

RESEARCH

Open Access



Visual analysis method for unmanned pumping stations on dynamic platforms based on data fusion technology

Zhen Liu¹, Sen Chen^{2,3*} , Zhaobo Zhang¹, Jiahao Qin¹ and Bao Peng⁴

*Correspondence:
chensen@m.scnu.edu.cn

¹ Guangdong GDH Pearl River Delta Water Supply Co., Ltd., Guangzhou 511458, Guangdong, China

² South China Academy of Advanced Optoelectronics, South China Normal University, Guangzhou 510006, China

³ College of Electron and Information, University of Electronic Science and Technology of China, Zhongshan Institute, Zhongshan 528402, China

⁴ Shenzhen Institute of Information Technology, Shenzhen 518172, Guangdong, China

Abstract

As the scale of water conservancy projects continues to expand, the amount and complexity of analytical data have also correspondingly increased. At present, it is difficult to realize project management decision support based on a single data source, and most manual analysis methods not only have high labor costs, but also are prone to the risk of misjudgment, resulting in huge property losses. Based on this problem, this paper proposes visual analysis method for unmanned pumping stations on dynamic platforms based on data fusion technology. First, the method uses the transfer learning method to enable ResNet18 obtain generalization ability. Secondly, the method uses ResNet18 to extract image features, and outputs fixed length sequence data as the input of long short-term memory (LSTM). Finally, the method uses LSTM outputs the classification results. The experimental results demonstrate that the algorithm model can achieve an impressive accuracy of 99.032%, outperforming the combination of traditional feature extraction and machine learning methods. This model effectively recognizes and classifies images of pumping stations, significantly reducing the risk of accidents in these facilities.

Keywords: Visual analysis, Transfer learning, ResNet18, LSTM

1 Introduction

With the increasingly significant changes in the environment, global resources are gradually decreasing, rendering it increasingly difficult for people to meet their basic living needs, particularly the demand for water resources [1–4]. In response to this challenge, an increasing number of water resource scheduling projects are emerging to tackle the complexities inherent in water resource management. Within the domain of water resource scheduling engineering, a variety of methods have been developed for the establishment of an integrated operation system that encompasses multiple reservoirs and pumping stations. These methods effectively mitigate regional water replenishment, address resource shortages and manage overflow problems [5–9]. However, these methods overlook monitoring pumping station safety. In practical applications, predicting and analyzing the operating status of pumping stations is crucial for ensuring the stability of water resource scheduling projects. With the continuous expansion of unmanned

pumping stations and the increasing complexity of equipment, it is necessary to develop real-time monitoring, early warning and efficient management systems to ensure the safety and stability of pumping station operation.

Unmanned monitoring technology is revolutionizing water conservancy engineering by addressing the limitations of conventional methods. Traditional methods rely on manual inspection and regular maintenance, which suffer from inefficiency, high costs and delayed responses to equipment failures. As an important direction of modern water conservancy engineering construction, unmanned monitoring technology achieves automated operation and efficient management of pumping stations through technologies such as sensor monitoring, remote monitoring of stations and unmanned aerial vehicle optimization scheduling and inspection [10–12]. Compared to manual methods, unmanned monitoring technology conserves human resources, enhances production efficiency and ensures timely responses to equipment failures. There have been a number of studies utilizing unmanned surveillance technology to monitor water projects. Gama Moreno et al. proposed a system to measure the water level of water tanks using ultrasonic sensors and Arduino equipment. The system provides real time access to the status of water tanks through the GSM network. This facilitates timely interventions, improving overall water supply efficiency [13]. Getu et al. proposed a water level sensor with a seven-segment display and a relay-based motor pump drive for automatic water level control, eliminating the need for manual monitoring. This design contributes to efficient water management and enhances automated system productivity [14]. Apte et al. proposed an automated monitoring system addressing tank overflow, automatic filling and water quality with integrated sensors. They enhanced pumping system efficiency by implementing a leak detection system and remote communication for identifying and alerting users about pipe leaks [15]. Remote monitoring stations are another crucial component of unmanned monitoring systems. These stations analyze on-site data collected by sensors, providing real-time information on pumping station operation status to users, central monitoring centers or relevant institutions. Klokov et al., for instance, improved the efficiency of sewage pumping stations by optimizing electric motor units and developing a modern sewage pumping station with real-time visualization technology and remote control capabilities [16]. Tlabu et al. proposed a centralized intelligent digital water management system that utilizes a data-centric pump infrastructure and a comprehensive system architecture model to achieve data security, real-time monitoring and digitization to provide comprehensive support for business operations [17]. Mahjoub implemented an intelligent automated solution that integrates algorithms into microcontroller units, connecting PLC, operation panels and industrial networks to web servers. This allows for online control, anomaly detection and decision-making to enhance the availability, reliability and safety of production systems [18]. Despite notable progress, traditional unmanned monitoring still faces limitations, necessitating regular maintenance by technical personnel. The complexity of hardware systems results in elevated costs for personnel and equipment. Moreover, the strong interdependence between modules can impede the accurate real-time monitoring of pumping station equipment.

In recent years, the rapid development of deep learning has introduced a paradigm shift in visual analysis methods. By applying deep learning techniques to analyze

pumping station equipment images, more efficient and reliable technical support for pumping station operation management and maintenance can be achieved. The deep learning method was first composed of neural networks proposed by Hinton et al. [19], which sparked a wave of deep learning. Alex Krizhevsky et al. proposed an AlexNet [20] based on the structure of convolutional neural network (CNN) that won the ImageNet image recognition competition, which also caused a huge response in the industry. He et al. proposed a convolutional neural network ResNet based on the shortcut structure [21], which solves the degradation phenomenon of network models after deep training, making deep learning models more expressive and suitable for complex tasks. Szegedy et al. from Google proposed an Inception V4 based on Inception and Residual structures [22], which reduces the parameter count of convolutional neural networks and improves the algorithm's running speed. With the continuous development of deep learning, a growing number of deep learning methods are being proposed, such as generative adversarial network (GAN) which improves the generalization ability of models through sample expansion [23], recurrent neural network (RNN) which captures correlations in sequence data and maintains the model's memory of historical information [24–26], LSTM is a special type of RNN that can learn long-term dependency information [27], the Unet network optimizes the performance of deep learning models in image segmentation tasks [28], while the Transformer network solves the problems of vanishing and exploding gradients, enabling deep learning models to better capture long-distance dependencies [29]. Therefore, deep learning models have a wide range of applications in fields such as image recognition, speech recognition and data analysis. Using deep learning for visual analysis can reduce the steps of manually designing feature extractors and achieve high accuracy and robustness after training on large-scale datasets.

At present, many studies are committed to applying deep learning technology to visual analysis tasks in the industrial field [30–33], promoting the rapid development of intelligent industry. The high automation and excellent performance of this technology provide new possibilities for visual analysis of unmanned pumping stations. However, there is relatively little research on visual analysis of unmanned pumping stations. By utilizing deep learning techniques to extract pumping station features, visual analysis of pumping stations can be conducted from multiple perspectives, which not only significantly reduces human resource costs but also achieves the goal of unmanned pumping stations.

Based on the current problems with unmanned monitoring technology, this paper proposes a visual analysis method based on deep learning, which augments the dataset by employing data augmentation techniques on unmanned pumping station images. The integration of transfer learning, ResNet18 [21] and LSTM is employed to enhance the model's generalization ability. This approach enables the identification, classification and analysis of the status of unmanned pumping stations, facilitating real-time monitoring and early warning of pumping station equipment status. Experimental results demonstrate the applicability of this method to automatic inspection, equipment fault diagnosis, maintenance monitoring and other facets of unmanned pumping stations. The method provides robust technical support for improving pumping station operation efficiency, reducing equipment maintenance costs and ensuring water resource safety.

2 Methods

2.1 Model framework

The model framework employed in this study is shown in Fig. 1. It consists of two models, ResNet18 and LSTM. ResNet18 serves to extract features from pump images, while LSTM is employed to capture contextual relationships among pump images and enhance the classification of pumping station status. ResNet18 undergoes transfer learning using the ImageNet large-scale dataset [20]. Following transfer learning, ResNet18 is utilized to extract features from pumping station images and these extracted features are then fed into LSTM for classification to yield prediction results. Transfer learning is widely utilized in computer vision, where inputting algorithm models into existing standard datasets for training enables the derivation of optimal weight models for practical projects. This method significantly reduces the complexity of model training and enhances convergence speed. Due to the scarcity of visual images in the pump station industry, the application of transfer learning endows the model with generalization ability, enhancing its robustness. It also enabled the model to maintain high accuracy, even in the case of pump station images that show significant differences. ResNet18 is an algorithmic model based on convolutional neural networks that prevents gradient vanishing and network degradation during deep training. This characteristic ensures improved model convergence and its lightweight nature facilitates deployment in real-time monitoring applications for pumping station status. LSTM, as a specialized recurrent neural network structure, is primarily employed for processing sequence data. In this study, ResNet18 is employed to generate sequence data as input to LSTM, effectively mitigating gradient vanishing and exploding issues. Furthermore, considering the large volume and spatial occupation of water pumps in images, with continuous pixels, utilizing LSTM to establish contextual relationships among pumping station images enhances the visual analysis of the pumping station's status.

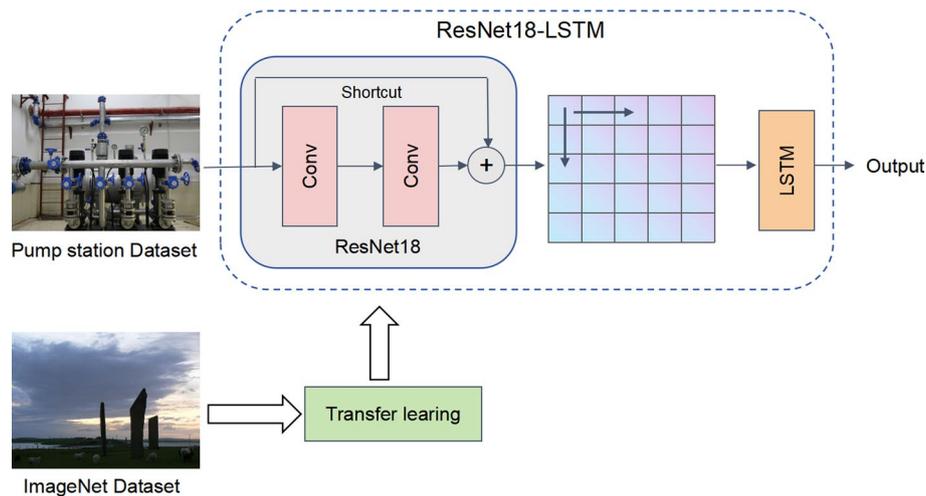


Fig. 1 Network model

2.2 Network model

2.2.1 Convolutional feature extractor

Feature extraction constitutes a pivotal phase in the visual analysis of images, and this article employs ResNet18 to undertake this crucial task. ResNet18 is trained on the ImageNet dataset and the weight parameters are upgraded by transfer learning. Given the limited number of pumping station state samples, the application of transfer learning methods proves instrumental in significantly reducing the time and sample size required for model training. This method not only enhances the model's generalization ability but also empowers ResNet18 to extract fundamental features.

ResNet18 consists a series of convolutional layers, pooling layers, global pooling layers and fully connected layers. The convolutional layer, characterized by a step size of 2 and a convolution kernel size of $7 * 7$, executes the multiplication of input feature values with weight values, followed by the addition of a bias value to yield the convolutional output value. This process completes the extraction of image features. This convolutional operation serves as a fundamental block for ResNet18, allowing it to effectively capture and represent intricate features in pumping station images. The convolutional formula is shown as:

$$y = w \times x + b \quad (1)$$

where w is the weight value, x is the input feature value, b is the bias value and y is the convolutional output value.

The pooling layer in ResNet18 employs a maximum pooling method with a step size of 2 and a size of 3×3 . This pooling operation serves to downsample the input feature map, effectively reducing its size while retaining crucial feature information. The down-sampling process contributes to decreased computational and memory requirements, resulting in a more lightweight model. The pooling formula is shown as:

$$y = \max \left(\begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{bmatrix} \right) \quad (2)$$

where x_{ij} is the input feature value, $\max(x)$ is the function that takes the maximum value of x , y is the pooled output value. ResNet18 primarily comprises stacked residual blocks, each consisting of two convolutional layers and a shortcut, as illustrated in Fig. 1 of ResNet18 architecture. The convolutional layers employ a step size of 1 and a 3×3 convolution kernel. Following convolution, the output feature values undergo ReLU activation to uphold network sparsity and mitigate the issue of gradient vanishing. This architectural design enhances the model's ability to capture intricate patterns in data while addressing challenges associated with gradient propagation. The ReLU formula is shown as:

$$y = \begin{cases} 1, & x > 0 \\ 0, & x \leq 0 \end{cases} \quad (3)$$

where x is the input feature value, y is the output value of the activation function. The shortcut in ResNet18 preserves the original input information by traversing the convolutional layer and adding it to the output value post-convolutional operation, subsequently

passing it to successive layers. This mechanism enhances the network's capacity to learn residual and detailed information, mitigates the issue of gradient vanishing, facilitates the swift transmission and retention of information, expedites network convergence and thereby enhances overall performance. The Short cut formula is shown as:

$$y = F(x) + x \quad (4)$$

where $F(x)$ is the output value after the convolutional operation, x is the input feature value and y is the shortcut output value. The global average pooling layer transforms the output layer of the final residual block into a fixed-length vector, thereby preserving information across channels and aiding in feature extraction. The formula for average pooling is shown as:

$$y = \text{avg} \left(\begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{bmatrix} \right) \quad (5)$$

where x_{ij} is the input feature value, $\text{avg}(x)$ is the function that takes the average value of x , and y is the pooled output value. The fully connected layer links all nodes within the feature vector produced by the global average pooling layer, mapping them to the requisite length of the input sequence data for LSTM.

2.2.2 LSTM classifier

The LSTM classifier is the process of predicting the pumping station status based on the pumping station image. LSTM is a variant of RNN structure, mainly used for modeling and prediction tasks of sequence data. Compared to traditional RNNs, LSTM has stronger memory and long-term dependency modeling capabilities, which can effectively solve the problems faced by traditional RNNs such as gradient vanishing and exploding. During normal pump operation, pixel information within the water pump remains continuous. However, in the event of a malfunction, faults between pixels emerge, leading to a noticeable divergence in the region surrounding the faulty pixel from the typical water pump operation. Leveraging LSTM to capture contextual relationships within the image sequence features, extracted by ResNet18, proves advantageous in classifying pumping station statuses. This approach enables a more comprehensive understanding of spatial and temporal information embedded in water pump images, thereby enhancing classification accuracy. A distinctive feature of LSTM lies in the introduction of a structure known as a "gate," which regulates the input, forgetting, and output of information in distinct ways. This gating mechanism effectively controls the flow of information and facilitates memory updates, contributing to the model's robust performance. A standard LSTM unit consists of forget gate, input gate, output gate and cell state, as shown in Fig. 1. In the figure, S represents the sigmoid function and T represents the tanh function are shown as:

$$f(x) = \frac{1}{1 + e^{-x}} \quad (6)$$

$$f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (7)$$

The forget gate of LSTM controls the update of cell state by using a sigmoid function to determine whether to retain or forget the previous cell state. The forget gate first fuses the sequence data x_t outputted by ResNet18 with the memory state h_{t-1} at time $t - 1$, and after passing through the sigmoid function, it is fused with the cell state c_{t-1} at time $t - 1$. The formula for the forget gate is shown as:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (8)$$

where σ is the sigmoid function, W_f is the weight matrix of the forget gate, b_f is the bias vector of the forget gate. The input gate controls the inflow of input information by using a sigmoid function to determine whether to fuse the current input information with the cell state. The input gate first passes the fused data of x_t and h_{t-1} through the sigmoid function and tanh function and outputs the input gate i_t and candidate memory unit \tilde{c}_t respectively. Then it multiplies i_t with \tilde{c}_t and f_t with c_{t-1} element by element and fuses the states of the two multiplied results to obtain the updated memory unit c_t . The calculation formulas for input gate i_t , candidate memory unit \tilde{c}_t and updated memory unit c_t are shown as:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (9)$$

$$\tilde{c}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (10)$$

$$c_t = i_t * \tilde{c}_t + f_t * c_{t-1} \quad (11)$$

where W_i is the feature matrix of the input gate, b_i is the bias vector of the input gate, W_c is the feature matrix of the candidate memory unit and b_c is the bias vector of the candidate memory unit. The output gate determines whether to output the current cell state through the sigmoid function to control the generation of output information. Firstly, the fused data of x_t and h_{t-1} is used to obtain the output gate o_t through the sigmoid function. Then, the memory unit c_t is controlled to be scaled through the tanh function. Finally, the scaled memory unit c_t is multiplied element by element with the output gate o_t to obtain the memory state h_t , which is the model prediction result. The calculation formulas for output gate o_t and memory state h_t are shown as:

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (12)$$

$$h_t = o_t * \tanh(c_t) \quad (13)$$

where W_o is the feature matrix of the output gate and b_o is the bias vector of the output gate.

2.3 Transfer learning

Transfer learning is a widely adopted technique in deep learning, expediting model training and enhancing performance by leveraging model parameters, weights and features trained in a source domain for tasks in a target domain. This approach proves particularly valuable in scenarios with limited data, where crucial parameters and features can be gleaned from extensive datasets and applied to target tasks, enabling models to

perform effectively even in situations of data scarcity. In the context of this article, the ImageNet dataset is chosen as a large-scale dataset for transfer learning on ResNet18. The pretrained ResNet18 is then applied to extract features from pumping stations, addressing both the shortage of pumping station data and the challenges associated with extracting meaningful features. Furthermore, the article proposes the redefinition of a fully connected layer to output sequence data with a length of 500 as the input for LSTM. This modification enhances the model's ability to articulate pumping station features while concurrently reducing the number of parameters, thereby accelerating the training speed of the model. Through this transfer learning methodology, the network model achieves improved generalization to the target task, learning more universal features and ultimately enhancing its performance on the pumping station-related tasks.

3 Results and discussion

3.1 Dataset

The dataset utilized in this study is self-collected and preprocessed, encompassing 1546 images capturing pumping stations at various time periods, angles and operational states. Each image boasts a resolution of 320 * 320 pixels. To organize the data effectively, images of the same category are grouped within the same folder, and labels are automatically generated using software. The label information encompasses three distinct operating states: complete failure, partial failure and normal operation. In the context of this dataset, the "normal operation" category denotes the absence of any abnormalities across all equipment and components. This includes a fully intact pumping station structure, stable operation of crucial equipment such as water pumps and valves, and so forth. "Partial faults" indicate malfunctions or abnormal conditions in specific equipment or components within the pumping station, such as structural damage, rusting, or minor issues with a particular water pump or valve. Lastly, "complete failure" signifies a severe malfunction in key equipment or components within the pumping station. Examples include significant structural damage, the inability to open, close crucial valves or an inability to pump water. The process of data preprocessing is shown in Fig. 2. The first

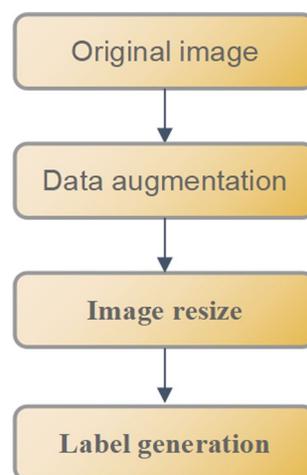


Fig. 2 The procedure of image processing

step is to obtain the original image. Subsequently, data augmentation techniques were applied to enrich the sample set. Then, standardize the size of the data enhanced images and generate corresponding labels.

3.2 Data augmentation

Data augmentation involves applying a series of transformations and perturbations to the original data to generate more diverse and enriched training samples, thereby enhancing the model's generalization ability and robustness. In this study, various augmentation techniques, including rotation, flipping, scaling and noise addition, are employed to augment the dataset. Rotation is implemented at angles of 30°, 60°, 90°, 120°, 150° and 180° to enhance the model's robustness to different angle transformations, enabling it to effectively handle variations in image orientation. Flipping is performed both horizontally and vertically, augmenting the dataset and improving the model's adaptability to changes in image direction. Scaling, with ratios of 0.5 and 1.5, simulates target objects at different scales, providing additional samples for scale changes and enhancing the model's ability to recognize variations in scale. The addition of noise, including Gaussian noise, salt, pepper noise and multiplicative noise, serves to simulate real-world environmental interference. This augmentation strategy improves the model's recognition and classification capabilities in complex and noisy environments.

3.3 Evaluating indicator

This study establishes a network model that integrates transfer learning, ResNet18 and LSTM for dynamic visual analysis of unmanned pumping stations. The comprehensive evaluation of the model encompasses four key indicators: accuracy, precision, recall and F1 score.

Accuracy refers to the proportion of correctly predicted samples by the classification model among the total number of samples, serving as a metric to gauge the overall precision of the model's predictions. The accuracy calculation formula is shown as:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (14)$$

where TP is the number of correctly identified positive samples, FP is the number of incorrectly identified negative samples, TN is the number of correctly identified negative samples, and FN is the number of incorrectly identified positive samples.

Precision denotes the ratio of positively predicted samples by the classification model that are genuinely positive, offering a measure of the model's precision in predicting positive samples. The precision calculation formula is shown as:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (15)$$

Recall represents the ratio of positive samples predicted by the classification model among all positive samples, providing insight into the model's ability to recognize positive instances. The recall calculation formula is shown as:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (16)$$

F1 score, as the harmonic mean of precision and recall, indicates the model's performance in achieving a balance between accuracy and recall. A higher F1 score implies superior model performance in this regard. The F1score calculation formula is shown as:

$$\text{F1score} = \frac{2 \times \text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} \quad (17)$$

These indicators collectively offer a comprehensive evaluation of the model's classification performance from various angles. Through a thorough analysis of these metrics, adjustments to model parameters can be made to enhance its overall classification performance.

3.4 Training data and parameters

This study employed a model that integrates transfer learning, ResNet18 and LSTM to predict the three different states of pumping stations. Transfer learning was utilized to enhance the model's generalization performance and expedite convergence, while ResNet18 played a crucial role in extracting image features, preventing gradient vanishing and addressing potential model degradation issues. LSTM was employed to establish contextual connections within the input sequence data, effectively managing long-term dependencies in the sequence and mitigating problems associated with gradient vanishing and exploding.

The tool parameters employed in this experiment are detailed in Table 1. Throughout the study, the dataset was partitioned into training, validation and testing sets, maintaining a ratio of 6:2:2. The training and validation sets were utilized during the training phase, enabling the algorithm model to learn and adjust its parameters. Specifically, the training set was employed for training the model's parameters and weights, ultimately enhancing the model's prediction accuracy. Concurrently, the validation set played a pivotal role in fine-tuning hyperparameters, preventing overfitting to the training set and improving the model's generalization capabilities. To further elaborate on the experimental setup, a batch size of 32, a learning rate of 0.0005 and 20 epochs were specified. The output sequence data length from ResNet18 was set to 500 and the LSTM output length was configured to 3, corresponding to the three different states of the pumping stations. The optimization process involved using the Adam optimizer and cross-entropy loss function to refine the weight values of the model.

Table 1 Tool parameters

Tool	Parameters
Graphics card	RTX 3060
Operating system	Windows 11
Cuda	Cuda 11.7
Data processing	Python3.8

3.5 Experiments

To demonstrate the efficacy of transfer learning in improving model convergence and enhancing generalization to the target task, this article visualizes the changes in loss values and accuracy throughout the training and validation stages, as shown in Fig. 3. Fig (a) illustrates the loss function's progression during training, showcasing changes in the model's loss values per training iteration. The horizontal axis represents the number of training batches, while the vertical axis denotes the loss function's values. As training advances, the loss function gradually decreases, exhibiting a rapid downward trend that stabilizes over time. This trend signifies that transfer learning has effectively trained the algorithm model using extensive datasets, resulting in the acquisition of robust feature representations. Fig (b) presents the accuracy graph during the training phase, illustrating changes in the model's accuracy across each training iteration. The horizontal axis represents the number of training epochs and the vertical axis depicts accuracy values. The figure reveals that the accuracy surpasses 90% by the fourth epoch, indicating the model's proficient classification of samples. Particularly in scenarios with limited sample sizes, transfer learning proves advantageous by avoiding the need to train the model from scratch, thereby conserving substantial data collection and training time.

To demonstrate the performance of the algorithm model in this study, the same dataset was used to train and validate LSTM, ResNet18 and ResNet18 + LSTM. Next, evaluation metrics scores and confusion matrices for each model are obtained using the test set and compared with the proposed algorithm model in this paper. The results of these indices are shown in Table 2.

The experimental results reveal that when utilizing only LSTM, the accuracy, precision, recall, and F1 score are all notably lower, suggesting that LSTM is not inherently suitable for direct deployment in image classification tasks. Given that ResNet18 operates on a CNN structure, it exhibits significant advantages in image classification, leading to substantial improvements in evaluation metrics compared to LSTM alone. Upon combining ResNet18 with LSTM, where the output of ResNet18 feeds into the input of LSTM, there is a further enhancement in model performance, with an approximate accuracy increase of 3.226%. This observation underscores the effectiveness of using ResNet18 to generate sequence data and subsequently employing

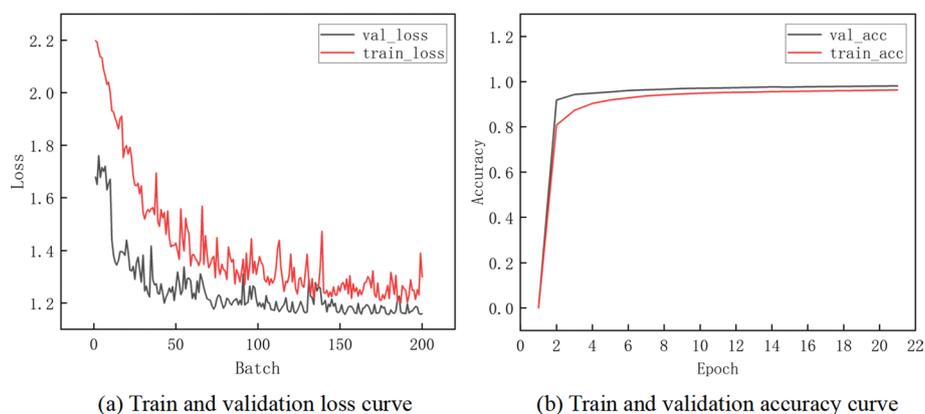


Fig. 3 Transfer learning ResNet18 + LSTM loss and accuracy during train and validation

Table 2 Performance comparison of deep learning models

Algorithm	Accuracy (%)	Precision (avg) (%)	Recall (avg) (%)	F1score (avg) (%)
LSTM	71.935	72.326	71.918	72.011
ResNet18	94.193	94.447	94.230	94.206
ResNet18 + LSTM	97.419	97.410	97.404	97.405
Transfer learning ResNet18 + LSTM	99.032	99.057	99.020	99.024

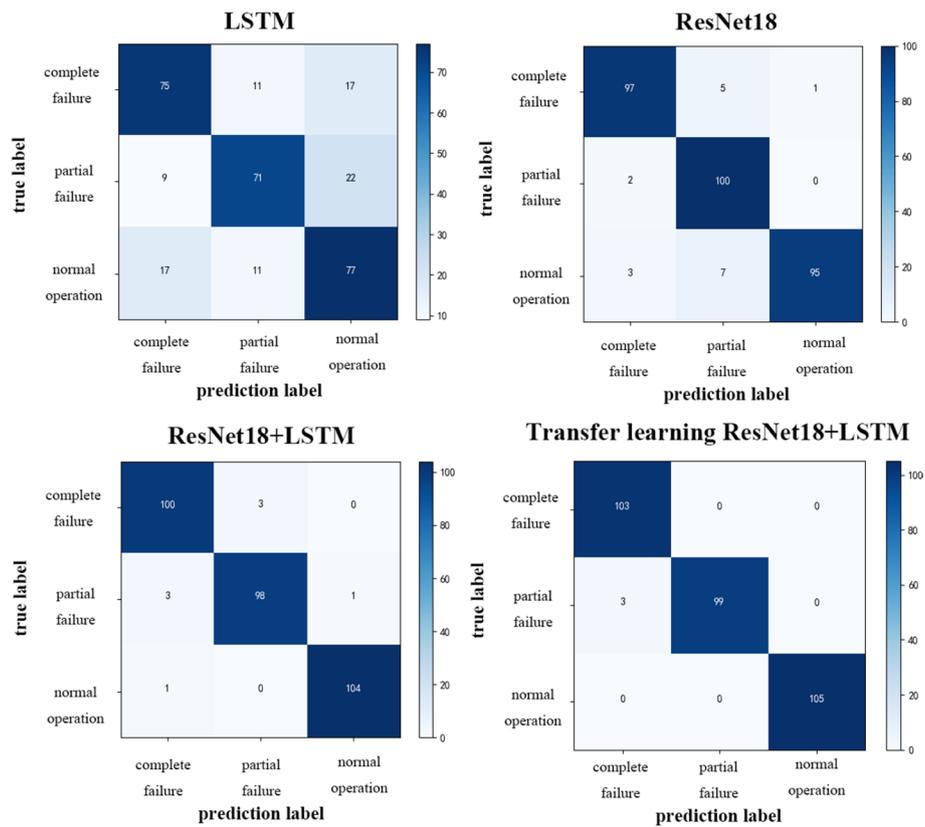


Fig. 4 The confusion matrix of LSTM, ResNet18, ResNet18 + LSTM and Transfer Learning ResNet18 + LSTM

LSTM for sequence data classification. Moreover, the application of transfer learning on ResNet18 before image feature extraction contributes to an additional accuracy improvement of 1.613%. This outcome underscores the positive impact of transfer learning in enabling the model to acquire general features, augment its generalization capabilities and enhance the accuracy of classification results in practical tasks.

To provide a more visual representation of each model’s performance, this article generated confusion matrices for LSTM, ResNet18, ResNet18 + LSTM and transfer learning ResNet18 + LSTM, depicted in Fig. 4. In these matrices, darker colors represent higher values within the rectangular blocks. A careful examination of the confusion matrices reveals that the model proposed in this article consistently achieves the highest prediction accuracy, effectively discerning between normal operation, partial

Table 3 Performance comparison of machine learning under HOG feature extraction

Algorithm	Accuracy (%)	Precision (avg) (%)	Recall (avg) (%)	F1score (avg) (%)
NB	56.058	55.250	56.037	54.057
DT	64.620	64.727	64.713	64.583
RF	80.775	80.892	80.891	80.748
LSVM	81.422	81.658	81.492	81.411
NLSVM	85.784	85.815	85.844	85.757
KNN	70.436	73.958	71.292	69.717

Table 4 Performance comparison of machine learning under SIFT feature extraction

Algorithm	Accuracy (%)	Precision (avg) (%)	Recall (avg) (%)	F1score (avg) (%)
NB	55.250	55.420	55.540	55.438
DT	66.721	66.945	66.713	66.729
RF	89.176	89.195	89.099	89.094
LSVM	68.982	69.364	68.953	67.886
NLSVM	92.407	92.653	92.186	92.285
KNN	95.477	95.447	95.415	95.411

faults and complete faults in water pumps. This demonstrates the model's robust capability to accurately classify and predict the different states of pumping stations.

To illustrate the superiority of deep learning in image visual analysis over traditional machine learning methods, this article incorporates two feature extraction techniques—histogram of oriented gradient (HOG) and scale invariant feature transform (SIFT). These features are coupled with traditional machine learning methods, including naive bayes (NB), decision tree (DT), random forest (RF), linear support vector machine (LSVM), nonlinear support vector machine (NLSVM) and k-nearest neighbors (KNN). The resulting evaluation index scores are compared with those obtained from deep learning models, and the comparative results are presented in Tables 3 and 4, respectively.

The experimental findings indicate that, on the whole, SIFT exhibits superior feature extraction capabilities for pumping station images compared to HOG. However, when employing the LSVM machine learning method, HOG proves more suitable for feature extraction than SIFT. For NB and DT machine learning methods, there is minimal disparity in classification performance for different features within the same image; nonetheless, these methods demonstrate suboptimal overall effectiveness when applied to small-scale pumping station datasets. In contrast, RF, NLSVM and KNN algorithms exhibit superior classification performance for distinct features within the same image, particularly when employing SIFT. Notably, KNN achieves the highest accuracy at 95.477%, yet it still falls short compared to the accuracy achieved by deep learning models, which stands at 99.032%. Additionally, traditional machine learning methods for feature extraction are comparatively intricate when juxtaposed with deep learning. Consequently, the results suggest that deep learning is more suited for the visual analysis of small-scale pumping station images.

4 Conclusions

In this paper, we propose a visual analysis method based on dynamic platforms to address the challenges of monitoring unmanned pumping stations. We utilized transfer learning to train a more generalized ResNet18 for extracting pumping station image features. These features are transformed into sequence data and analyzed through LSTM to understand the contextual relationship between image sequences, resulting real-time monitoring and early warning of equipment status. This method outperforms traditional methods in terms of autonomy, performance and scalability. Our experiments demonstrate its effectiveness in unmanned pumping station analysis, offering significant support for water resource management.

Abbreviations

LSTM	Long short term memory
CNN	Convolutional neural network
GAN	Generative adversarial network
RNN	Recurrent neural network
HOG	Histogram of oriented gradient
SIFT	Scale Invariant Feature Transform
NB	Naive Bayes
DT	Decision tree
RF	Random forest
LSVM	Linear support vector machine
NLSVM	Nonlinear support vector machine
KNN	K-nearest neighbors

Acknowledgements

Not applicable.

Author contributions

ZL participated in the design and writing of the study, the authors supervised the study, suggested changes to the manual, and provided comments on revisions to the draft manuscript. SC and ZBZ contributed to the data. JHQ provided some comments on the manuscript. BP provided some comments on revisions to the manuscript. All authors have read and agreed to the manuscript.

Funding

This work was supported in part by the Project of Shenzhen University Stability Support Plan under Grant 20200829114939001, in part by the Project of Shenzhen Institute of Information Technology School-level Innovative Scientific Research Team under Grant TD2020E001 and in part by the Project of The Pearl River Delta water resources allocation engineering scientific research project (CD88-QT01-2022-0068).

Declarations

Availability of data and materials

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interests.

Received: 31 December 2023 Accepted: 21 February 2024

Published online: 29 February 2024

References

1. Y. Pokhrel, F. Felfelani, Y. Satoh et al., Global terrestrial water storage and drought severity under climate change. *Nat. Clim. Change* **11**(3), 226–233 (2021). <https://doi.org/10.1038/s41558-020-00972-w>
2. C. He, Z. Liu, J. Wu et al., Future global urban water scarcity and potential solutions. *Nat. Commun.* **12**(1), 4667 (2021). <https://doi.org/10.1038/s41467-021-25026-3>
3. X. Li, D. Long, B.R. Scanlon et al., Climate change threatens terrestrial water storage over the Tibetan Plateau. *Nat. Clim. Change* **12**(9), 801–807 (2022). <https://doi.org/10.1038/s41558-022-01443-0>
4. D.R. Rounce, R. Hock, F. Maussion et al., Global glacier change in the 21st century: every increase in temperature matters. *Science* **379**(6627), 78–83 (2023). <https://doi.org/10.1126/science.abo1324>
5. Z.H. Gong, X.H. Jiang, J.L. Cheng, Y. Gong, X. Chen, H.M. Cheng, Optimization method for joint operation of a double-reservoir-and-double-pumping-station system: a case study of Nanjing, China. *J. Water Supply Res. Technol. AQUA* **68**(8), 803–815 (2019). <https://doi.org/10.2166/aqua.2019.094>

6. A. Ahmad, A. El-Shafie, S.F.M. Razali, Z.S. Mohamad, Reservoir Optimization in Water Resources: a Review. *Water Resour. Manage* **28**(11), 3391–3405 (2014). <https://doi.org/10.1007/s11269-014-0700-5>
7. B. Durin, Some aspects of the operation work of pump station and water reservoir. *Period. Polytech. Civ. Eng.* **60**, 345–353 (2016). <https://doi.org/10.3311/PPci.7983>
8. J. Reca, A. Garcia-Manzano, J. Martinez, Optimal pumping scheduling model considering reservoir evaporation. *Agric. Water Manage.* **148**, 250–257 (2015). <https://doi.org/10.1016/j.agwat.2014.10.008>
9. S. Nabinejad, S. Jamshid Mousavi, J.H. Kim, Sustainable basin-scale water allocation with hydrologic state-dependent multi-reservoir operation rules. *Water Resour. Manage* **31**(11), 3507–3526 (2017). <https://doi.org/10.1007/s11269-017-1681-y>
10. X. Liu, B. Lai, B. Lin, V.C.M. Leung, Joint communication and trajectory optimization for multi-UAV enabled mobile internet of vehicles. *IEEE Trans. Intell. Transp. Syst.* **23**(9), 15354–15366 (2022). <https://doi.org/10.1109/TITS.2022.3140357>
11. X. Liu, Z. Liu, B. Lai et al., Fair energy-efficient resource optimization for multi-UAV enabled Internet of Things. *IEEE Trans. Veh. Technol.* **72**(3), 3962–3972 (2022). <https://doi.org/10.1109/TVT.2022.3219613>
12. X. Liu, Y. Yu, B. Peng et al., RIS-UAV enabled worst-case downlink secrecy rate maximization for mobile vehicles. *IEEE Trans. Veh. Technol.* **72**(5), 6129–6141 (2022). <https://doi.org/10.1109/TVT.2022.3231376>
13. L.A. Gama-Moreno, A. Corralejo, A. Ramirez-Molina et al., A Design of a Water Tanks Monitoring System Based on Mobile Devices. Paper presented at the 2016 International Conference on Mechatronics, Electronics and Automotive Engineering (ICMEAE), Cuernavaca, Mexico, 22–25 November 2016
14. B.N. Getu, H.A. Attia, Automatic water level sensor and controller system. Paper presented at the 2016 5th International Conference on Electronic Devices, Systems and Applications (ICEDSA), Ras Al Khaimah, United Arab Emirates, 6–8 December 2016
15. P.A. Apte, S.B. Naseem, IOT based Research Proposal on Water Pump Automation System for Turbidity, Pipeline Leakage and Fluid Level Monitoring. Paper presented at the 2022 5th International Conference on Advances in Science and Technology (ICAST), Mumbai, India, 2–3 December 2022
16. O.A. Klovov, A.A. Pushkina, Modernization of the Electric Drive and Automation System of the Sewage Pumping Station. Paper presented at the 2020 International Multi-Conference on Industrial Engineering and Modern Technologies (FarEastCon), Vladivostok, Russia, 6–9 October 2020
17. S.P. Tlabu, A. Telukdarie, B.G. Mwanza, Maintenance 4.0 for Water Pumping Infrastructures. Paper presented at the 2022 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM), Kuala Lumpur, Malaysia, 7–10 December 2022
18. R. Mahjoub, A Smart Control and Monitoring of a Pumping System. Paper presented at the 2021 International Conference Design and Modeling of Mechanical Systems, Hammamet, Tunisia, 20–22 December 2021
19. G.E. Hinton, R.R. Salakhutdinov, Reducing the dimensionality of data with neural networks. *Sci.* **313**(5786), 504–507 (2006). <https://doi.org/10.1126/science.1127647>
20. A. Krizhevsky, I. Sutskever, G.E. Hinton, ImageNet classification with deep convolutional neural networks. *Commun. ACM* **60**(6), 84–90 (2017). <https://doi.org/10.1145/3065386>
21. K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition. Paper presented at the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, USA, 27–30 June 2016
22. C. Szegedy, S. Ioffe, V. Vanhoucke, A.A. Alemi, Inception-v4, inception-resnet and the impact of residual connections on learning. Paper presented at the 31st AAAI Conference on Artificial Intelligence, San Francisco, USA, 4–9 February, 2017
23. I.J. Goodfellow, J. Pouget-Abadie, M. Mirza et al., Generative Adversarial Nets. Paper presented at the 28th Conference on Neural Information Processing Systems (NIPS), Montreal, Canada, 8–13 December, 2014
24. D.E. Rumelhart, G.E. Hinton, R.J. Williams, Learning representations by back-propagating errors. *Nature* **323**, 533–536 (1986). <https://doi.org/10.1038/323533a0>
25. J.L. Elman, Finding structure in time. *Cognit. Sci.* **14**(2), 179–211 (1990). [https://doi.org/10.1016/0364-0213\(90\)90002-E](https://doi.org/10.1016/0364-0213(90)90002-E)
26. M.I. Jordan, Chapter 25 - Serial Order: A Parallel Distributed Processing Approach, ed. by JW Donahoe. *Advances in Psychology*, vol 121 (Elsevier, North-Holland, 1997), p. 471–495
27. S. Hochreiter, J. Schmidhuber, Long short-term memory. *Neural Comput.* **9**(8), 1735–1780 (1997). <https://doi.org/10.1162/neco.1997.9.8.1735>
28. O. Ronneberger, P. Fischer, T. Brox, UNet: Convolutional Networks for Biomedical Image Segmentation. Paper presented at the 2015 Medical Image Computing and Computer-Assisted Intervention (MICCAI), Munich, Germany, 5–9 October 2015
29. A. Vaswani, N. Shazeer, N. Parmar et al., Attention Is All You Need. Paper presented at the 31st Annual Conference on Neural Information Processing Systems (NIPS), Long Beach, USA, 4–9 December, 2017
30. J. Yang, C. Wang, B. Jiang, H. Song, Q. Meng, Visual perception enabled industry intelligence: State of the Art, Challenges and Prospects. *IEEE Trans. Ind. Inf.* **17**(3), 2204–2219 (2021). <https://doi.org/10.1109/TII.2020.2998818>
31. N.N. Misra, Y. Dixit, A. Al-Mallahi, M.S. Bhullar, R. Upadhyay, A. Martynenko, IoT, Big Data, and Artificial Intelligence in Agriculture and Food Industry. *IEEE Internet Things J.* **9**(9), 6305–6324 (2022). <https://doi.org/10.1109/JIOT.2020.2998584>
32. A. Darko, A. Chan, M. Adabre et al., Artificial intelligence in the AEC industry: Scientometric analysis and visualization of research activities. *Autom. Constr.* (2020). <https://doi.org/10.1016/j.autcon.2020.103081>
33. B.I. Oluleye, D.W. Chan, P. Antwi-Afari, Adopting Artificial Intelligence for enhancing the implementation of systemic circularity in the construction industry: a critical review. *Sustain. Prod. Consum.* **35**, 509–524 (2023). <https://doi.org/10.1016/j.spc.2022.12.002>

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.