Research Article

Research on Travel Reimbursement Behavior Management Based on Deep Learning in Financial Sharing Mode

Qian Zhang [1] **and Bowei Feng** [2]

[1] *Jilin International Studies University, Jilin Changchun 130117, China*
[2] *State University of Information Technology, Mechanics and Optics, St. Petersburg 197376, Russia*

Correspondence should be addressed to Qian Zhang; zhangqian@jisu.edu.cn

The standardization, transformation and upgrading of financial management plays an important supporting role in promoting the standardized management and healthy operation of corporate expense reimbursement behaviors. This paper starts with the behavioral portrait of enterprise personnel travel expense reimbursement. Based on the problems of lengthy process and complex financial accounting in most reimbursement behaviors at this stage, an efficient and efficient expense reimbursement processing method is proposed, that is, reimbursement through collection Information images, using the convolutional neural network algorithm under deep learning, to detect target images and identify valid information, and rely on the financial shared service center's screening function of expense reimbursement behavior portraits under big data, the integrity of expense reimbursement data. Compliance is judged, thus forming a complete set of standardized and procedural personal self-service reimbursement model, which optimizes the management of enterprise expense reimbursement behavior, avoids subsequent losses and resource waste caused by human risks, and improves enterprise operation efficiency and expenses. Control has certain value and reference significance.

1. Introduction

In order to adapt to the changing requirements of the development strategy and management system of the enterprise, further promote the healthy and rapid development and transformation and upgrading of the enterprise, promote the standardization of financial management, and a virtuous circle of standardized management of expense reimbursement behavior. The first task is to diagnose the problem of employee reimbursement behavior, that is, reimbursement content and expense data, to ensure the compliance and effectiveness of reimbursement behavior. The data terminal based on the enterprise financial sharing center has the information identification function of the user's consumption operation behavior, which can realize the transformation of enterprise expense reimbursement to intelligent management and control and value creation; the construction of risk prediction system to highly controllable,

full-process supervision and early warning will become a Focus on the focus, so how to manage the expense reimbursement behavior from user reimbursement to reimbursement is the focus of this paper. With the rapid development of computer vision technology and deep learning technology, more and more work is done by humans instead of by computers. Based on the current situation of travel reimbursement behavior, this paper will use the convolutional neural network image processing method under deep learning to carry out business travel through the functions of whole-process accounting, centralized standardized supervision and real-time data support provided by the Financial Shared Service Center. Research on the detection and identification of reimbursement picture information, and research on the compliance and validity of the reimbursement content list and expense data, trying to solve the problems of manual submission of expense reimbursement vouchers and the tedious verification of ledger

information by financial personnel, and formed a complete set of standardized research. And programmed personal self-service reimbursement mode, in order to achieve standardized and efficient reimbursement behavior management of user expense reimbursement behavior.

2. Related Work

With the advancement of science and technology and the development trend of business and financial integration, the concept of "financial sharing" was born. In order to achieve the goal of improving work efficiency, reducing operating costs, and improving service quality, the Financial Shared Service Center centralizes the processing of various scattered operating functions. Realize the centralized analysis of financial-related information of each business unit of the group enterprise, which greatly promotes the efficiency of corporate decision-making [1].

As one of the most common content of financial shared service application scenarios, travel reimbursement business has attracted the attention of academics and practical scholars. Cloud computing, OCR, electronic reimbursement information, etc. have brought development opportunities for expense reimbursement business, helping companies realize the goals of electronic bills, online approval, intelligent auditing, and automation of fund payment. As a key data source of accounting information, it is the beginning of corporate accounting. Link [2–5].

From the current situation and needs of travel reimbursement business, Guo Fang discussed the issue of expense reimbursement in the traditional branch management of group companies, focusing on the development of centralized reimbursement process organization, internal approval mechanism and information system construction [6–10]. Xing Ying analyzes the key issues that arise in the multiple business scenarios of financial reimbursement, and establishes a compliance control design through the discussion of the expense reimbursement system and reimbursement process [8]. With the gradual improvement of informatization construction, online reimbursement and reimbursement systems have been developed and applied successively to reshape the original reimbursement nodes and improve the efficiency of reimbursement and the level of financial management. However, in practice, online reimbursement still has a large amount of reimbursement documents, various types, low comprehensive level of reimbursement staff, and misconduct of reimbursement, which are the pain points under the current status of the reimbursement business market [11].

From the perspective of the reform and development trend of travel reimbursement, Qian Yijun and others believe that online reimbursement is a technical iteration of traditional reimbursement business from the perspective of mobile office needs. The three parties are directly connected to each other to realize the automatic processing of accounting and payment [12–15]. Luo Xiangming further mentioned that online expense reimbursement is operated through the network. Employees and leaders of an enterprise can interact with financial reimbursement in a modern office without being restricted by time, space and methods. After the original vouchers are verified, they can pass directly. Presigned banks conduct online payment settlement [16].

Expense reimbursement behavior management is the core content of enterprise reuse, and the formulation of individualized plans based on the classification of problem behaviors has high practical and practical significance. The optimization implementation involves intelligent budget planning and credit system construction, as well as the optimization concept of intelligent declaration path. The classification method of early warning mechanism based on behavioral characteristics and the level management based on scoring method are embedded, and financial intelligent information technology such as voice print is added to the declaration. This makes up for the lack of artificial intelligence technology research in the field of accounting. Focusing on business embedded technology, deepening joint multidisciplinary research will be an important direction in the future. Based on the perspective of deep learning, this paper will discuss the research on information identification and detection in travel reimbursement under the financial sharing model, and provide new ideas and directions for the in-depth development of standardized reimbursement behavior management and embedded technology.

3. Related Theories

3.1. Financial Sharing. Financial sharing is actually a process of reengineering and standardizing the financial business that is scattered in various projects, has high operational repeatability, and is easy to standardize. A financial management model that improves efficiency, reduces costs, allocates resources rationally, and increases the real-time and accuracy of information transmission. The financial shared service center is to comprehensively transform the traditional financial operation mode. The core principle is "professional stratification, business integration", relying on the shared service platform to create the integration of shared finance and business finance.

In terms of functional composition, the financial shared service center is supported by whole-process accounting, centralized standardized supervision and real-time data support. In addition, as a data support center, it can take advantage of its information advantages to provide enterprises with comprehensive, multi-dimensional and real-time accounting information services such as accounting and accounting reports. need. In the guarantee system supported by the financial shared service center, its core business platforms mainly include ERP system and SAP system, its auxiliary business platform includes the whole-process online reimbursement and payment system, and the information management system includes performance management system, expense management and control system, IT audit system, expense budget management system, etc. For the differences in business processing methods, the financial shared service center can provide a stable underlying interactive system and build a complete set of standardized systems, which is of great importance to realize the unified and centralized management of the financial
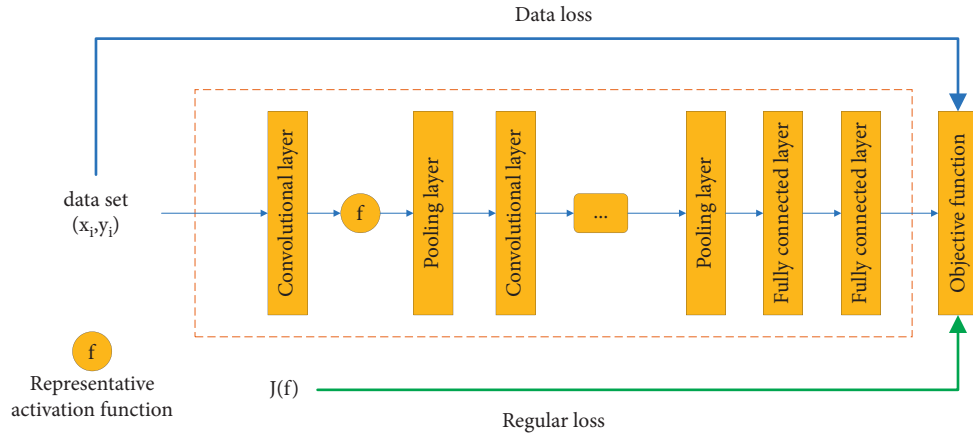
Figure 1: Convolutional neural network structure.

shared service business and ensure the compliance operation of various businesses. effect.

3.2. Convolutional Neural Network.

Among the deep learning techniques that have emerged in recent years, the most widely used convolutional neural network is undoubtedly the convolutional neural network [17–21]. Convolutional neural network is a hierarchical model. It is mainly composed of an input layer, a convolutional layer, an activation function, a pooling layer, and a fully connected layer. Its input is similar to the original data of the image. Its feedforward operation is a process of acquiring and abstracting feature information layer by layer from the input layer through operations such as convolution, pooling, and nonlinear activation function mapping. Among them, different operations generally correspond to their related layers one-to-one. For example, the convolution operation corresponds to the convolution layer and so on [22–24]. Finally, its last layer converts target tasks such as classification and regression into objective functions, and uses back-propagation algorithm to feed back and update the difference between the predicted value and the true value from the last layer. Figure 1 shows the structure of the convolutional neural network.

3.2.1. Input Layer.

The input layer is the beginning of the network model and is mainly used to input the data required for training. When the image is operated, it is generally stored and calculated in the form of a matrix for the color input image. Therefore, for a black and white input image, the input to the network is a two-dimensional matrix; like the number of channels of a color image, each channel corresponds to a two-dimensional matrix, so the input network is three two-dimensional matrices.

3.2.2. Design of DNQ Algorithm Structure Based on Pointer Network.

In a convolutional neural network, each neuron in the hidden layer can be regarded as a convolution kernel, and each convolution kernel will perform a sliding convolution operation on the image. The convolution kernel is used to extract the features of the image, thanks to its sparse connection and weight sharing.

(1) Sparse Connection. For a general artificial neural network, neurons in adjacent layers are closely connected. That is to say, every neuron in the current layer maintains a connection with all neurons in the front and rear layers, and each connection means a weight that needs to be learned and saved. This will lead to when the network model has a deeper number of layers. The network is difficult to train and converge, and consumes a lot of memory, which makes it difficult to complete the task.

In a convolutional neural network, each neuron only needs to save the weight in its own structure. For example, for a convolution kernel with a size of 3×3, it only needs to store its own 9 weights. This means that the number of connections between each neuron and the previous layer is only related to its own size. Compared with the full connection, the convolution kernel has few connections, so it is called sparse connection.

(2) Weight Sharing. It can be seen from the sparse connection that each convolution kernel only stores a small amount of weights. For the same convolution kernel, in each round of training, the weights used in sliding convolution are the same. Only when one iteration is completed, will its internal weights be updated.

Therefore, for the same convolution kernel, in the same round of iteration, the weight of each convolution is constant, so it is called weight sharing.

The size of the image after the convolution operation is related to factors such as the size of the convolution kernel, the step size, and the pooling size. The specific formula is

$$O = \frac{I - F + 2P}{S} + 1. \tag{1}$$

Among them, I and O represent the size of the input and output respectively, F is the size of the convolution kernel, and P and S represent the size of the filled edge pixel and the convolution step size, respectively.

The size of the convolution kernel determines the size of the receptive field obtained after each convolution operation. The filled edge pixels can prevent the size of the image from becoming smaller and smaller due to the convolution operation and can avoid the loss of boundary information. Convolution step size The setting of directly causes the image size to change, and can also reduce the amount of calculation.

3.2.3. Pooling Layer. Usually several consecutive convolutional layers are used to extract more features, but this also means a large amount of calculation and parameters. Therefore, in order to reduce the amount of calculation and compress the image feature map, a pooling layer is generally added in the middle of the continuous convolutional layer. The operation of the pooling layer is very similar to that of the convolutional layer. Without filling, you only need to adjust the size of the convolution kernel to 2 in the convolution formula to realize that the size of the output image is half the size of the input image. According to different needs, there are two main operations of the pooling layer, namely maximum pooling and average pooling.

(1) Maximum pooling. Taking the case where the size of the pooling layer is 2×2 as an example, the maximum pooling is to take the largest pixel value in the 2×2 area as the pixel value of the area, and at the same time the size of the area changes from 2×2 to lxl.

(2) Average pooling. Similarly, taking the case where the size of the pooling layer is 2×2 as an example, the average pooling is to use the average value of all pixel values in a 2×2 area as the pixel value of the area, and the size of the area is changed from 2×2 to lxl.

It can be seen from the two pooling methods that the maximum pooling retains the maximum pixel value in the region, which is the texture feature in the image, while the average pooling retains the average value of the pixel values in the region, which is integral, Which is the overall characteristics of the image.

3.2.4. Active Layer. The essence of convolutional neural network training is to make the model have a good fit to the data, and at the same time have a good generalization ability. The convolution operation is essentially a linear operation. In order to make the model have better expressive ability, it is often necessary to add a certain nonlinearity, that is, add an activation layer after the convolution layer.

The activation layer structure is relatively simple, generally just an activation function, used to add nonlinearity to the output result of the convolutional layer. Commonly used activation functions include Sigmoid function, Tanh function and ReLU function.

(1) Sigmoid Function. The Sigmoid function is a typical S-shaped curve, and its function and derivative expressions are respectively

$$S(x) = \frac{1}{1 + e^{-x}},$$
$$S(x) = \frac{e^{-x}}{(1 + e^{-x})^2} = S(x)(1 - S(x)). \tag{2}$$

From the Sigmoid function and its derivative, it can be found that the output result of the function is in. It is monotonic and has central symmetry, so it is very suitable as an activation function.

The characteristics of the sigmoid function are very suitable for use in binary classification problems, but because of the large amount of calculation of the function and the slow change of the derivative at both ends of the function, the gradient disappearance phenomenon is prone to occur during back propagation, which makes the deep model difficult train.

(2) Tanh Function. The Tanh function is the hyperbolic tangent function, and its function and derivative expressions are

$$T(x) = \frac{e^x - e^{-x}}{e^x - e^{-x}},$$
$$T(x) = 1 - T^2(x). \tag{3}$$

It can be found from the Tanh function and its derivative that it is very similar to the Sigmo id form, and the function image is very similar. It can be further found that the conversion relationship is

$$T(x) = 2.S(2x) - 1. \tag{4}$$

(3) ReLU Function. The ReLU function is the modified linear unit, which has the simplest function expression form, and its function and derivative expressions are

$$R(x) = \max(0, x),$$
$$R(x) = \begin{cases} 0 \; x < 0 \\ 1 \; x \geq 0 \end{cases} \tag{5}$$

From the ReLU function and its derivative, it can be found that the output of the ReLU function is 0 on the left side of the vertical axis, and the input itself is on the right side. It is only a function that takes the input and 0 to the maximum value, and the amount of calculation is very small. Observing the derivative of the ReLU function at the same time, it can be found that the derivative form is simpler, with a piecewise function with 0 as the boundary, and each segment is a constant, and the amount of calculation in the back propagation is also very small.

As shown in Figure 2, compared with the Sigmoid function and the Tanh function, it can be found that the ReLU function has a small amount of calculation, so that the model can converge faster; and the output on the left side of the vertical axis is 0, which means that it can play a certain role on the neuron. The most important thing is that the derivative of the ReLU function has nothing to do with the
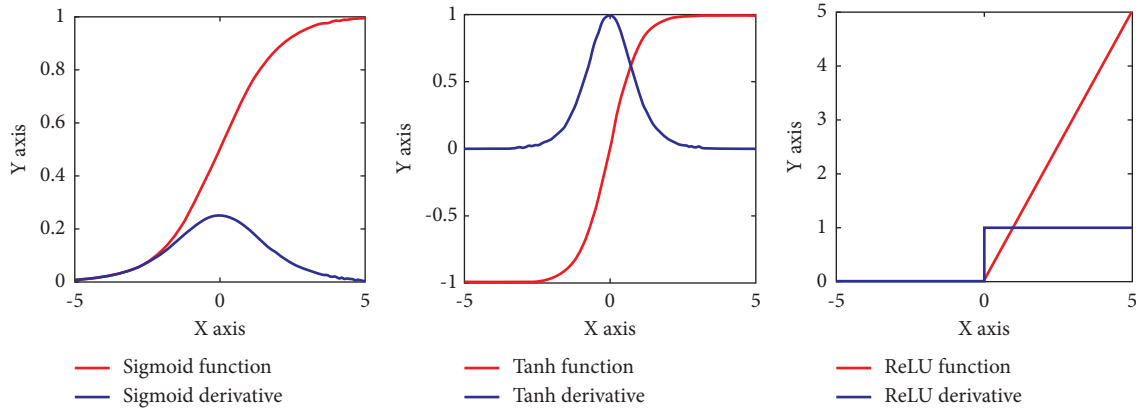
FIGURE 2: Comparison of three activation functions and their derivatives.

input and is a constant, thus eliminating the phenomenon of the disappearance of the gradient.

3.2.5. Fully Connected Layer.

The function of the convolutional layer, the pooling layer and the activation layer can be seen as extracting features of the input image, that is, mapping the input original data into its feature space, and the function of the fully connected layer is to use the extracted features for classification tasks. That is, the feature is mapped to the label space. In other words, the function of the fully connected layer is to integrate the local features extracted from the previous layers to obtain overall feature information.

The operation performed by the fully connected layer is also very simple, that is, the basic matrix operation, and the result obtained is a one-dimensional vector. But because of its fully-connected nature, it often needs to store a lot of parameters and consumes a lot of memory.

In order to reduce the memory, considering the advantages of the convolutional layer, the convolutional layer can be used to replace the fully connected layer, thereby reducing the training difficulty and memory of the network model.

4. Construction of a Travel and Reimbursement Information Detection and Recognition Model Based on Convolutional Networks under Financial Sharing

4.1. The Overall Process of Travel Reimbursement under Financial Sharing.

Electronic reimbursement information and paper reimbursement information. For electronic reimbursement information, due to its relatively regular form, uniform templates and printing standards, the information can be extracted and identified based on its location information using traditional methods. As for paper-based reimbursement information, with the massive use of mobile devices, more and more reimbursement information images are captured by mobile devices such as smart phones. Due to the different shooting habits and the shooting quality of mobile devices, the reimbursement information images

taken by them are difficult to quickly identify with traditional methods. This section designs a method based on deep learning. In the reimbursement behavior, the reimbursement information images taken by smartphones, Perform target detection and information identification, synchronize the integrity of the identified information and data to the financial sharing service center for file integrity and data compliance determination, and through cost control, the final termination of the reimbursement behavior is completed. The overall process is shown in Figure 3:

The process framework is based on the identification of information for travel reimbursement, and then extends to the normative guidance of the financial reimbursement process, the three major management requirements of the discrimination of fraud and risk early warning management, and the support of deep learning algorithms under the big data platform. Provide intelligent technical solutions. Based on the difficulty in distinguishing the types of expense reimbursement behaviors faced by enterprises, and the difficulty in identifying and preventing fraudulent behaviors and operational risks, the reimbursement behavior management life cycle links are optimized and designed from the source.

4.2. Creation of Data Set of Travel and Reimbursement Information under Financial Sharing.

The method based on deep learning is a kind of machine learning method. The essence is to let the model fit a certain data set and complete the corresponding task through its generalization ability. The convolutional neural network model used in this paper also needs to be trained to fit the travel and reimbursement data set. Therefore, the quality of the data set determines the upper limit of the task completion quality, and the training model is only constantly approaching this upper limit. In order to obtain enough travel reimbursement data image information and compliant travel reimbursement list information, use crawler technology, use keywords in Google pictures to obtain corresponding travel reimbursement information images and extract them according to the standardization of the Financial Shared Service Center The processed travel reimbursement list data information image. Because the amount of data is relatively small compared to the model, it causes overfitting problems. Therefore, it is
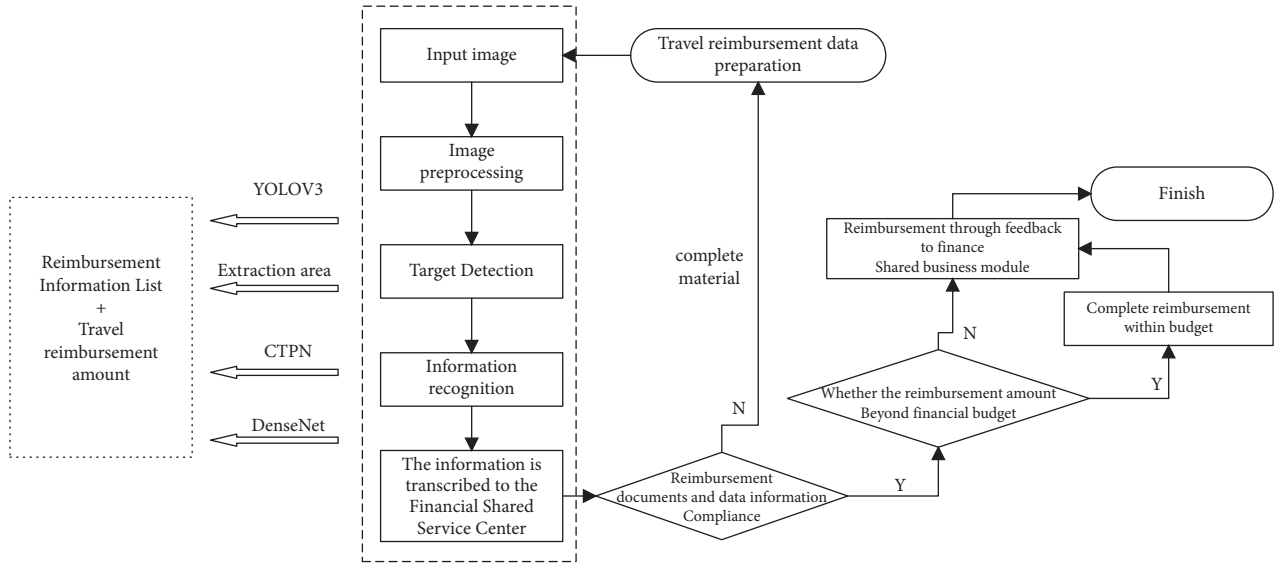
FIGURE 3: Overall flow chart of travel reimbursement under financial sharing.

necessary to perform data enhancement on the extracted image information, that is, use traditional image processing methods to expand the image. Since the features extracted by the convolutional neural network have nothing to do with the position, this section again uses Gaussian noise, horizontal and vertical flipping and rotation methods to expand the data set. After the data set is expanded, because it is a problem of target detection, different areas of the reimbursement information data set need to be labeled differently, that is, labeling. LabelImage software is used to label the reimbursement information images. After the data set is fully labeled, each piece Images are generated in the corresponding XML tag file, which contains the image size, the position information of the four rectangular boxes and their tag names.

There are a total of 1250 images in the data set. First, the data set is randomly divided according to a ratio of 9 : 1, of which 90% of the images, that is, 1125 images, are used as the training set. Then the remaining 125 images are randomly divided according to the ratio of 9 : 1, of which 90%, that is, 112 reimbursement information images, are used as the verification set, and the remaining 10%, that is, 13 images are used as the test. set. At the same time, the generated annotation data is divided into training set, verification set and test set accordingly.

4.3. Construction of Target Detection Model for Travel Expense Reimbursement Information under Financial Sharing. The POLO-V3 model is based on the Darknet-53 model and detects the outputs of the Darknet-53 model at three different depths respectively, so as to achieve multi-scale detection, and use the regression method to get the position of the detected target.

4.3.1. Darknet-53 Model. The probability that an image belongs to a certain category can be determined according to

	Type	Filters	Size	Output
	Convolutional	32	3×3	256×256
	Convolutional	64	$3 \times 3 / 2$	128×128
1×	Convolutional	32	1×1	
	Convolutional	64	3×3	
	Residual			128×128
	Convolutional	128	$3 \times 3 / 2$	64×64
2×	Convolutional	64	1×1	
	Convolutional	128	3×3	
	Residual			64×64
	Convolutional	256	$3 \times 3 / 2$	32×32
8×	Convolutional	128	1×1	
	Convolutional	256	3×3	
	Residual			32×32
	Convolutional	512	$3 \times 3 / 2$	16×16
8×	Convolutional	256	1×1	
	Convolutional	512	3×3	
	Residual			16×16
	Convolutional	1024	$3 \times 3 / 2$	8×8
4×	Convolutional	512	1×1	
	Convolutional	1024	3×3	
	Residual			8×8
	Avgpool		Global	
	Connected		1000	
	Softmax			

FIGURE 4: Darknet-53 structure diagram.

the value between 0 and 1. In the end, it is only necessary to find the category corresponding to the value with the highest probability, which is the classification result of the input image by the model. The basic model for extracting image features mainly uses Darknet-53, and its structure is shown in Figure 4.

It can be seen from the structure diagram that the Darknet-53 model contains a total of 53 convolutional layers, of which the first 52 convolutional layers are used to

extract features, and the last convolutional layer has a size of 1×1 and a step size of 1, which is used to replace the entire convolutional layer. The connection layer is classed to reduce the amount of calculation, so the model is named Darknet-53.

The core structure in the model is the residual unit. Each residual unit is mainly composed of two consecutive convolutional layers. At the same time, the input of the residual unit and the output of the convolutional layer are added, and the sum is used as the residual unit Output. The structure diagram of the residual unit is shown in Figure 5.

Among them, x is the input of the residual unit, ReLU is the activation function, Fix) is the output of the two convolutional layers, and $H(x)$ is the output of the residual unit, that is, the sum of $F(x)$ and x. It can be seen that the residual unit directly adds the input and the output as the final output to ensure that the worst case of its structure, that is, the performance of the previous layer can be maintained when the current convolutional layer has zero output. This type of gradient disappears and gradients explode. Therefore, the role in the deep convolutional neural network. At the same time, the deep mode is well resolved, and the residual unit plays an important role.

There are five residual modules in the model, and each module is a stack of several residual units to extract deeper features. The number of residual units in the five modules are 1, 2, 8, 8, and 4 respectively. After the residual module has extracted the features, it uses the global average pooling method to fuse the features to facilitate subsequent classification.

4.3.2. POLO-V3 Model.

The model can correspond to different feature scales and can be used to detect objects of different sizes. The POLO-V3 model architecture is based on the Darknet-53 model. The first 52 convolutional layers of the Darknet-53 model are used to extract the features of the input image, and then the three feature maps of different scales in the extraction process are respectively performed Detection, using the idea of regression, so as to achieve the task of detecting and positioning the input image. The overall structure of the POLO-V3 model is shown in Figure 6.

It can be seen that the part marked by the red dashed box is the first 52 layers of the Darknet-53 model, that is, the feature extraction part. Using its feature extraction model, the detection is performed after the third, fourth and fifth residual modules of the feature extraction model. The size of the feature map is 52×52, 26×26 and 13×13, which correspond to different feature scales used to detect targets of different sizes.

4.4. Construction of Target Recognition Model of Travel Expense Reimbursement Information under Financial Sharing.

Information recognition is divided into two parts: text detection and text recognition. The text detection part uses the CTPN model to initially detect the text area in the reimbursement information image, and then uses postprocessing methods to optimize the preliminary detection results; the
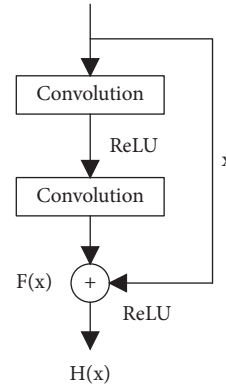

FIGURE 5: Residual unit structure diagram.

text recognition part uses the DenseNet model Recognize the detected text area.

4.4.1. CTPN Model.

The overall structure of the CTPN model includes the VGG-16 model, the BiLSTM model and the full convolutional layer. It combines the VGG-16 model and the BiLSTM model to extract both the spatial information of the image and the sequence features of the text, which is more suitable for detection For the text in the image, the fully convolutional layer is finally used to replace the fully connected layer, which simplifies the calculation and improves the efficiency. The overall result of the CTPN model is shown in Figure 7.

First, use the VGG-16 model to extract spatial features of the input image. Since only features need to be extracted and no classification is required, only the convolutional layer part is retained in the V GG-16 model, and its structure is shown in Figure 8.

The dimension of the feature map extracted by VGG-16 is NxCxHxW, where N is the number of output feature maps, C is the number of channels of the feature map, and H and W are the dimensions of the feature map, that is, height and width. Since the text information in the reimbursement information image is arranged horizontally from left to right, the 3×3 check feature map is used to slide row by row, and each pixel and the surrounding eight pixels are combined as a feature vector to obtain the dimension It is the feature map of Nx9CxHxW.

The size of the full convolution layer is set to 512, and the feature map dimension of the previous layer output is Nx256xHxW, then the output of the full convolution layer is the feature map dimension of Nx512xHxW. Three outputs can be obtained by classification and regression respectively, which are rectangular boxes. Position, rectangular box category score and rectangular box boundary calibration. Since the width of each rectangular box is fixed at 16 pixels by default, the position of the rectangular box is the ordinate and height of the center point of the detected rectangular box. The type of the rectangular box is whether the rectangular box contains text information. The boundary of the rectangular box is calibrated as a rectangle. The offset of the
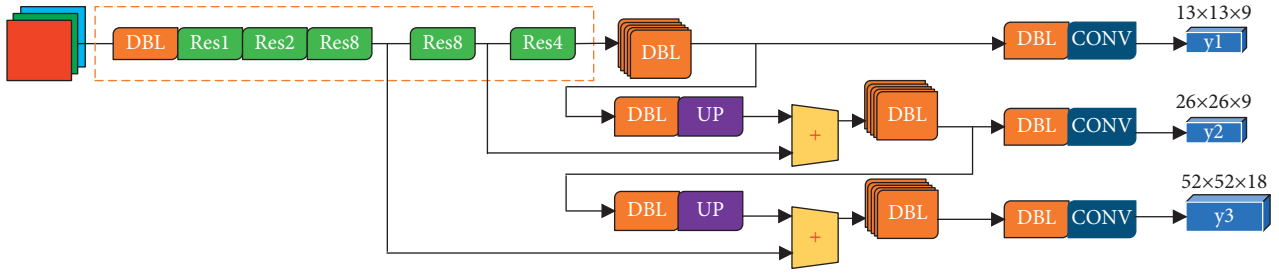
FIGURE 6: POLO-V3 model structure diagram.

frame in the horizontal direction. The loss function of the model is

$$L\left(S_i, V_j, O_k\right) = \frac{1}{N_x} \sum_i L_s^{cl}\left(S_i, S_i^*\right) + \frac{\lambda_1}{N_v} \sum_i L_v^{re}\left(V_j, V_j^*\right)$$

(6)

$$+ \frac{\lambda_2}{N_O} \sum_i L_o^{re}\left(O_k, O_k^*\right).$$

Among them, S_i, V_j, and O_k are the model category, the vertical coordinate and the horizontal offset of the L_s^{cl} is the cross-dimension loss function, which is used to calculate the category error, while L_v^{re} and L_o^{re} are both smooth L1 functions, which are used to calculate the vertical coordinate and horizontal offset errors of the rectangular box, respectively. $N_s.$, N_v, and N_o respectively represent the category, the ordinate of the rectangular frame, and the total number of anchor frames used for the horizontal offset of the rectangular frame. λ_1 and λ_2 are the weights for balancing different tasks, which are set to 1 and 2, respectively.

As for the horizontally arranged text, 10 anchor boxes with the same width are used to locate the text more accurately. The width of each anchor frame is 16 pixels, and the heights are 283, 198, 139, 97, 68, 48, 33, 23, 16, and 11, respectively, decreasing by a ratio of 0.7. Correct the 10 anchor boxes detected in each group, leaving only one anchor box as the text detection box at the current position.

4.4.2. DenseNet Model. After obtaining the text detection frame, train the DenseNet model to recognize the text in the image in the detection frame, so as to obtain the text information in the image. The biggest feature of the DenseNet model structure is that the input of each layer comes from the output of all the previous layers. That is to say, the output of each layer takes into account the results of each layer before it, which greatly enhances the characteristics of the image. Delivery in the model. In order to facilitate training, the DenseNet model is simplified, and its structure is shown in Figure 9.

5. Simulation Experiment and Result Analysis

5.1. Simulation. After the data set is prepared and the model structure is determined, the model can be simulated and
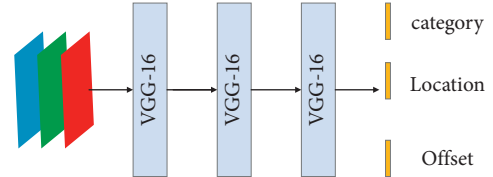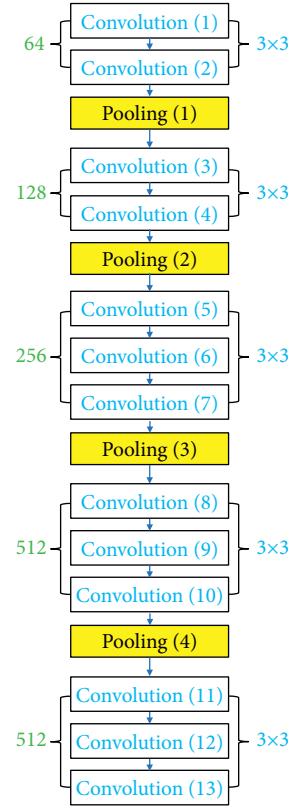

FIGURE 7: CTPN model structure diagram.

FIGURE 8: VGG-16 model structure diagram.

tested so that it can complete the task of automatically identifying the text detection frame of the reimbursement image. The experiment was run on a GeForce GTX 1080Ti GPU, and its video memory was about 11G.

Because text detection and text recognition use CTPN and DenseNet models respectively, the training process can be separated, and the test is performed sequentially. The specific parameters are set as follows.
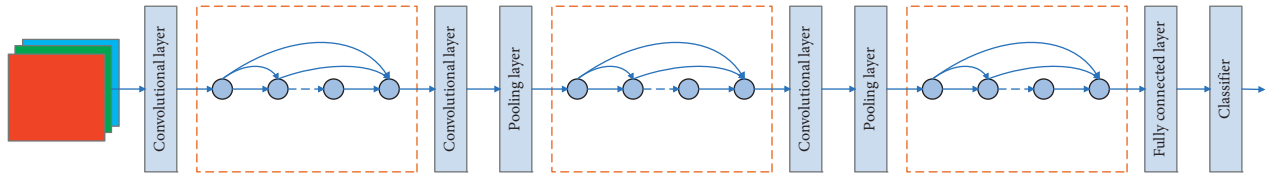
FIGURE 9: DenseNet model structure.

5.1.1. Input Image Size. When the CTPN model is trained, the image size is uniformly set to 900×900, that is, the width and height are both 600 pixels. When training the DenseNet model, the input image size is uniformly set to 280×32, in which the image width is 280 pixels, and the height is in pixels.

5.1.2. Training Batch. In this simulation experiment, a detection and recognition method of reimbursement information image information based on POLO-V3 model, CTPN model and DenseNet model is designed. In this section, the detection and recognition effect, accuracy and operating speed of the method are analyzed through experiments.

The CTPN model sets the maximum number of iterations to 50,000, and each batch trains 300 images. The DenseNet model sets the maximum number of training rounds to 10 rounds, and 128 images are trained in each batch.

5.1.3. Learning Rate. When the CTPN model is trained, the initial learning rate is 0.00001. When training 10,000 times, the learning rate is reduced to one-tenth of the current one. When the DenseNet model is trained, the initial learning rate is 0.0005, and every time it is trained, the learning rate is reduced to two-fifths of the current one.

5.1.4. Convolution Kernel. The specific structures of the CTPN model and the DenseNet model have been introduced in the previous section, and the specific convolutional layer parameter settings are detailed in the model structure introduction in the previous section.

5.1.5. Optimization Method. Both the CTPN model and the DenseNet model are trained using the Adam optimization method, that is, the adaptive moment estimation optimization algorithm.

5.1.6. Label Length. For the DenseNet model, since each image in the data set contains only 10 characters, the maximum length of its label is also set to 10.

5.2. Simulation Experiment Results and Analysis

5.2.1. Detection and Recognition Effect. This section will be based on the POLO-V3 model, C TPN model and DenseNet model of reimbursement image information detection and
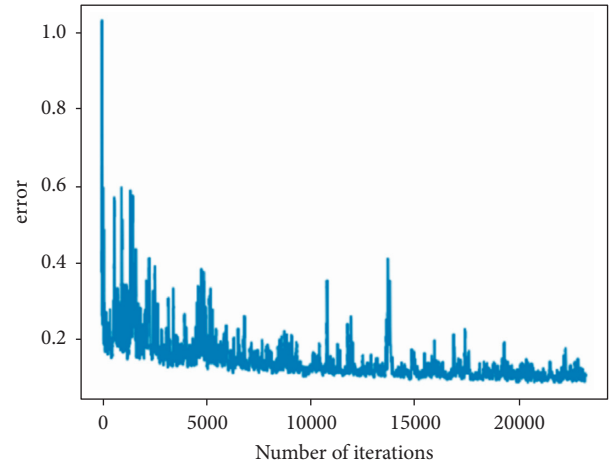

FIGURE 10: Training error graph.

recognition methods, through experiments to analyze the detection and recognition effect, accuracy and operating speed of the method. The input reimbursement information image is passed through the POLO-V3 model, the CTPN model and the DenseNet model in sequence, and the information area image, the text detection image and the text recognition result in the reimbursement information image are obtained respectively.

First, use the POLO-V3 model to perform target detection on the input reimbursement information image, detect the area where the relevant information is located in the reimbursement information image, and mark it with different colored rectangular boxes according to the category, and label the category label.

Then, based on the position information of the detected rectangular frame, the images of the relevant information area are respectively extracted.

Then, the CTPN model is used for text detection on the extracted information area images, and the positions of the text areas arranged in rows are obtained, and they are marked with a rectangular frame. It can be seen that the detection effect of the line text area is very good. For the "name" area, because the two characters are far apart, they are not all detected on the same line.

The DenseNet model is used for the recognition of the detected line text area image. The method designed in this chapter has a good effect on the detection and recognition of the reimbursement information image. It is very accurate except for the recognition of individual words.

5.2.2. Accuracy. In the process of detection and recognition of reimbursement information images, its accuracy is the
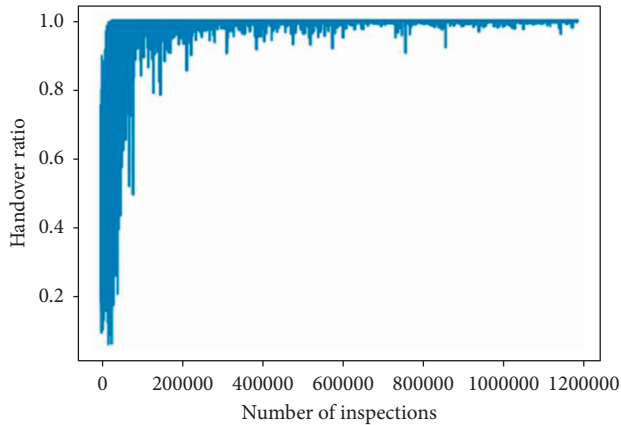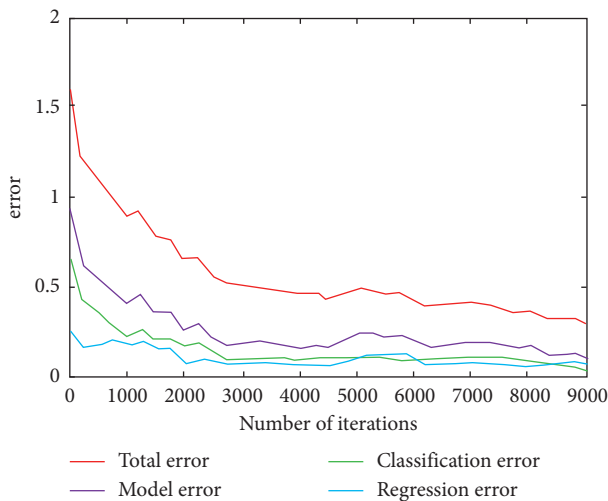
FIGURE 11: Training cross-combination diagram.

FIGURE 13: DenseNet model training accuracy curve.

FIGURE 12: CTPN model training error diagram.
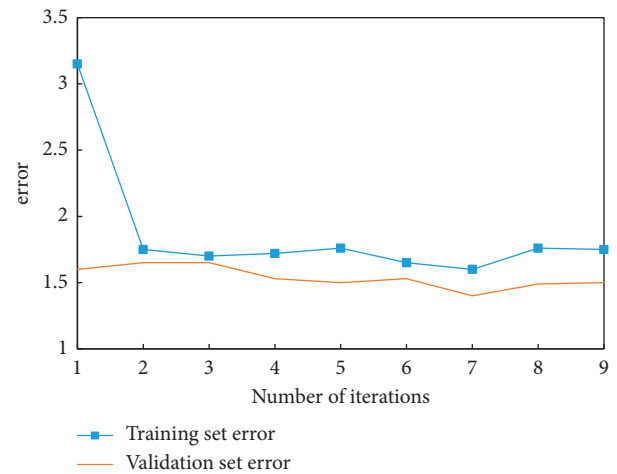
FIGURE 14: DenseNet model training error curve.

most important. This section conducts experiments and analyzes on the accuracy of detection and recognition of reimbursement information images.

For the problem of reimbursement information area detection, the training error loss and intersection ratio of the model are shown in Figures 10 and 11, respectively.

In the error loss diagram, the abscissa is the training batch, and the ordinate is the error loss. It can be seen that the final model converges and the error is small; The intersection ratio when the model converges is very high, close to 1, indicating that the accuracy of model detection is very high.

Since 16 images are trained in each batch, and the same image will be detected in three scales of large, medium and small, there is a multiple relationship between the number of detections and the training batch of 48, which can be obtained from the relationship between the abscissas of the two images Be confirmed.

For the recognition of the reimbursement information area, comparing the recognized characters with the original image of the reimbursement information, the accuracy rate is about 95.18%, and the recognition accuracy rate is better.
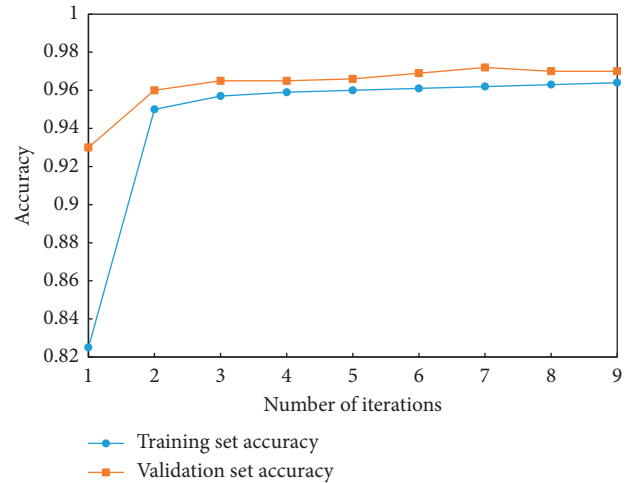
The error of CTPN model training is shown in Figure 12, where the overall error includes model error and horizontal offset error, and the model error includes classification error and regression error. It can be seen that after about 10,000 iterations of model training, the error curves all converge.

The accuracy and error of DenseNet model training are shown in Figures 13 and 14, respectively. It can be seen that the accuracy and error curves of the model converge after training, and the performance is very good.

5.2.3. Running Speed. In practical applications, high accuracy is often not enough, and a good user experience also has certain requirements for running speed. Since the detection and recognition methods of reimbursement information images are based on deep learning, the training of the model can be performed separately offline, so the training speed of the model is not considered, only the running speed of the tested model. The specific test results are shown in Table 1.

For the detection of the reimbursement information area, ten reimbursement information images were tested, and the average detection time for each reimbursement

TABLE 1: Running speed test.

	Detection time (s)	Recognition time per sheet (s)	Recognition time per line (s)
Image 1	0.68	1.346	0.056083
Image 2	0.63	1.312	0.50462
Image 3	0.66	1.264	0.046815
Image 4	0.63	1.107	0.04813
Image 5	0.71	1.233	0.047423
Image 6	0.59	1.273	0.05092
Image 7	0.67	1.126	0.048957
Image 8	0.62	1.09	0.047391
Image 9	0.68	4.192	0.182261
Image 10	0.7	1.578	0.068609
Average value	0.66	1.55	0.06

information was about 0.66 seconds. There are four information areas in each reimbursement information image that need to be detected, and the average detection of each area only takes about 0.17 seconds. Compared with the entire reimbursement information reimbursement process, the detection time is minimal, and it fully meets the requirements of real-time detection.

Therefore, the deep learning-based method designed in this section has a very fast running speed for the detection and recognition of reimbursement information images, and the detection and recognition of each area can meet the real-time requirements. In actual applications, the detection and recognition can be performed as needed. Modifying the identification area can greatly improve the efficiency of detection, while ensuring the compliance and effectiveness of reimbursement information, and enriching the data accumulation of the financial sharing service platform.

6. Conclusion

In order to transform and develop, enterprises need to use the financial sharing model to improve the implementation plan of internal control, so as to improve the implementation of informatization in financial management and deepen the goal of enterprise modernization and transformation. Therefore, this paper discusses the research on the validity detection and identification of travel reimbursement content and expense data in the enterprise financial management business. process design, the following conclusions can be drawn:

(1) According to the simulation of the training model, the CTPN model has a high success rate in detecting reimbursement vouchers. At the same time, the DenseNet model can effectively identify the detected content with high accuracy, which verifies the validity and feasibility of the model. sex.

(2) By testing the training speed of CTPN and DenseNnet, it is further verified that users will get a good experience in collecting fee vouchers independently.

(3) Based on the effective identification of expense reimbursement content and data, it provides guarantee and data support for the compliance determination of expense data information carried out by the

Financial Shared Service Center, enabling the efficient operation of the service model of individual expense reimbursement.

Data Availability

The dataset can be accessed upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] Y. Yang, Q. Liu, J. Song, and M. Zhou, "The influence mechanism of financial shared service mode on the competitive advantage of enterprises from the perspective of organizational complexity: a force field analysis," *International Journal of Accounting Information Systems*, vol. 42, Article ID 100525, 2021.

[2] Q. Yao, A. Au, T. Auang, and J. Hu, "TAe ResearcA of Implementing Enterprise Financial SAared Service Center Information System," in *Proceedings of the 2012 International Conference on Computing, Measurement, Control and Sensor Network*, pp. 388–391, Taiyuan, Shanxi, China, 2012.

[3] N. Dokoohaki and M. Matskin, "Reasoning about weighted semantic user profiles through collective Confidence analysis: a fuzzy evaluation," *Advances in Soft Computing*, vol. 67, pp. 71–81, 2010.

[4] B. Shi, X. Bai, and C. Yao, "An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 11, pp. 2298–2304, 2017.

[5] R Smith, "An overview of the Tesseract OCR engine," in *Proceedings of the International Conference on Document Analysis and Recognition(ICDAR)*, pp. 629–633, Curitiba, Brazil, September 2007.

[6] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[7] G. Hinton, L. Deng, D. Yu et al., "Deep neural networks for acoustic modeling in speech recognition: the shared views of four research groups," *IEEE Signal Processing Magazine*, vol. 29, no. 6, pp. 82–97, 2012.

[8] A Krizhevsky, I Sutskever, and G. E Hinton, "Imagenet classification with deep convolutional neural networks," in *Proceedings of the Advances in Neural Information Processing*

Systems, pp. 1097–1105, San Francisco, CA, United States, 2012.

[9] A Graves, A Mohamed, and G Hinton, "Speech recognition with deep recurrent neural networks," in *Proceedings of the International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 6645–6649, Singapore, 2013.

[10] G. Hinton, "Deep belief networks," *Scholarpedia*, vol. 4, no. 5, p. 5947, 2009.

[11] C Szegedy, W Liu, Y Jia et al., "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1–9, Boston, MA, 2015.

[12] S. J. Kaplan, "Organization of behavior," *Yale Journal of Biology & Medicine*, vol. 23, no. 1, p. 79, 1950.

[13] F. Rosenblatt, "The perceptron: a probabilistic model for information storage and organization in the brain," *Psychological Review*, vol. 65, no. 6, pp. 386–408, 1988.

[14] E Karami, M Shehata, and A Smith, "Image Identification Using SIFT Algorithm: Performance Analysis against Different Image Deformations," 2017, http://arxiv.org/abs/1710.02728.

[15] H. M. Khan, "Automated breast cancer diagnosis using artificial neural network (ANN)," in *Proceedings of the Iranian Conference on Intelligent Systems and Signal Processing*, pp. 54–58, Shahrood, Iran, 2017.

[16] T. Li and H. Zhang, "Digital Image Enhancement System Based on MATLAB GUI," in *Proceedings of the International Conference on Software Engineering and Service Science (ICSESS)*, pp. 296–299, Beijing, China, 2017.

[17] A. Kumari, P. J. Thomas, and S. K. Sahoo, "Single image fog removal using gamma transformation and median filtering," in *Proceedings of the IEEE International Conference on INDICON*, pp. 1–5, Pune, India, 2014.

[18] G. A. Abdenova, "Scaling of Input-Output Data and Number of Conditionality of a Matrix of Grama," in *Proceedings of the Actual Problems of Electronics Instrument Engineering*, pp. 16–18, Novosibirsk, Russia, October 2012.

[19] S. H. Songe, M. Antonelli, T. W. K. Fung, B. D. Armstrong, A. Chong, A. Lo, and B. E. Shi, "Developing and assessing MATLAB exercises for active concept learning," *IEEE Transactions on Education*, vol. 99, pp. 1–9, 2018.

[20] K. He, G. Gkioxari, P. Dollar, and R. Girshick, "Mask R-CNN," in *Proceedings of the IEEE International Conference on Computer Vision (ICC)*, pp. 2961–2969, Cambridge, MA, USA, 2017.

[21] E. Meng, S. Huang, Q. Huang, W. Fang, L. Wu, and L. Wang, "A robust method for non-stationary streamflow prediction based on improved EMD-SVM model," *Journal of Hydrology*, vol. 568, pp. 462–478, 2019.

[22] J Redmon, S Divvala, R Girshick, and A. Farhadi, "You only look once: unified, real-time object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 779–788, NV, USA, June 2016.

[23] J Redmon and A Farhadi, "YOL09000: better, faster, stronger," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 7263–7271, Honolulu, Hawaii, 2017.

[24] J Redmon and A. Farhadi, "Yolov3: An Incremental Improvement," 2018, http://arxiv.org/abs/1804.02767.