, which makes for an explanatory rule-setting process. Moreover, the multiple criteria are distributed on the subspace by reducing the hierarchy in order to reduce the restriction of the model on the number of criteria. As described by Hu and Wang [22], the hierarchical structure shows the causal relationship between criteria and makes it possible to consider as many variables as demanded. Bacchetti [23] created a hierarchical multi-criteria spare parts classification method and verified it in an Italian household appliances manufacturing company. With the constructed structure, the criteria can be colored for specific data features, which transfers the classification into an image-identification problem.

The CNN is a deep learning method [24–26] for image identification and classification, and several networks based on CNN have been explored in the past years, such as LeNet, AlexNet, VGG (Visual Geometry Group)-Net, ResNet and DenseNet. Wang [27] proposed a CNN-based symbolic hierarchical cluster method for operation trend prediction problems, where the CNN structure is used to classify trends. Ren [28] used AlexNet to classify different disturbances in a 3D-printing heating map process. All developed networks from CNN have their own excellent specifics. The core of ResNet model is to establish skip connection between the front layer and the back layer, which is helpful in relieving the gradient problem by residual fitting in the back propagation of training process [29]. DenseNet is mainly distinctive in the dense connection and feature reuse that improve the calculation capacity [30]. However, AlexNet is a deep CNN structure which can solve gradient problem better with the Relu function. Additionally, both dropout and GPUs (Graphics Processing Units) technology are adopted to improve the calculation capacity when considering the overfitting problem [31].

This paper is organized as follows: Section 2 introduces the deep CNN principle, which is the basis of the proposed algorithm. The proposed G-DCNN method and its stepwise operation process are presented in Section 3. Section 4 reports the case study conducted to validate the proposed method, as well as a comparison between the proposed method and other similar methods. Section 5 gives a summary, concluding remarks, and a brief statement regarding future research directions.

2. Preliminaries

The deep CNN used in this study is called AlexNet, which has a structure first proposed in 2012 by Alex Krizhevsky. The overall structure of the AlexNet network is similar to that of LeNet. Both networks are convolved first and then fully connected, though the details are rather different.

2.1. Convolution Layer

Convolution operation is the most important link in the CNN structure. Local features with different dimensions are obtained using different convolutional kernels. The convolution layer has three parts: convolutional calculation, residual calculation, and gradient calculation.

2.1.1. Convolutional Calculation of the Convolution Layer

Assuming that the *i*th layer is the convolutional layer and the (i + 1)th layer is the subsampling layer, the *j*th feature map of the *i*th layer can be calculated as follows:

$$x_j^i = f\left(\sum_{i \in M_j} x_j^{i-1} * k_{ij}^i + b_j^i\right)$$
(1)

where *M* is the feature map, *k* is the convolutional kernel, *b* is the bias, and * is the convolutional operation.

2.1.2. Residual Calculation of the Convolution Layer

Subsampling layers are always placed after the convolutional layer. A one-to-one non-overlapping sampling method is adopted. The residual of *jth* feature map of the *ith* layer can be expressed as follows:

$$\delta_j^i = \beta_j^{i+1} \left(f'\left(u_j^i\right) \circ up\left(\delta_j^{i+1}\right) \right)$$
(2)

where δ is the residual, β is equal to the "weight" defined at a downsampling layer map, $f(\cdot)$ is the output activation function, u is the output of the certain layer, \circ is the element-wise multiplication, and $up(\cdot)$ is an upsampling operation.

2.1.3. Gradient Calculation of the Convolution Layer

Gradient descent calculation can provide the minimum loss function and model parameter values. The derivation of the bias *b* is:

$$\frac{\partial E}{\partial b_i} = \sum_{m,v} \left(\delta^i_j \right)_{mv} \tag{3}$$

where *E* is the error and (m, v) is the position of the output convolution map. The derivation process of the kernel *k* is:

$$\frac{\partial E}{\partial k_{ij}^i} = \sum_{m,v} \left(\delta_j^i\right)_{mv} \left(p_i^{i-1}\right)_{mv} \tag{4}$$

where *p* is the patch, which is multiplied elementwise by the kernels during convolution.

2.2. Subsampling Layer

The subsampling layer (also called the "pooling layer") is operated similarly to the convolutional layer. However, the convolution kernels of the subsampling layer only take the maximum value or average value of the corresponding position; the subsampling layer does not take any modification due to BP into consideration either.

2.2.1. Convolution Calculation of the Subsampling Layer

Assuming that the *lth* layer is the subsampling layer, the (l-1)th layer is the convolution layer, and the sampling size is 2 * 2, the convolution process can be described as follows:

$$x_j^l = f\left(\beta_j^l down\left(x_j^{l-1}\right) + b_j^l\right) \tag{5}$$

where $down(\cdot)$ represents the sum of the 2 * 2 size of the previous convolutional layer. The results are multiplied by the weight *w* and the activation function is applied after adding the bias.

2.2.2. Residual Calculation of the Subsampling Layer

The residual calculation formula used here is constructed in MatLab.

$$\delta_j^l = f'\left(u_j^l\right) \circ conv2\left(\delta_j^{l+1}, rot180\left(k_j^{l+1}\right), full'\right) \tag{6}$$

where *rot*180 represents rotating the kernel to make the convolution function perform cross-correlation, *'conv2'* is the convolution operation, and *'full'* is the full convolution process.

2.2.3. Gradient Calculation of the Subsampling Layer

The bias *b* derivation process of the subsampling layer is the same as the convolution layer. The weight β derivation process is:

$$\frac{\partial E}{\partial \beta_j} = \sum_{u,v} \left(\delta_j^l \circ down\left(x_j^{l-1} \right) \right)_{uv} \tag{7}$$

2.3. AlexNet Structure

For original AlexNet structure, as shown in Figure 1, there are a total of eight layers including five convolutions and three fully connected layers in this network. It is obvious that the AlexNet structure proceeds in both the upper and the down layer with 2GPUs at the same time, which improves the computing efficiency greatly [32].



Figure 1. The original AlexNet structure (extracted from [32]).

The first layer is the input layer wherein RGB images with pixels of $224 \times 224 \times 3$ are preprocessed into $227 \times 227 \times 3$. In this layer, 96 convolution kernels 11 * 11 * 3 in size are used to extract the features. With the stride movement of four pixels, FeatureMap generates $55 \times 55 \times 96$. FeatureMap is input into the first ReLu activation function and then processed via max-pooling with a pooling unit of size 3 * 3 to obtain pixel sets of $27 \times 27 \times 96$. Local response normalization is adopted in this layer with the output $27 \times 27 \times 96$, and the data is divided into two groups as $27 \times 27 \times 48$ because of two GPUs.

Next, for the second layer, the input has been two sets of $27 \times 27 \times 48$. For each set, 128 convolution kernels 5 * 5 * 48 bring about the output of $27 \times 27 \times 128$. There are the same calculation processes in this layer as the first layer, including ReLu function and local response normalization, finally with the output of two sets $13 \times 13 \times 128$.

The third layer is fully connected to the two GPUs in second layer. However, only convolution and ReLu are processed in this layer with 384 convolution kernels 3 * 3 * 256, with the FeatureMap of $13 \times 13 \times 384$ and $13 \times 13 \times 192$ for each GPU.

Same process happen in the fourth layer with the output of $13 \times 13 \times 192$.

In the fifth layer, the input data are two sets of $13 \times 13 \times 192$ with 128 convolution kernels of size 3 * 3 * 192. The scale of the pooling operation is 3 * 3, the step size of the operation is 2, and then the output of $6 \times 6 \times 256$ is obtained.

The sixth layer is a fully connected layer with the input of $6 \times 6 \times 256$. There are a total of 4096 filters with size of 6 * 6 * 256 that can perform convolution operations. Since the size of convolution kernel is same as the size of input, there is only one value after convolution operation. Therefore, the size of the convolution pixel layer is $4096 \times 1 \times 1$. Namely, there are 4096 neurons. The 4096 neurons are processed by ReLu function and output 4096 values.

There are same fully connected, ReLu function and Dropout process in the seventh layer with the 4096 input. Finally, the 4096 data are fully connected to the 1000 neurons in the eighth layer. The probability of the 4096 data moving toward the 1000 categories can be calculated accordingly. The category corresponding to the largest probability value is selected as the category of the initial input image.

Unlike the SVM method, which has an uncalibrated computation process and scores for all classes that are not easily interpretable, the Softmax classifier allows the user to compute "probabilities" for all labels. For example, given an image, the SVM classifier might give scores [12.5, 0.6, -23.0] for the classes "cat", "dog", and "ship". The Softmax classifier can instead compute the probabilities of the three labels as [0.9, 0.09, 0.01], which allows the confidence in each class to be interpreted

3. Proposed G-DCNN Method Based on Modified AlexNet

3.1. Modified AlexNet Structure for Classification

Transfer learning is an important machine-learning tool for applying knowledge learned in a domain to different but related domains or issues [33]. Thus, transfer learning thought is used in this study to modify the original AlexNet structure for spare parts classification task. During transfer learning, only the dense layers are restructured and



retrained to identify features that are specific to the spare parts classification. The modified structure and dimensions of crucial layers are shown in Figure 2.

Figure 2. Modified AlexNet structure.

Specially, the input dataset are the spare parts hierarchical structure images with the preprocessed resolution of 227×227 pixels. Additionally, since the spare parts are usually classified into four classes (A, B, C and D), we mainly modify the final softmax layer and change the 1000 neurons to four, which is the same size as the number of the target categories, as shown in Figure 3.



Figure 3. Modified softmax layer and classification.

The spare part's hierarchical structure images are processed by convolutional layers and pooling layers which can recognize the graphic features through trained weighted matrices. During the process, the filters expand overall dimensions of the input data through the creation of multiple feature maps. Additionally, max-pooling approach is used to down-sample the input information and extract the features. Each extracted feature from the convolutional layers is assigned weights with respect to each classification category. Then, the fully connected layers use the weight matrices to compute the confidence level of the input belonging to each category.

3.2. Proposed G-DCNN Model Based on Modified AlexNet

The proposed G-DCNN model can be divided into two main phases: constructing the hierarchical structure and classification using the modified AlexNet structure.

Feature selection is always the basis of a classification problem. In terms of feature selection for hierarchical classification, the expert knowledge base is necessary to obtain the most closely related criteria to the target. Traditional methods provide only limited selections of characteristic criteria that cannot accurately represent the samples. Therefore,

it is necessary to layer the multiple criteria according to the relationship in a particular task and to determine the final target from each layer with as little feature space as possible. The hierarchical structure effectively exploits qualitative and quantitative information and solves complex decision-making problems with multiple criteria. As shown in Figure 4, the selected criteria are built into layers here according to their causal relationships.



Figure 4. The hierarchical classification structure.

The structure has clearly explainable rules and provides a workable basis for the subsequent image conversion and classification. Then, all the involved criteria are colored according to their level status. The multi-criteria hierarchical structure can be converted into an image database, then the modified AlexNet structure can be used to obtain further classification results. As shown in Figure 5, the G-DCNN model is operated in a stepwise process:

- (1) Select the criteria related to the target.
- (2) Construct the hierarchical structure according to the causal relationships among criteria.
- (3) Grade the actual data into different levels.
- (4) Color the nodes according to the criteria level status.
- (5) Converse the hierarchical structure diagrams into images.
- (6) Initialize the parameters of the neural network.
- (7) Input the built spare parts level structure graphs.
- (8) Resize the input images to 227×227 to suit the modified neural network.
- (9) Label the images and build the dataset artificially.
- (10) Complete the parameters initialization and function activation.
- (11) Divide the input images into training (70%) sets and testing (30%) sets.
- (12) Train the modified AlexNet structure with the training sets and check the results;
- (13) Determine whether to proceed testing step; otherwise, adjust parameters and return to the training step.
- (14) During the testing procedure, the accuracy can be used to determine whether to save the model and otherwise, return to initialization step and proceed the whole learning process again.
- (15) Save the model and end the classification process.



Figure 5. General process of the proposed approach.

4. Case Study

In this section, the proposed G-DCNN model is applied and validated with a realworld case study.

4.1. Problem Description

Ships are a typical long-cycle equipment system with an average life span of 30–50 years. Throughout the life cycle, they are subject to regular repairs and maintenance requiring extensive spare parts storage and supply. A regular ship has more than 10 systems consisting of 4000+ components as shown in Figure 6. Some of the components are critical to the reliable operation but highly expensive to keep in stock at all times. It is also common for the ship to carry a large number of spare parts that may not be replaced for years.



Figure 6. Complex ship system and spare parts.

There are a total of 73 ships served on the trading routes in this case study company, which demand vast types and huge quantities of spare parts to ensure reliable operation over the equipment's whole life cycle. Therefore, it is essential and challenging for the shipping company to make strategic purchasing and stocking decisions according to the spare parts classes.

4.2. Hierarchical Structure of Ship Spare Ship Parts Classification

There are three main categories of classification criteria most commonly used in cases such as: (1) the inherent properties of the spare parts (criticality and cost); (2) consumption characteristics including failure rate and inventory turnover; (3) supply chain characteristics (lead time and supply reliability). Fourteen criteria are selected for this case study under advisement of the spare parts manager to determine the final class of spare parts: X₁ (average lead time), X₂ (number of suppliers), X₃ (system monitoring level), X₄ (criticality), X₅ (annual consumption), X₆ (average unit cost), X₇ (safety inventory), X₈ (supply reliability), X₉ (demand predictability), X₁₀ (out-of-stock risk), X₁₁ (out-of-stock cost), X₁₂ (annual consumption cost), X₁₃ (inventory holding cost), X₁₄ (in-stock cost). Table 1 provides further description of the 14 criteria.

Table 1. Multiple spare parts classification criteria.

Criteria	Name	Description
X ₁	Average lead time	The time between order replacement and arrival of the demanded spare parts.
X ₂	Number of the suppliers	The number of suppliers that can supply some kind of spare parts.
X ₃	System monitoring level	Assessed by the maintenance manager who has long experience on the ships.
X_4	Criticality	Assessed by the maintenance manager who has long experience on the ships.
X5	Annual consumption	Calculated from the historical data of past spare parts usage.
X ₆	Average unit cost	The spare price.
X ₇	Safety inventory	A buffer stock prepared to prevent the uncertain factors of future demand.
X ₈	Supply reliability	The suppliers' fulfillment to orders.
X9	Demand predictability	The predictability of the spare parts demand.
X ₁₀	Out-of-stock risk	The possibility of unavailable spare parts.
X ₁₁	Out-of-stock cost	The penalty cost caused by out-of- stock of the spare parts.
X ₁₂	Annual consumption cost	Calculated by average unit cost multiply annual consumption.
X ₁₃	Inventory holding cost	The cost to hold the safety inventory.
X ₁₄	In-stock cost	Calculated by annual consumption cost plus inventory holding cost.

All the 14 criteria are collected based on expert knowledge and the literature in view of inventory, supply, and technology information. There are seven input variables in this case, as shown in Table 1, wherein X_1 (average lead time), X_2 (number of suppliers), X_5 (annual consumption), X_6 (average unit cost), and X_7 (safety inventory) are quantitative values gained from the actual dataset while X_3 (system monitoring level) and X_4 (criticality) are qualitative values assessed by the spare parts manager based on their personal experience. The hierarchical structure of spare parts classification in the ship system is constructed here according to both causal relationship and expert knowledge. As shown in Figure 7, the relationship of multiple criteria can be explained clearly.

For criteria, they can be scaled into different levels according to the data feature space. The original structure and distribution of criteria data for the spare parts are shown in Table 2.

In the hierarchical structure, the variable relationship is constructed into layers and a final level emerges that is determined by the inventory cost and out-of-stock cost. The target output of hierarchical structure falls into four classes: A, B, C and D. The spare parts in class A are always highly expensive to keep in stock but are seldom needed. These out-of-stock spare parts will bring about the system heavy losses. Thus, the corresponding component is focused on closely monitoring and prognosis. The spare parts in class B are also highly expensive to keep in stock but not as much as class A, and the replacement of spare parts in class B happens more frequent. It is often strategic to decide the inventory amount of B spare parts and guarantee supplies. The spare parts in class C are always those that are used as part of normal consumption of scheduled maintenance that are ordered in regularly. They are neither expensive nor difficult to obtain when necessary. Spare parts in class D can be the daily consumption items that have little influence on the operation.



Figure 7. Hierarchical structure of the ship spare parts classification.

Code	X ₇	X ₆	X ₁	X ₂	X ₅	•••	X ₁₂	X ₁₃	X ₁₄
50502-08H-095	2	68.00	6	3	0.50		34.00	136.00	170.00
50502-08H-117	4	11,250.00	25	1	0.00		0.00	45,000.00	45,000.00
50502-08H-129	8	8.00	6	5	0.00		0.00	64.00	64.00
50502-08H-130	60	3.00	7	5	25.00		75.00	180.00	255.00
50502-08H-178	7	365.00	6	3	0.25		91.25	2555.00	2646.25
50502-08H-191	82	3.00	5	5	0.25		0.75	246.00	246.75
50502-08H-201	60	3.32	5	5	87.50		290.50	199.20	489.70
50502-08H-237	12	52.00	5	5	1.25		65.00	624.00	689.00
50502-08H-249	28	3.21	5	5	35.50		113.96	89.88	203.84
50501-11H-184	12	65.00	5	5	4.75		308.75	780.00	1088.75
50501-11H-196	24	8.07	5	5	9.00		72.63	193.68	266.31
50501-11H-206	24	9.50	5	5	2.50		23.75	228.00	251.75
50501-11H-219	21	6.16	5	5	12.25		75.46	129.36	204.82
50501-11H-243	12	219.58	5	5	16.50		3623.07	2634.96	6258.03
50501-11H-255	10	16.28	5	5	17.50		284.90	162.80	447.70
50501-11H-267	5	27.43	5	5	1.00		27.43	137.15	164.58
50501-11H-279	3	1317.33	7	3	0.50		658.67	3951.99	4610.66
50610-03H-092	15	9.92	6	5	8.75		86.80	148.80	235.60
50801-02H-111	46	3.00	6	5	6.00		18.00	138.00	156.00
50801-02H-028	2	50.00	6	5	0.25		12.50	100.00	112.50
						•••			•••

Table 2. The structure and distribution of the spare parts.

4.3. Hierarchical Classification Structure Image Conversion

To convert the spare parts classification into an image classification problem, 15 total variables involved in the hierarchical structure are represented by different colors according to their level status. Table 3 describes the color assignments corresponding to specific criteria.

An example of spare part No. 144 (Oil temperature control valve) is given in Table 4 and Figure 8. All of its criteria are colored according to their actual status, the part appears to fall into class B.

Citeria	Scale	Level	Color Assignment
	$0 \leq X_1 \leq 3$	Short	\bigcirc
X_1	$3 \leq X_1 \leq 7$	Medium	\bigcirc
(Day)	$7 \leq X_1 \leq 10$	Long	\bigcirc
	$X_1 \ge 10$	Extreme long	
	$0 \leq X_2 \leq 2$	Low	•
X ₂	$2 \leq X_2 \leq 6$	Medium	\bigcirc
	$X_2 \ge 6$	High	ightarrow
	1	Low	\bigcirc
X ₃	2	Medium	\bigcirc
	3	High	•
	1	Low	\bigcirc
X_4	2	Medium	\bigcirc
	3	High	•
	$0 \leq X_5 \leq 0.25$	Extreme low	ightarrow
Xr	$0.25 \leq X_5 \leq 0.75$	Low	\bigcirc
(Piece)	$0.75 \leq X_5 \leq 10$	Medium	\bigcirc
	$10 \leq X_5 \leq 50$	High	\bigcirc
	$X_5 \ge 50$	Extreme high	•
	$0 \le X_6 \le 25$	Extreme low	ightarrow
Xc	$25 \leq X_6 \leq 200$	Low	\bigcirc
(Yuan)	$200 \leq X_6 \leq 1000$	Medium	\bigcirc
	$1000 \le X_6 \le 50,000$	High	\bigcirc
	$X_6 \ge 50,000$	Extreme high	
	$0 \leq X_7 \leq 10$	Extreme low	ightarrow
X ₇	$10 \leq X_7 \leq 20$	Low	\bigcirc
(Piece)	$20 \leq X_7 \leq 60$	Medium	\bigcirc
	$X_7 \ge 60$	High	
	1	Low	lacksquare
X ₈	2	Medium	\bigcirc
	3	High	
	1	Low	
X9	2	Medium	\bigcirc
	3	High	

 Table 3. Level and color assignment of variables.

Citeria	Scale	Level	Color Assignment
	1	Low	\bigcirc
X ₁₀	2	Medium	\bigcirc
	3	High	•
	1	Low	\bigcirc
X ₁₁	2	Medium	\bigcirc
	3	High	•
	1	Low	\bigcirc
X ₁₂	2	Medium	\bigcirc
	3	High	•
	1	Low	\bigcirc
X ₁₃	2	Medium	\bigcirc
	3	High	•
	1	Low	\bigcirc
X ₁₄	2	Medium	\bigcirc
	3	High	•
		А	•
V		В	\bigcirc
Ŷ		С	\bigcirc
		D	\bigcirc

Table 3. Cont.



Figure 8. The hierarchical structure of spare part No. 144.

Criteria	x ₇	x ₆	x ₁	x ₃	x ₂	x ₅	x ₄	x ₈	X9	x ₁₀	x ₁₁	x ₁₂	x ₁₃	x ₁₄	Y
value grade	1 Extreme low	1304.76 high	6 medium	2 medium	3 medium	6 medium	2 medium	2 medium	2 medium	2 medium	2 medium	7828.56 high	1304.76 high	9133.32 high	В
color	\bigcirc	\bigcirc	\bigcirc	\bigcirc	\bigcirc	\bigcirc	\bigcirc	\bigcirc	\bigcirc	\bigcirc	\bigcirc				\bigcirc

Table 4. The original data information of spare part No. 144.

4.4. Supervised Learning Analysis Based on Proposed Method

After the variables are assigned with different colors, the hierarchical structure could be updated and saved as images. In the first phrase, the learning process is conducted under a supervised environment, and the model is designed to be trained by two different datasets to determine how the sample quantity affects the results.

4.4.1. Classification Results Analysis with Few-Shot Samples

In the primary learning process, 189 labeled spare parts from the auxiliary system are selected: four from class A, 26 from class B, 43 from class C, and 116 from class D. The dataset is divided into training and testing sets.

All the spare parts data are transferred into colored hierarchical structure and saved as images. Subsequently, all the images are processed in the modified AlexNet model as discussed in Section 3.2. The modified AlexNet structure has five convolution layers and three fully connected layers. Every image is 227×227 in size; the ReLU activations ensure that no outputs after the convolution or fully connected layers have overfitting problems. Once the output is calculated and output by the last fully connected layer, a Softmax layer determines the final category of the image. The final classification results in this case split the images into four different classes: A, B, C, and D.

For the 189 images in this step, the first random 100 images are trained using the modified AlexNet, as shown in Figure 9. Some of the classification results are shown in Figure 8. It is indicated that the four results are D, B, C, D respectively. It is obvious that the listed four spare parts images are all classified with an accuracy of 100%. Additionally, for the overall training result, the average accuracy of the modified model reaches 99.8%, which can fully satisfy the practical demand.



Figure 9. Partial training results of initial dataset. (**a**,**d**) indicate these two spare parts are 100% classified into class D; (**b**) shows this spare part is 100% classified into class B; (**c**) shows this spare part is 100% classified into class C.

Then, the constructed model is also used to test the remaining 89 images, and the commonly used indicator recall ratio r and precision ratio p are adopted to evaluate its performance:

$$r = \frac{TP}{TP + FN} \tag{8}$$

$$p = \frac{IP}{TP + FP} \tag{9}$$

where TP is a true positive classified correctly as positive, FN is a false negative misclassified as false, and FP is a false positive misclassified as positive. The test accuracy of each category is determined here as follows:

TΠ

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$
(10)

where *TN* is a true negative classified correctly as negative.

The classified results for *TP*, *FP*, *TN*, and *FN* are shown in Table 5, and the results of *r* and *p* for the testing are shown in Table 6.

Class	ТР	FP	TN	FN	Accuracy
А	1	1	0	0	50%
В	13	1	0	1	86.7%
С	21	1	0	1	87.5%
D	47	1	0	0	97.9%

Table 5. The classified results for TP, FP, TN, and FN.

Table 6. The results of *r* and *p* for the testing.

Class	Α	В	С	D
r	100%	92.9%	95.5%	100%
p	50%	92.9%	95.5%	97.9%

For practical application in the spare parts classification, as shown in Table 5, there is one sample in class A that is classified as false positive. It is the true spare part No. 23 that is easily classified to be A because both its unit price and annual inventory cost are high. However, in the actual management, this item is classified into B class since it is in high monitoring and the supply reliability is high.

As shown in Table 5, the overall accuracy for B, C and D is 86.7%, 87.5% and 97.9%, respectively. However, the accuracy for class A is merely 50%, since there are only two samples in the testing set. However, when the results are analyzed with precision and recall, it indicates the modified AlexNet structure appears to perform well. Especially for class A, the recall rate is 100%, although the precision ratio is only 50%. It is obvious that *p* and *r* metrics contradict each other. One of the main reasons is that class A has the fewest number of training samples, which may cause an under-fitting problem in the machine learning models. Thus, for the unbalanced dataset of class A, we explore the F-measure value to assess its performance:

$$F1 = \frac{2pr}{p+r} \tag{11}$$

For class A, the *F*1 value is 0.667, which can still prove the effectiveness of the model. The characteristic of the result is consistent with the fact that category A are always the most critical, expensive, and highly reliable spare parts; accordingly, these parts have limited stock, which complicates the training results in this test.

4.4.2. Re-Classified Results Analysis with Additional Samples

As mentioned above, few-shot problems emerged in the training process. Meanwhile, to consider and research the model learning ability, the samples are extended to 804 spare parts of the whole ship to cover a sufficient number of class A parts to verify the model. For the data distribution, there are 70 spare parts from class A, 122 spare parts from class B, 174 spare parts from class C, and 438 spare parts from class D. The whole spare parts are divided into two different datasets; the first 450 parts are regarded as the training set and the remaining parts as the test set. The overall training accuracy is 97% in this case. A portion of the training results are shown in Figure 10.



Figure 10. Partial training results of additional samples. (**a**) shows this spare part is 100% classified into class C; (**b**,**d**) indicate these two spare parts are 100% classified into class B; (**c**) shows this spare part is 100% classified into class A.

The training results are acceptable. An additional test model is built using the remaining spare parts dataset. The classification results for *TP*, *FP*, *TN*, and *FN* are shown in Table 7.

Class	ТР	FP	TN	FN	Accuracy
А	39	0	0	1	97.5%
В	75	0	0	0	100%
С	115	0	0	0	100%
D	220	0	0	0	100%

Table 7. Re-classified results for *TP*, *FP*, *TN*, and *FN*.

The accuracy of class A has been improved to 97.5% after these additional samples are added, while class B, C, and D maintained 100% accuracy. The G-DCNN algorithm appears to yield ideal classification results when it is applied to labeled samples.

4.5. Semi-Supervised Learning Analysis Based on Proposed Method

However, it is difficult to achieve the exact level descriptions of spare parts in practice. A semi-supervised learning analysis technique is developed based on the proposed method to verify the generality of the rules. The final color boxes of the structure are removed from the level judgment in the hierarchy to produce un-labeled samples, as shown in Figure 11.



The remaining variables are still used to test the effectiveness of the modified AlexNet algorithm in this case.

Figure 11. Un-labeled structure.

A total of 324 images are selected from all the 804 spare parts of the whole ship with a random ratio of labeled samples for further verification. A portion of the training results are shown in Figure 12. The overall training accuracy is 95.2% in this case, which is still acceptable. Then, the selected spare parts are used to test, and the test results for *TP*, *FP*, *TN*, and *FN* in the semi-supervised learning process are shown in Table 5.



Figure 12. Partial training results of the un-labeled structure classification. (**a**) shows this spare part is 98.4% classified into class B; (**b**,**d**) indicate these two spare parts are 100% classified into class D; (**c**) shows this spare part is 100% classified into class C.As shown in Table 8, all four spare part levels have acceptable accuracy. For class A, the testing accuracy is 89.7% (higher than 50% in the supervised learning process) with the few-shot samples. This is slightly lower than the 97.5% achieved in the supervised learning process with additional samples, but still confirms that the proposed method is effective. The *r* and *p* results of this semi-supervised process are shown in Figure 13.

	ТР	FP	TN	FN	Accuracy
А	26	3	0	0	89.7%
В	42	0	0	2	95.5%
С	70	0	0	0	100%
D	179	2	0	0	98.9%

Table 8. Semi-supervised learning results.



Figure 13. Results of *r* and *p* for semi-supervised learning.

Figure 13 shows that with an increasing number of class A spare parts, the precision ratio increases to a very high level. Even without the final color box for the level judgment, the spare parts of different levels still have high *r* and *p*. Compared to supervised learning, the non-supervised learning shows lower precision for certain spare parts but the error is still within an acceptable range. The non-supervised results, in other words, do also demonstrate the effectiveness of the proposed method.

4.6. Comparative Study

We further compared the proposed method to other traditional classification methods including SVM, BP Neural Network (BPNN), and K-Nearest Neighbor (KNN) using the same dataset and accuracy indicator.

The SVM parameter with the cost coefficient and kernel function coefficient is 1 and $1/n_{feature}$ initially, where $n_{feature}$ is the number of the features. The BPNN has one input layer, one hidden layer, and one output layer. The numbers of input neurons, hidden neurons, and the output neurons are $N \times D$, 1*K*, and $N \times 4$ respectively, wherein *N* represents the number of the images and *D* is the characteristics of the image. There are 10 neighbors in the KNN. Table 6 shows detailed information gathered in this comparative analysis, where the modified AlexNet structure shows the best average accuracy.

Table 9 shows that the average accuracy of the modified AlexNet structure for spare part classes A, B, C, and D is 0.90, 0.95, 1.0, and 0.99 respectively; the proposed method outperforms all other methods tested here in terms of average accuracy. The SVM algorithm produced relatively close results at 0.88, 0.90, 0.84, and 0.98 for classes A, B, C, and D, respectively. As discussed above, class A had the fewest samples while class C and D had more samples. However, the performance of the modified AlexNet structure is consistent for all four classes. This suggests that the proposed approach is effective regardless of the number of available datasets.

		Т	Р			-	FP			Т	'N			F	'N			Accu	iracy	
Method	Α	В	С	D	Α	В	С	D	Α	В	С	D	Α	В	С	D	Α	В	С	D
AlexNet	26	42	70	179	3	0	0	2	0	0	0	0	0	2	0	0	0.90	0.95	1	0.99
SVM	26	44	59	177	3	0	0	4	0	0	0	0	0	0	11	0	0.88	0.90	0.84	0.98
BPNN	25	39	62	167	3	0	0	9	0	0	0	0	1	5	8	5	0.86	0.89	0.89	0.92
KNN	25	40	66	171	3	0	0	8	0	0	0	0	1	4	4	2	0.86	0.91	0.94	0.94

Table 9. Comparative analysis of average accuracy.

Neither SVM nor KNN is an effective tool for practical spare parts management because of their "black boxes"; they cannot offer understandable explanations to decisionmakers regarding the rule's assignment of criteria. The proposed method, conversely, makes the extracted criteria and hierarchical structure explicable and modifiable according to the expert knowledge or feedback based on the results. In the learning phase, the proposed method changes the traditional classification problem into an image classification problem which is helpful to the data visualization. It also outranks other methods tested here in terms of its accuracy and calculation speed.

5. Conclusions

For accurate spare parts classification, advanced condition monitoring technology should be developed and applied to the critical components, which will benefit a reliable maintenance logistics system. Moreover, this image identification method will also be considered to use for operation dashboard monitoring and fault diagnosis.

The graph-based deep convolutional neural network (G-DCNN) method developed in this study transforms the traditional classification problem into an image identification problem, realizing the visible and accurate classification of spare parts. The hierarchical classification structure is constructed to take unlimited criteria into consideration, which will be adjusted according to the practical demand. Additionally, the causal relationship between the criteria can be explained well through the hierarchical structure. Due to the transfer learning thought, AlexNet is modified to solve the spare parts classification problem by its excellent image identification ability. Both ReLu function and Dropout are adopted to eliminate overfitting problems under the condition of a two-GPU deep network. Through the case study results analysis, we can see a well performed model by cross validation and comparison. In particular, the influence of dataset amount to the learning model is obvious by A-class spare parts.

However, there are some problems unelaborated in this study that will be continued in the future. Regarding to the hierarchical structure, we intend to propose a quantitative approach that can deal with the causal relationship and status level division. With respect to the AlexNet application in spare parts classification, we may try to simplify its structure for more convenience in understanding and calculation in the future.

Author Contributions: Conceptualization, K.Y.; methodology, K.Y.; software, K.Y. and Y.W.; validation, K.Y. and Y.W.; formal analysis, K.Y.; data curation, K.Y.; writing—original draft preparation, K.Y.; writing—review and editing, K.Y.; visualization, K.Y.; supervision, S.F., and A.M.; funding acquisition, S.F. All authors have read and agreed to the published version of the manuscript.

Funding: This research is funded by National Natural Science Foundation of China under Grant: 51679178 and 52071240.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors would like to acknowledge the financial support of the first Author from China Scholarship Council (CSC) and B. John Garrick Institute for the Risk Science, University of California, Los Angeles.

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

References

- 1. Hu, Q.; Chakhar, S.; Siraj, S.; Labib, A. Spare parts classification in industrial manufacturing using the dominance-based rough set approach. *Eur. J. Oper. Res.* 2017, 262, 1136–1163. [CrossRef]
- 2. Xie, W.; Liao, H.; Jin, T. Maximizing system availability through joint decision on component redundancy and spares inventory. *Eur. J. Oper. Res.* **2014**, 237, 164–176. [CrossRef]
- Costantino, F.; Di Gravio, G.; Tronci, M. Multi-echelon, multi-indenture spare parts inventory control subject to system availability and budget constraints. *Reliab. Eng. Syst. Saf.* 2013, 119, 95–101. [CrossRef]
- Molenaers, A.; Baets, H.; Pintelon, L.; Waeyenbergh, G. Criticality classification of spare parts: A case study. *Int. J. Prod. Econ.* 2012, 140, 570–578. [CrossRef]
- Teixeira, C.; Lopes, I.; Figueiredo, M. Multi-criteria Classification for Spare Parts Management: A Case Study. *Procedia Manuf.* 2017, 11, 1560–1567. [CrossRef]
- Lengu, D.; Syntetos, A.; Babai, M. Spare parts management: Linking distributional assumptions to demand classification. *Eur. J.* Oper. Res. 2014, 235, 624–635. [CrossRef]
- Looy, G.; Paul, M. An inventory policy for slow and fast movers in a petrochemical plant: A case study. J. Oper. Res. Soc. 1978, 29, 867–874.
- Trafalis, T.; Ince, H. Support Vector Machine for Regression and Applications to Financial Forecasting. In Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks. IJCNN 2000. Neural Computing: New Challenges and Perspectives for the New Millennium, Como, Italy, 24–27 July 2000; Volume 6, pp. 348–353.
- 9. Duchessi, P.; Belardo, S.; Seagle, J.P. Artificial Intelligence and the Management Science Practitioner: Knowledge Enhancements to a Decision Support System for Vehicle Routing. *Interfaces* **1988**, *18*, 85–93. [CrossRef]
- 10. Kraljic, P. Purchasing must become supply management. Harv. Bus. Rev. 1983, 61, 109–117.
- 11. Braglia, M.; Grassi, A.; Montanari, R. Multi-attribute classification method for spare parts inventory management. *J. Qual. Maint. Eng.* **2004**, *10*, 55–65. [CrossRef]
- 12. Cakir, O.; Canbolat, M.S. A web-based decision support system for multi-criteria inventory classification using fuzzy AHP methodology. *Expert Syst. Appl.* **2008**, *35*, 1367–1378. [CrossRef]
- 13. Zeng, Y.; Wang, L.; He, J. A Novel Approach for Evaluating Control Criticality of Spare Parts Using Fuzzy Comprehensive Evaluation and GRA. *Int. J. Fuzzy Syst.* **2012**, *14*, 392–401.
- 14. Kabir, G.; Hasin, M.A.A. Multiple criteria inventory classification using fuzzy analytic hierarchy process. *Int. J. Ind. Eng. Comput.* **2012**, *3*, 123–132. [CrossRef]
- 15. Ng, W.L. A simple classifier for multiple criteria ABC analysis. Eur. J. Oper. Res. 2007, 177, 344–353. [CrossRef]
- 16. Hadi-Vencheh, A. An improvement to multiple criteria ABC inventory classification. *Eur. J. Oper. Res.* **2010**, 201, 962–965. [CrossRef]
- 17. Yu, M.-C. Multi-criteria ABC analysis using artificial-intelligence-based classification techniques. *Expert Syst. Appl.* **2011**, *38*, 3416–3421. [CrossRef]
- 18. Partovi, F.Y.; Anandarajan, M. Classifying inventory using an artificial neural network approach. *Comput. Ind. Eng.* **2002**, *41*, 389–404. [CrossRef]
- Cai, H.; Yu, T.; Xia, C. Quality-Oriented Classification of Aircraft Material Based on SVM. *Math. Probl. Eng.* 2014, 2014, 273906. [CrossRef]
- 20. Liu, X.; Qiu, L.; Wang, W. Spare parts inventory management in special steel enterprise. *Comput. Integr. Manuf. Syst.* 2007, 13, 1756–1761.
- 21. Guo, G.; Wang, H.; Bell, D.; Bi, Y.; Greer, K. KNN Model-Based Approach in Classification; Springer: Berlin/Heidelberg, Germany, 2003; pp. 986–996. [CrossRef]
- 22. Hu, Q.; Wang, Y.; Zhou, Y.; Zhao, H.; Qian, Y.; Liang, J. Review on hierarchical learning methods for large-scale classification task. *Sci. Sin. Inf.* **2018**, *48*, 487–500. [CrossRef]
- 23. Bacchetti, A.; Plebani, F.; Saccani, N.; Syntetos, A. Spare Parts Classification and Inventory Management: A Case Study. In Proceedings of the International Symposium on Inventories, Budapest, Hungary, 23–27 August 2010. [CrossRef]
- 24. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. *Commun. ACM* 2017, 60, 84–90. [CrossRef]
- Chen, Y.; Zhu, K.; Zhu, L.; He, X.; Ghamisi, P.; Benediktsson, J.A. Automatic Design of Convolutional Neural Network for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote. Sens.* 2019, 57, 7048–7066. [CrossRef]
- Fang, L.; Wang, C.; Li, S.; Rabbani, H.; Chen, X.; Liu, Z. Attention to Lesion: Lesion-Aware Convolutional Neural Network for Retinal Optical Coherence Tomography Image Classification. *IEEE Trans. Med. Imaging* 2019, 38, 1959–1970. [CrossRef]
- Wang, Y.; Zhang, Y.; Wu, Z.; Li, H.; Christofidesc, P.C. Operational Trend Prediction and Classification for Chemical Processes: A Novel Convo-lutional Neural Network Method Based on Symbolic Hierarchical Clustering. *Chem. Eng. Sci.* 2020, 225, 115796. [CrossRef]
- Ren, Y.M.; Zhang, Y.; Ding, Y.; Wang, Y.; Christofides, P.D. Computational fluid dynamics-based in-situ sensor analytics of direct metal laser solidification process using machine learning. *Comput. Chem. Eng.* 2020, 143, 107069. [CrossRef]

- 29. Szegedy, C.; Ioffe, S.; Vanhoucke, V.; Alemi, A. Inception-v4, inception-resnet and the impact of residual connections on learning. In Proceedings of the AAAI Conference on Artificial Intelligence, Phoenix, AZ, USA, 12–17 February 2016.
- 30. Iandola, F.; Moskewicz, M.; Karayev, S.; Girshick, R.; Darrel, T.; Keutzer, K. DenseNet: Implementing Efficient ConvNet Descriptor Pyramids. *arXiv* 2014, arXiv:1404.1869.
- Wang, S.; Xie, S.; Chen, X.; Guttery, D.; Tang, C.; Sun, J.; Zhang, Y.-D. Alcoholism Identification Based on an AlexNet Transfer Learning Model. Front. Psychiatry 2019, 10, 205. [CrossRef] [PubMed]
- 32. Alex, K.; Ilya, S.; Geoffrey, E. ImageNet Classification with Deep Convolutional Neural Networks. *Adv. Neural Inform. Process. Syst.* **2012**, 25, 1097–1105.
- 33. Pan, S.; Yang, Q. A survey on transfer learning. IEEE Trans. Knowl. Data Eng. 2009, 22, 1345–1359. [CrossRef]