An Improved Deep Canonical Correlation Fusion Method for Underwater Multisource Data

KUIYONG SONG1,2, NIANBIN WANG1, AND YUN ZHANG1

1College of Computer Science and Technology, Harbin Engineering University, Harbin 150000, China
2Department of Information Engineering, Hulunbuir Vocational Technical College, Hulunbuir 021000, China

Corresponding author: Nianbin Wang (wangnianbin@hrbeu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61772152, in part by the National Key Research and Development Program of China under Grant 2018YFC0806000, in part by the Technical Basic Research Project under Grant SQB2017206C002, in part by the Pre-research Project under Grant 10201050201, in part by the Project funded by China Postdoctoral Science Foundation under Grant 019M651262, in part by the Youth Fund Project of Humanities and Social Sciences Research of the Ministry of Education of China under Grant 0YJCZH172, in part by the Postdoctoral Foundation of Heilongjiang Province under Grant BH-Z19015, and in part by the Fundamental Research Funds for the Central Universities under Grant 3072020CF0605.

ABSTRACT In complex underwater environments, the single mode of a single sensor cannot meet the precision requirement of object identification, and multisource fusion is currently the mainstream research approach. Deep canonical correlation analysis is an efficient feature fusion method but suffers from problems such as not strong scalability and low efficiency. Therefore, an improved deep canonical correlation analysis fusion method is proposed for underwater multisource sensor data containing noise. First, a denoising autoencoder is used for denoising and to reduce the data dimension to extract new feature expressions of raw data. Second, given that underwater acoustic data can be characterized as 1-dimensional time series, a 1-dimensional convolutional neural network is used to improve the deep canonical correlation analysis model, and multilayer convolution and pooling are implemented to decrease the number of parameters and increase the efficiency. To improve the scalability and robustness of the model, a stochastic decorrelation loss function is used to optimize the objective function, which reduces the algorithm complexity from $O(n^3)$ to $O(n^2)$. The comparison experiment of the proposed algorithm and other typical algorithms on MNIST containing noise and underwater multisource data in different scenes shows that the proposed algorithm is superior to others regardless of the efficiency or precision of target classification.

INDEX TERMS Convolutional neural network, deep canonical correlation analysis, denoising autoencoder, multisource fusion, underwater data.

I. INTRODUCTION

Underwater environments are complex and changing, and it is difficult to identify underwater objects [1], [2]. Recent years have experienced an increasing number of water vessels, and background noise from their frequent activities makes object identification more difficult. Single sensors are powerless under these conditions to meet the functional requirements of object identification. Multisource sensor fusion can efficiently improve the precision of object identification [3]–[5] by utilizing multisource information associated with the system function to conduct fusion and reasoning to combine the strengths of various types of information to obtain more accurate information than that obtained by a single sensor.

The associate editor coordinating the review of this manuscript and approving it for publication was Lefei Zhang.

In terms of fault detection [6]–[8], medical diagnosis [9], robot technology [10] and integrated navigation [11], multisensor fusion is widely researched and applied. According to the stages of data fusion, multisource data fusion can be divided into pixel-level fusion, feature-level fusion and decision-level fusion. Feature-level fusion is the process of feature extraction, comprehensive analysis and processing of multisource raw information. Typical feature fusion methods are based on data dimension reduction, such as principal component analysis (PCA) [12], linear discriminant analysis (LDA) [13], and linear locally embedding (LLE) [14]. In recent years, with the continuous deepening of deep learning research, feature-level fusion has been widely researched and has received wide attention. Models such as convolutional neural networks (CNNs), recurrent neural networks (RNNs) and long short-term memory (LSTM) have
been proposed and widely applied to feature extraction and fusion [15], [16]. As a traditional nonlinear feature extraction and fusion method, the autoencoder (AE) has also received attention [17], [18]. Many autoencoder models have been proposed, such as the basic autoencoder (BAE), denoising autoencoder (DAE), contractive autoencoder (CAE) and robust autoencoder (RAE).

Canonical correlation analysis (CCA), as a statistical method of calculating the maximum correlation of two sets of variables, is widely researched and applied in data fusion [19]. Kernel CCA (KCCA), the nonlinear expansion of CCA, is a nonparametric method that is subject to poor scalability and efficiency [20]. Deep CCA (DCCA) is a CCA method based on a deep neural network that uses two sets of variables to train a neural network model to make them have the maximum correlation [21]. Compared to CCA, is a nonparametric method that is subject to poor scalability and efficiency [20]. Deep CCA (DCCA) is a CCA method based on a deep neural network that uses two sets of variables to train a neural network model to make them have the maximum correlation [21]. Compared to CCA, is a nonparametric method that is subject to poor scalability and efficiency [20]. Deep CCA (DCCA) is a CCA method based on a deep neural network that uses two sets of variables to train a neural network model to make them have the maximum correlation [21]. Compared to CCA, is a nonparametric method that is subject to poor scalability and efficiency [20]. Deep CCA (DCCA) is a CCA method based on a deep neural network that uses two sets of variables to train a neural network model to make them have the maximum correlation [21].

The use of deep neural networks by DCCA leads to lower efficiency. Compared to DNN, the network structure of CNN weight sharing significantly reduces the complexity of the model and reduces the number of weights [15]. The convolution operation extracts deep-level features of the data, and the pooling operation reduces the number of model parameters and reduces the complexity of the network. On the other hand, DCCA is computationally expensive due to the matrix inversion or SVD operations required for each decorrelation in each training iteration. In the soft CCA algorithm [22], exact decorrelation is replaced by soft decorrelation via a mini-batch-based stochastic decorrelation loss (SDL) to be optimized jointly with the other training objectives. Soft decorrelation is more efficient and scalable by avoiding computationally expensive operations such as matrix inversion and SVD.

In this paper, an improved DCCA (IDCCA) is developed for underwater multisource data fusion. We introduce the 1-dimensional CNN (1DCNN) and SDL function into IDCCA. The main contributions of this paper can be summarized as follows.

1) To address the substantial noise and high dimensionality of underwater multisource experimental data, DAE is used to eliminate noise, reduce the dimensionality and learn new feature expressions of data.

2) To improve the efficiency and model scalability of DCCA, a 1DCNN is used to improve DCCA. Multi-layer convolution and pooling reduce the number of parameters in the training models and increase the efficiency. To improve the scalability and robustness of the model, an SDL function is used to optimize the objective function, which reduces the algorithm complexity from $O(n^5)$ to $O(n^2)$.

3) An experimental comparison with CCA, KCCA, soft CCA, and DCCA on the MNIST dataset containing noise, anechoic water tank data and authentic underwater monitoring data shows that the proposed method achieves high classification accuracy.

This thesis consists of 5 parts. In the second part, the AE, CCA, and DCCA algorithms are introduced. In the third part, the algorithm proposed in this thesis is introduced. The fourth part presents the experimental comparison analysis, and the fifth part is the conclusion.

II. RELATED WORK

A. AUTOENCODER

Autoencoders are neural network models with symmetric structures. Fig. 1 shows a 3-layer autoencoder with an input layer, hidden layer and output layer. The number of nodes of the input layer is the same as that of the output layer and more than that of the hidden layer. The hidden layer recodes the input data, and the output layer is the reconstitution of the hidden layer. Given a training sample $x = [x_1, x_2, \ldots, x_N]^T \in \mathbb{R}^N$, the encoding function $f$ and decoding function $g$ of the autoencoder in Fig. 1 are as follows:

$$y = f(x) = s_1(w^{(1)}_x + b^{(1)}),$$

$$(1)$$

$$z = g(y) = s_2(w^{(2)}_y f(x) + b^{(2)}),$$

$$(2)$$

where $w^{(1)}$ and $w^{(2)}$ are weight matrices, $b^{(1)}$ and $b^{(2)}$ are bias vectors, and $s_1$ and $s_2$ are nonlinear activation functions. The objective of an AE is to optimize $w$ and $b$ to minimize the reconstruction error. Traditionally, the mean squared error or cross entropy is used to compute the reconstruction error. The mean squared error function is $l_{MSE}(x, z) = \|x - z\|_2^2$, and the cross entropy function is

$$l_{CE}(x, z) = -\sum_{k=1}^{d} x_k \log z_k + (1 - x_k) \log(1 - z_k)$$

$$(3)$$

B. DEEP CANONICAL CORRELATION ANALYSIS

Canonical correlation analysis (CCA) [23] originates from multivariate statistical analysis. Given two sets of variables $x = [x_1, x_2, \ldots, x_N] \in \mathbb{R}^{d \times N}$ and $y = [y_1, y_2, \ldots, y_N] \in \mathbb{R}^{d \times N}$, each pair $(x_i, y_i)$ is a data sample including two modalities. CCA aims to find two sets of basis vectors $w_x$ and $w_y$ to map multimodal data into a shared $d$-dimensional subspace, that is, $p_x = w_x^T x$ and $p_y = w_y^T y$, where $x$ and $y$ have been normalized to mean 0 and variance 1. The objective function can be written as (4), where $\rho$ denotes the correlation

\[
\rho(x, y) = \frac{\text{cov}(x, y)}{\sqrt{\text{var}(x) \text{var}(y)}}
\]
FIGURE 2. The structure of DCCA.

coefficient and \( C \) denotes the covariance matrix.

\[
\rho = \max_{w_x, w_y} \frac{w_x^T c_{xy} w_y^T}{\sqrt{(w_x^T c_{xx} w_x)(w_y^T c_{yy} w_y)}} \tag{4}
\]

Because \( \rho \) is invariant to the scale of \( w_x \) or \( w_y \), and \( (4) \) is further improved to be the constrained optimization for \( (5) \).

\[
\max w_x^T c_{xy} w_y^T \quad s.t. \quad w_x^T c_{xx} w_x = 1, \ w_y^T c_{yy} w_y = 1 \tag{5}
\]

CCA projects two sets of variables into linear space, while DCCA [21] projects data into nonlinear space utilizing deep neural networks. DCCA consists of two deep neural networks whose output layers have the maximum correlation. Let \( H_x = f_x(x, \theta_x) \) and \( H_y = f_y(y, \theta_y) \) be nonlinear transformation functions implemented by a neural network that maps \( x \) and \( y \) into a shared subspace. The optimization objective is to maximize the cross-modality correlation between \( H_x \) and \( H_y \) formulated as follows [15]:

\[
\max \text{corr}(H_x, H_y) = \max \text{corr}(f_x(x, \theta_x), f_y(y, \theta_y)) \tag{6}
\]

Fig. 2 shows a DCCA structure composed of two 3-layer neural networks in which the output layer has the greatest correlation. The neural network has 3 nodes in the input layer, 4 nodes in the hidden layer, and 2 nodes in the output layer.

III. IMPROVED DEEP CANONICAL CORRELATION ANALYSIS

To improve the underwater object identification rate in noisy environments, an improved deep canonical correlation analysis (IDCCA) is proposed. Fig. 3 shows the overall structure of the proposed algorithm. The left part of Fig. 3 uses DAE to extract features, where red and blue circles are new features extracted, and these feature data are sent to the right part of Fig. 3 for feature fusion using IDCCA.

A. DAE FEATURE EXTRACTION

A denoising autoencoder (DAE) [24] can reconstruct raw data from partially destroyed samples; that is, DAE can construct a potential feature space for corrupted data and has good robustness to corrupted inputs. With a similar structure to the common autoencoder, the DAE is different in the use of corrupted data in the input stage, and it randomly selects a certain number of input nodes and sets the node value to 0. The DAE is thus trained to guess missing values. The reconstructions of the AE are, however, compared to the original, uncorrupted inputs.

If a raw datum is \( x \), then \( \tilde{x} \) is a datum containing noise. Similar to the basic autoencoder, the objective function for DAE is defined as:

\[
l_{DAE}(x, \tilde{z}) = -\sum_{k=1}^{d} x_k \log \tilde{z}_k + (1 - x_k) \log(1 - \tilde{z}_k) \tag{7}
\]

With

\[
\tilde{z} = g(\tilde{y}) = s_2(w^{(2)}f(\tilde{x}) + b^{(2)}) \tag{8}
\]

\[
\tilde{y} = f(\tilde{x}) = s_1(w^{(1)}\tilde{x} + b^{(1)}) \tag{9}
\]

\[
\tilde{x} = x + N(0, \sigma^2I) \tag{10}
\]

where \( w \) and \( b \) represent network parameters, noise datum \( \tilde{x} \) is subject to a Gaussian distribution, and \( \sigma \) is noise-level parameter that represents the degree of corruption.

A 3-layer DAE architecture has a simple structure and good performance [8]. The 3-layer structure shown in Fig. 1 is used to extract data features. In the whole algorithm, DAE is on the left side of Fig. 3, where corrupted data are input and optimized parameters are set. The red and blue packed data in the middle are new feature expressions. Further fusion of these new features is performed via IDCCA.

B. THE IMPROVED DEEP CANONICAL CORRELATION ANALYSIS

An IDCCA that utilizes two 1DCNN models to train network parameters is proposed. To improve the scalability and robustness of the model, SDL is used to optimize the objective function. The basic structure is shown on the right side of Fig. 3.

A CNN is a kind of neural network. The weight sharing network structure significantly reduces the complexity of the models and the number of weights. Convolution is a
We set one branch of a deep CCA network over a mini-batch as $Z \in R^{m \times k}$, where $m$ is the mini-batch size and $k$ indicates the number of feature channels, and assume that $Z$ has been batch-standardized. The mini-batch covariance matrix $C_{\text{mini}}^t$ for the $t$-th training step is then given as $C_{\text{mini}}^t = 1/(m-1)Z^T Z$. We first compute an accumulative covariance matrix for $C_{\text{accu}}^t = \alpha C_{\text{accu}}^{t-1} + C_{\text{mini}}^t$, where $\alpha \in [0, 1)$ is a decay rate and $C_{\text{accu}}^{t=0}$ is initialized with an all-zero matrix. A normalization factor is also computed accumulatively as $C_{\text{appx}}^t = C_{\text{appx}}^{t-1} + (c^t - 1)$, where $\mathcal{O}^t = c^t_{ij}$ is the element in $C_{\text{appx}}^t$ at $(i,j)$. $L_1$ loss is used here to encourage sparsity in the off-diagonal elements. SDL is soft because it only penalizes the correlation across activations instead of enforcing exact decorrelation. It will be jointly optimized with any other losses the model may have. With the proposed SDL, the constrained optimization problem in (6) can be reformulated as the following unconstrained objective:

$$L_{\text{SDL}} = \sum_{i=1}^{k} \sum_{j \neq i} \left| \phi_{ij} \right|^2$$

(12)

where $\mathcal{O}^t_{ij}$ is the element in $C_{\text{appx}}^t$ at $(i,j)$. $L_1$ loss is used here to encourage sparsity in the off-diagonal elements. SDL is soft because it only penalizes the correlation across activations instead of enforcing exact decorrelation. It will be jointly optimized with any other losses the model may have. With the proposed SDL, the constrained optimization problem in (6) can be reformulated as the following unconstrained objective:

$$\operatorname{arg} \min_{(\theta_1, \theta_2)} L_{\text{dist}}(f_1(x_1 : \theta_1), f_2(x_2 : \theta_2)) + \lambda (L_{\text{SDL}}(f_1(x_1 : \theta_1)) + L_{\text{SDL}}(f_2(x_2 : \theta_2)))$$

(13)

where $L_{\text{dist}}(f_1(x_1 : \theta_1), f_2(x_2 : \theta_2))$ is the $L_2$ distance and $\lambda$ weights the alignment versus decorrelation losses. The soft CCA architecture is illustrated in Fig. 5. Note that both SDL and $L_2$ loss are mini-batch-based losses. Therefore, soft CCA can be realized using standard SGD optimization for end-to-end learning. The overall computational complexity of SDL is $O(n^3)$. In contrast, the existing exact decorrelation computation has a complexity of $O(n^5)$ due to the use of SVD [22].

After the collection of multisource underwater data, data segmentation and standardized processing must be performed. Then, DAE is used for denoising and to extract features, IDCCA is used to fuse features, and SVM is used to classify the features.

## IV. EXPERIMENTAL ANALYSIS

In this experiment, Python2.7 is used for development under the frameworks of TensorFlow 2.0, theano1.0.4 and...
Keras 2.0.6. To verify the efficiency of the algorithm proposed in this thesis, the MNIST containing noise [20] and underwater experimental data are selected. Underwater experimental data include those in anechoic water tanks and authentic underwater environments, which are ideal underwater environments and authentic underwater environments.

A. INTRODUCTION TO DATASET

1) MNIST CONTAINING NOISE

MNIST is a typical handwriting dataset that contains 60K training pictures and 10K test pictures. The size of each picture is 28 × 28 pixels. Thesis [28] generates a more challenging two-view version of the MNIST dataset, as shown in Fig. 6. That is, the pixel values are first rescaled to [0, 1] and the images are then randomly rotated at angles uniformly sampled from [−π/4, π/4]. The resulting images are then used as view 1 inputs. For each view 1 image, we randomly select an image of the same identity (0-9) from the original dataset and add independent random noise uniformly sampled from [0, 1] to each pixel to obtain the corresponding view 2 sample.

2) ANECHOIC WATER TANK DATA

This experiment simulates an authentic underwater target sound environment by arranging different underwater target sound equipment in an anechoic water tank. The targets are placed across from the hydrophones, which are orderly arranged around the tank. There are 18 hydrophones in the array, and each of them corresponds to the same number signal. The end of the hydrophone is 0.5 meters from the bottom of the tank, the space between hydrophones is 0.25 meters, and the length of the array cable is 4.5 meters. The measurement time is 6 min, and the frequency band is 25.6 kHz. All the collected signals are voltage values. To achieve refined underwater target information perception, the experiment uses different power amplifiers to simulate the noise of different vessels; the vessel sailing powers are 20%, 50% and 80%.

According to the targets’ engine powers and speeds, this experiment classifies the targets into three categories. Target 1 has 4800 segments, target 2 has 10000 segments, and target 3 has 3470 segments. The percentage of the data in the training set is 72.5%, and that in the test set is 27.5%.

3) AUTHENTIC UNDERWATER MONITORING DATA

The data are collected by arranging several underwater target noise collectors in authentic water areas. The target vessels are sailed at a constant speed on the surface of the water at different powers to collect data under different sound conditions. In the experiment in the authentic underwater scene, the natural environment messages are temperature: 6°C; humidity: 39%; wind speed: 4 m/s; air pressure: 1.029*105 Pa; and water speed: 7 m/s.

The authentic water experimental data are shown in Table 1. The vessels are sailing at constant speeds of 8 HP and 12 HP, and the movement distance is 20-40 meters. The distance from vessels to sound sensors is 100-350 meters. The experiment classifies the target vessels into 8 categories according to HP and distance. At the beginning and end of the experiment, the environmental background noise is recorded.

B. EXPERIMENTAL VERIFICATION AND COMPARISON ANALYSIS

1) EXPERIMENTAL ANALYSIS ON MNIST

The DAE model parameters include input dimensions, encoding dimensions, weight decay and activation function. In this experiment, the number input dimensions is 784. The encoding layer compresses and extracts the features according to 75% and 50% of the input dimensions. If the encoding dimensions are set to 75% of the input dimensions, then the number of encoding dimensions is 588; if the encoding dimensions are set to 50% of the input dimensions, then the number is 392. Weight decay is TRUE, and the activation function is tanh.

The classification comparison analysis of the IDCCA model, CCA, DCCA, and soft CCA is performed on MNIST containing noise. Before fusion, two kinds of data are used:

| NO. | Working condition | Movement distance (M) |
|-----|-------------------|-----------------------|
| 1   | 8 Horsepower-100 m| 80-110                |
| 2   | 8 Horsepower-160 m| 150-180               |
| 3   | 8 Horsepower-200 m| 200-221               |
| 4   | 8 Horsepower-260 m| 257-266               |
| 5   | 8 Horsepower-350 m| 340-360               |
| 6   | 12 Horsepower-350 m| 340-360              |
| 7   | 12 Horsepower-260 m| 220-255             |
| 8   | 12 Horsepower-150 m| 130-152            |
TABLE 2. Accuracy of different methods on MNIST.

| NO  | Algorithm | Classification accuracy |
|-----|-----------|-------------------------|
| 1   | CCA       | 81.2%                   |
| 2   | DCCA      | 94.4%                   |
| 3   | Soft CCA  | 95.2%                   |
| 4   | IDCCA     | 97.4%                   |
| 5   | DAE(50%)+DCCA | 96.8%          |
| 6   | DAE(75%)+DCCA | 97.2%         |
| 7   | DAE(50%)+Soft CCA | 97.1%       |
| 8   | DAE(75%)+Soft CCA | 97.5%       |
| 9   | DAE(50%)+IDCCA | 98.1%     |
| 10  | DAE(75%)+IDCCA | **98.5%** |

Table 2 indicates that IDCCA outperforms soft CCA, DCCA, and CCA overall. The classification accuracy of 81.2% of CCA is the lowest among all methods. The accuracy of fusion classification with DCCA, soft CCA, and IDCCA on data processed by DAE is higher than that on data not processed by DAE; thus, the classification performance is improved after denoising by DAE. The classification accuracy with DCCA is 96.8% and 97.2% for DAE (50%) and DAE (75%), respectively, the classification accuracy with soft CCA is 97.1% and 97.5% for DAE (50%) and DAE (75%), respectively, and the classification accuracy with IDCCA is 98.1% and 98.5% for DAE (50%) and DAE (75%), respectively. The classification accuracy with DAE (75%) is higher than that with DAE (50%). Among all algorithms, the highest accuracy of 98.5% is achieved with DAE(75%)+IDCCA.

2) UNDERWATER DATA EXPERIMENTAL ANALYSIS

The data should be split before the use of hydrophones because of the large amount of data to be detected. The split size is based on powers of 2: in this thesis, 512, 1024, and 2048. Then, a preprocessing function is used to normalize the data to [0, 1]. Last, the dataset is split into a training set and a test set with a proportion of 6:4.

The input dimensions of the DAE are 512, 1024 and 2048, respectively; the encoding dimensions are 50% and 75% of the input dimensions. If the encoding dimensions are set to be 75% of the input dimensions, then the numbers of encoding dimensions correspond to 384, 768, and 1536, respectively. If the encoding dimensions are set to be 50% of the input dimensions, then the numbers of encoding dimensions correspond to 256, 512, and 1024, respectively. Weight decay is TRUE, and the activation function is tanh.

To further improve the classification performance, CCA, DCCA, soft CCA, and IDCCA are used to fuse the data, and SVM is then used to classify the data. The classification results are shown in Table 3.

Table 3 shows that the accuracy after feature extraction by DAE is higher than that without DAE. After extraction by DAE, the accuracy of data with a compression proportion of 75% is higher than that with a compression proportion of 50%, and the accuracy with IDCCA is higher than that with CCA, DCCA, and soft CCA. The highest data classification accuracy is 98.9% for anechoic tank data and 90.8% for the authentic underwater data with DAE (75%)+IDCCA. Table 3 shows that the classification accuracy after fusion by CCA, DCCA, soft CCA and IDCCA is improved and that the effect of fusion is obvious. Then, comparison analysis is performed via a histogram.

From raw data to DAE (50%) to DAE (75%), Fig. 7(a) shows that the classification accuracy of data after the fusion with IDCCA is highest, the second highest is with soft CCA, and the lowest is with CCA. The results shown in Fig. 7(b) are similar to those in Fig. 7(b). However, owing to the noise, the accuracy for authentic water data is lower than that in...
FIGURE 8. Classification accuracy of different classification methods.

FIGURE 9. Comparison of the algorithm time complexity before and after processing by DAE.

FIGURE 10. Comparison of the efficiency of CCA, DCCA, soft CCA and IDCCA without data.

anechoic tanks overall. Fig. 7 shows that the classification accuracy is improved by DAE, and the classification accuracy with DAE (75%) is higher than that with DAE (50%).

In addition, comparison of the classification accuracy of soft CCA and IDCCA in Table 3 and Fig. 7 indicates that regardless of the raw data, DAE (50%) or DAE (75%), IDCCA is better than soft CCA. IDCCA and soft CCA use the same objective function: the difference is that IDCCA is trained with a 1DCNN and soft CCA is trained with a deep neural network. In this case, the classification accuracy of the model trained with a 1DCNN is better.

To further verify the classification accuracy, SVM, LNN, and MLP are used to compare the classification accuracy of MNIST, anechoic water tank data (AWTD), and authentic underwater monitoring data (AUMD). The AWTD and AUMD were processed by DAE (75%). The classification results of the three classification algorithms are shown in Fig. 8.

As seen from Fig. 8, LNN and MLP obtain similar results to SVM, but SVM achieves the best classification accuracy, followed by MLP and then LNN.

3) TIME COMPLEXITY

After processing of the experimental data by DAE, the raw data are compressed and reduced in dimension; thus, the data dimension is reduced. Therefore, the time required for model training and data classification is reduced. A comparison of the fusion classification efficiency of the raw data, DAE (50%) data and DAE (75%) data is shown in Fig. 9.

Fig. 9 shows that raw data require the most time in the model training stage. DAE (75%) considers three-quarters of the raw data, and DAE (50%) considers half of the raw data. In the model test stage, similar results were obtained to those in the model training stage; that is, a higher compression proportion of data requires less time.

To compare the efficiency of CCA, DCCA, soft CCA, and IDCCA, comparative analysis was carried out on MNIST containing noise, anechoic water tank data and authentic underwater data: the results are shown in Fig. 10.

Fig. 10 shows that the three MNIST datasets containing noise, anechoic water tank data and authentic underwater data show the same trend: the efficiency of CCA is lower than that of DCCA, and the efficiency of DCCA is lower than that of IDCCA. Although soft CCA and IDCCA use the same loss function, the efficiency of deep neural networks is not as good as that of CNNs, so the efficiency of soft CCA is lower than that of IDCCA. IDCCA uses 1DCNN to train the network, which reduces the number of parameters, and uses the SDL function to reduce the time complexity of the optimization function from $O(n^3)$ to $O(n^2)$. Thus, the efficiency of IDCCA is the highest.

V. CONCLUSION

To assess the efficiency of the fusion algorithm, the project group has conducted numerous experiments in different scenarios. To address underwater multisource data containing noise, an IDCCA fusion method is proposed in this thesis. The method uses a denoising autoencoder for denoising and to reduce the data dimension and extract new feature expressions of the raw data. Given the features of a 1-dimensional time series of underwater acoustic data, a 1DCNN is used to improve the deep canonical correlation model. To improve the scalability and robustness of the model, an SDL function is used to optimize the objective function, which reduces the algorithm complexity. The tests of underwater data in different scenes verify the efficiency of the algorithm proposed in this thesis. There is space for improvement in model efficiency, and future work will be done in target optimization.
to improve the algorithm efficiency. Moreover, the algorithm in this paper uses DAE to extract features and IDCCA to perform feature fusion, which leads to the separation of feature extraction and feature fusion. We hope to combine feature extraction and feature fusion in the next step and believe that improved results will be achieved.

REFERENCES

[1] A. Pezeshki, M. R. Azimi-Sadjadi, L. L. Scharf, and M. Robinson, “Underwater target classification using canonical correlations,” in Proc. Oceania. Celebrating Past Teaming Toward Future, Sept. 2003, pp. 1906–1911.

[2] A. Pezeshki, M. R. Azimi-Sadjadi, and L. L. Scharf, “Undersea target classification using canonical correlation analysis,” IEEE J. Ocean. Eng., vol. 32, no. 4, pp. 948–955, 2007.

[3] P. K. Atrey, M. A. Hossain, A. El Saddik, and M. S. Kankanahalli, “Multimodal fusion for multimedia analysis: A survey,” Multimedia Syst., vol. 16, no. 6, pp. 345–379, Nov. 2010.

[4] T. Baltrusaitis, C. Ahuja, and L.-P. Morency, “Multimodal machine learning: A survey and taxonomy,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 41, no. 2, pp. 423–443, Feb. 2019.

[5] H. Wang, W. Yang, Z. Lin, and Y. Yu, “TMDA: Task-specific multi-source domain adaptation via clustering embedded adversarial training,” in Proc. IEEE Int. Conf. Data Mining (ICDM), Nov. 2019, pp. 1372–1377.

[6] H. Hua, H. Jiang, F. Wang, and H. Zhao, “An enhancement deep feature fusion method for rotating machinery fault diagnosis,” Knowl.-Based Syst., vol. 119, pp. 200–220, Mar. 2017.

[7] Y. Wu, B. Jiang, and N. Lu, “A descriptor system approach for estimation of incipient faults with application to high-speed railway traction devices,” IEEE Trans. Syst., Man, Cybern., vol. 49, no. 10, pp. 2108–2118, Oct. 2019.

[8] Y. Wu, B. Jiang, and Y. Wang, “Incipient winding fault detection and diagnosis for squirrel-cage induction motors equipped on CRH trains,” ISA Trans., vol. 99, pp. 488–495, Apr. 2020.

[9] A. Cheerla and O. Gevaert, “Deep learning with multimodal representation for pancancer prognosis prediction,” Bioinformatics, vol. 35, no. 14, pp. 1446–1454, Jul. 2019.

[10] E. Sheppard and K. S. Lohan, “Multimodal representation learning for robot human interaction,” in Proc. Companion ACM/IEEE Int. Conf. Hum.-Robot Interact., Mar. 2020, pp. 445–446.

[11] W. H. Chin, Y. Toda, N. Kubota, C. K. Loo, and M. Seera, “Episodic memory multimodal learning for robot sensorimotor map building and navigation,” IEEE Trans. Cognit. Develop. Syst., vol. 11, no. 2, pp. 210–220, Jun. 2019.

[12] S. Wold, K. Esbensen, and P. Geladi, “Principal component analysis,” Chemometrics Intell. Lab. Syst., vol. 2, nos. 1–3, pp. 37–52, 1987.

[13] H. Xie, J. Li, and H. Xue, “A survey of dimensionality reduction techniques based on random projection,” 2017, arXiv:1706.04371. [Online]. Available: http://arxiv.org/abs/1706.04371

[14] Y. M. Winlaw, L. S. Dehkordy, and A. Ghodsi, “Robust locally linear embedding using penalty functions,” in Proc. Int. Joint Conf. Neural Net., Jul. 2011, pp. 2305–2312.

[15] W. Guo, J. Wang, and S. Wang, “Deep multimodal representation learning: A survey,” IEEE Access, vol. 7, pp. 63373–63394, 2019.

[16] Z. Wang, B. Du, and Y. Guo, “Domain adaptation with neural embedding matching,” IEEE Trans. Neural Netw. Learn. Syst., vol. 31, no. 7, pp. 2387–2397, Jul. 2020.

[17] D. Charte, F. Charte, S. García, M. J. del Jesus, and F. Herrera, “A practical tutorial on autoencoders for nonlinear feature fusion: Taxonomy, models, software and guidelines,” Inf. Fusion, vol. 44, pp. 78–96, Nov. 2018.

[18] F. J. Pulgar, F. Charte, A. J. Rivero, and M. J. del Jesus, “Choosing the proper autoencoder for feature fusion based on data complexity and classifiers: Analysis, tips and guidelines,” Inf. Fusion, vol. 54, pp. 44–60, Feb. 2020.

[19] B. Thompson, J. Cartmill, M. R. Azimi-Sadjadi, and S. G. Schock, “A multichannel canonical correlation analysis feature extraction with application to buried underwater target classification,” in Proc. SPIE Int. Joint Conf. Neural Netw., Jun. 2019, pp. 4413–4420.

[20] S. Akaho, “A kernel method for canonical correlation analysis,” in Proc. Int. Meeting Psychometric Soc., 2001, pp. 1–5.

[21] G. Andrew, R. Arora, J. Bilmes, and K. Livescu, “Deep canonical correlation analysis,” in Proc. ICML, 2013, pp. 1247–1255.

[22] X. Chang, T. Xiang, and T. M. Hospedales, “Scalable and effective deep CCA via soft decorrelation,” in Proc. IEEE/CVF Comput. Vis. Pattern Recognit., Jun. 2018, pp. 1488–1497.

[23] B. Thompson, “Canonical correlation analysis,” in Encyclopedia of Statistics in Behavioral Science. Hoboken, NJ, USA: Wiley, 2005. [Online]. Available: https://doi.org/10.1002/0470013192.bsa068

[24] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P.-A. Manzagol, “Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion,” J. Mach. Learn. Res., vol. 11, no. 12, pp. 3371–3408, Dec. 2010.

[25] H. Kianyaz, T. Ince, and M. Gabbouj, “Real-time patient-specific ECG classification by 1-D convolutional neural networks,” IEEE Trans. Biomed. Eng., vol. 63, no. 3, pp. 664–675, Mar. 2016.

[26] W. Wang, R. Arora, K. Livescu, and J. A. Bilmes, “Unsupervised learning of acoustic features via deep canonical correlation analysis,” in Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP), Apr. 2015, pp. 4590–4594.

[27] F. Yan and K. Mikolajczyk, “Deep correlation for matching images and text,” in Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2015, pp. 3441–3450.

[28] W. Wang, R. Arora, K. Livescu, and J. Bilmes, “On deep multi-view representation learning,” in Proc. Int. Conf. Mach. Learn., Jun. 2015, pp. 1083–1092.

KUIYONG SONG received the B.E. degree in computer science and technology from Liaoning Normal University, China, in 2011. He is currently pursuing the Ph.D. degree with the School of Computer Science and Technology, Harbin Engineering University. He is also an Associate Professor with the Department of Information Engineering, Hulunbuir Vocational Technical College, China. His major research interests include deep learning and information fusion.

NIANBIN WANG received the Ph.D. degree in computer science and technology from the Harbin Institute of Technology, China, in 2001. He is a member of CCF, and he has been a Professor with the Department of Computer Science and Technology, Harbin Engineering University. His interests include datasets, deep learning, and data integration.

YUN ZHANG was born in 1988. He received the B.E. and M.E. degrees in computer science from the Harbin University of Science and Technology, China. He is currently pursuing the Ph.D. degree with the College of Computer Science and Technology, Harbin Engineering University, China. His main research interests include pattern recognition, domain adaptation, transfer learning, and image processing.