CSRS: A Chinese Seal Recognition System With Multi-Task Learning and Automatic Background Generation

ZHENYU WANG¹, JIE LIAN¹, CHUNFENG SONG², WEI ZHENG², SHAOYONG YUE¹, AND SENRONG JI¹

¹School of Control and Computer Engineering, North China Electric Power University, Beijing 102206, China
²National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing 100080, China

Corresponding author: Zhenyu Wang (zywang@ncepu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61573139, and in part by the Fundamental Research Funds for the Central Universities under Grant 2018ZD05.

ABSTRACT

As an important part of the Chinese painting and calligraphy, the seals not only have a high value of art but also contain a lot of information about the artwork itself. At this digital age, we would like not only to be able to represent the seals in the digital format, but also like to use image processing techniques to help us better understand them. With the development of deep learning, convolutional neural networks have been widely used in the fields of feature learning, object localization, and classification. Based on deep learning technology, this paper proposes a highly accurate Chinese seal recognition system (CSRS). With our CSRS, users could simply input a single seal image into the system, then CSRS would automatically recognize the seal and report the relevant information in real-time. The CSRS mainly contains three units. 1) A new Siamese network with multi-task learning (Siamese-MTL), which can effectively solve the similarity measurement problem and improve the generalization of the model. 2) A new online data generation algorithm called automatic background generation (ABG) which could generate numerous seal images with different backgrounds for effective training. 3) A new training method for Siamese network which based on a central constraint. In order to validate the effectiveness of the proposed method, we have established two large scale seals image databases, including 15,000 Chinese seal images and 1,700 background images, respectively. We evaluate our method and compare with the variant methods on these datasets, achieving the highest performance. The extensive experimental results indicate that our proposed method is effective and has a great potential for the practical application in Chinese seals recognition.

INDEX TERMS

Multi-task learning, Siamese network, Chinese seal recognition.

I. INTRODUCTION

As an important part of Traditional Chinese painting (TCP) and calligraphy, the seals not only have a high value of art, but also contain a lot of information about the artwork itself. As a special form of art, the seals play an important role in helping people appreciate and learn about the culture of different dynasties. In this paper, we develop a highly accurate Chinese seals recognition system named CSRS (Figure 1). First, we input the training samples into the CSRS, and train the recognition model with these data. Second, we input a new seal database into the CSRS and generate a feature database. Finally, users only need to input a single seal into the system, then CSRS would automatically recognize the corresponding seal from the feature database. In particular, the Siamese-MTL and ABG are the two key units of our system.

There are few works related with the Chinese seals recognition in the most similar domains, such as artificial intelligence, computer vision and art analysis. Tan et al. [1] used the Spectral Angle Mapping (SAM) method to carry on the classification of the seal area. They combined the spectral imaging technology and SAM to investigate the similarities between spectral vectors at pixel level and recognized the seal area of painting and calligraphy. In the literature of [2], a seal image detection method was proposed, they employed both color and structure features to extract the seal image regions from the TCPs. Firstly, the original TCP image was transformed into the CIE Lab color space, then a template...
matching method was used to refine the initial candidate seal image regions. Leung [3] proposed an analysis method to better understand traditional Chinese seals, so that their model could describe them in a more semantic way. In addition, they adopted a synthesis method to generate new seal images. For a given handwritten character, their method could consider the information obtained from the analysis of the traditional Chinese seals.

By using morphology technique, Su [4] developed a Chinese seals classification algorithm. The proposed method could find the precise locations of the Chinese seals which contain the useful geometry features. With geometry features, their model could classify different Chinese seal images. Based on this work [4], the author proposed an improved method later in the literature of [5]. The new method could find the gray levels around the edge, then the entire edge gray levels were recorded in a vertical strip. Finally, they used the vector magnitude invariant transform technique to transfer the gray level quantity into an invariant vector magnitude quantity to recognize the object. In order to effectively categorize the seal images, Yu et al. [6] proposed a seal image recognition method based on Krawtchouk moment and radial basis function (RBF [7]) neural network. First, the Krawtchouk moment invariants for part of the standard seal images were calculated, which were then used to train the RBF neural network. Finally, the characteristics of all the questioned seal images were applied for reorganization. Different from the above works, Lee et al. described an overlay metric method [8] for verifying the authenticity of a seal impression imprinted on a document. The overlay metric is the ratio of an effective seal impression pattern and the noise in the neighborhood of the reference impression region.

Most existing works dealing with Chinese seals like as [1]–[6], [8], not much attention has been paid to develop the corresponding system to realize one-to-one seal recognition. Moreover, most of the works adopt the traditional image processing methods and do not apply the deep learning theory to the application. The CSRS developed in this paper is based on Siamese network, with high speed and good performance.

For the seal images, they have the characteristics of high similarity, it is difficult to identify them by using common image processing methods. Even for powerful CNN model, the final layer of general classification network is Soft-max layer, which is used for probability distribution estimation, while cannot measure the similarity of different data. The Siamese architecture could help learn a similarity metric from data instead of learning the discrete classes. Hence, the distance of different classes is enlarged and the distance of the same class becomes closer. In this paper, we combine the two architectures and propose a new seal recognition model named Siamese-MTL (Siamese Network Combined with Multi-task Learning). The loss of the proposed model consists of two parts, i.e., Softmax loss and Contrastive loss. Learning a similarity metric while increasing the distance between different classes, this mechanism can effectively solve the similarity measurement problem and improve the generalization of the model.

In our CSRS, each seal is treated as a separate class. The main challenge is that there exists a large number of classes, while only a small number of training examples.
is available. In order to cope with the challenges from the few sample learning and improve the feature extraction capability of the model, we propose a new online data generation method called ABG (Automatic Background Generation). While training, we only need to feed each subnetwork with one standard seal sample, then the one can generate multiple seal images with different backgrounds. Moreover, we add the data augmentation operation to each training sample to increase the diversity of data, such as rotation, brightness varying, adding noises and so on. In traditional Siamese network, the input data pairs of any channel are selected randomly. Different from above operation, based on the central constraint, we fix the input data of channels as default center, then randomly select the other input data for another channel. In this way, our model could learn features invariant to the backgrounds.

The main strategies of our work (Multi-task learning and ABG algorithm) are relevant to some methods in face recognition. Similar to the idea of multi-task learning, for face recognition, Sun et al. [9] proposed to use both face identification and verification signals as supervision. The face identification task used softmax loss to increase the inter-personal variations, while the face verification task used contrastive loss to reduce the intra-personal variations. The difference is that we train the model in stages, and they train the model directly by using the combination of two loss functions. Likewise, due to the specificity of the training data (little data in each class and special usage scenarios), Ding and Tao [10] used data augmentation strategy to increase the data diversity. Different from the way we used image combination for augmentation, they used the method of blurring images for generating video-like face data from existing large-scale still face image databases. The same is that we all encourage each original image and its enhanced version to be classified into the same class. In summary, although we have different working backgrounds, the ideas to solving the problems are similar.

The rest of the paper is organized as follows. In section II, we summarize the related work in the field of Siamese network. Then in section III, we introduce the new seal retrieval model Siamese-MTL and the new online data generation algorithm. In section IV, we present the experimental results. Finally we conclude this paper in section V.

II. RELATED WORKS

Traditional Siamese network usually consists of two identical subnetworks with shared weights to learn characteristic differences between images. Their subnetworks are multilayer convolutional neural networks. The advantages of such an architecture are the ability to automatically learn and extract relevant features. In addition, the number of trainable parameters is manageable and independent of the number of input data. Siamese network has improved the network learning capabilities by exploiting pair-wise information about how the training samples are related, with the abilities to minimize or enlarge the distances of pairs. Siamese networks are used in many fields such as the signature matching, image originality verification and so on.

In recent years, researchers put a lot of efforts to apply siamese network into different fields [11]–[13]. In the literature of [14], Juan et al. proposed a method that automatically detect the spinal metastasis in the magnetic resonance imaging (MRI). In order to accommodate the large variability in metastatic lesion sizes, they used a siamese deep neural network approach comprising three identical subnetworks for multi-resolution analysis and spinal metastasis detection. Sabri textitet et al. [15] used siamese and triplet networks for facial expression intensity estimation. Firstly, they extracted the sequential relationship in the temporal domain that appeared due to the natural onset and offset variations in the patterns of facial expressions, then the siamese and triplet networks were used to estimate the emotional intensity in facial image sequence. Siamese network also plays an important role in the field of visual object tracking. Jiang et al. [16] proposed an Ensemble Siamese Tracker (EST), which used convolutional neural network features and compared the features of recent frames with the target features in the first frame. The candidate region with the highest similarity score was considered as the tracking result, and tracking results in recent frames were used to adjust the model for continuous target changing. Similar to this work [16], in order to improve the performance of tracker, Han et al. [17] applied a Siamese region proposal network to identify potential targets across the whole frames. Region proposal network is able to mine hard negative examples to make the network more discriminative for the specific sequence.

Apart from the field of image processing, Siamese network is also widely applied in other fields, such as the natural language processing. Ichida et al. [18] developed a Siamese neural network architecture to measure the semantic similarity between two sentences through metric learning. Given a representation that encodes the semantic and syntactic information about the words, their proposed approach can measure the semantic similarity which not depend on the linguistic information of the sentences. In the paper of [19], authors proposed a methodology to investigate the singing style. The proposed approach utilized convolutional neural networks in a Siamese architecture, ResNet based convolutional blocks were used to process spectral inputs. The feed-forward attention layer was used to handle temporal dependencies and fully-connected dense layers to learn the non-linear embeddings. Inspired by deep learning techniques, Shaham and Lederman [20] applied Siamese neural networks to solve the problem of learning a common source of variability in data which are synchronously captured by multiple sensors. This approach is useful in the exploratory, data-driven applications, where neither a model nor the label information is available.

Liu et al. [21] also applied Siamese network to solve the problem of person re-identification. In many person re-id works [22]–[24], the models achieve the purpose of
person re-identification by judging the image similarity. This mechanism is very similar to our seal recognition. The main challenges of the person re-id are illumination and various application scenarios, in addition, the body movements of the person also have impacts on identification. In our work, the main problem is that different seals have highly similar textures, it will pose a challenge to recognition.

The mechanism of Siamese network is to utilize multiple parallel paths to find similarities or differences between inputs. In addition, the unified and clear objective function help learn the optimal metric towards the target automatically. Siamese networks have good applications in many fields, but few about art. In this paper, we introduce the Siamese network to art recognition task. Chinese seals have great values of art and seal recognition system has great commercial and application value. Existing works about Chinese seals do not pay much attention to develop the recognition system, the proposed CSRS fill this gap. The key challenge of seals recognition is to measure the similarity of different seals, especially for large-scale datasets, yet Siamese network has the advantage of solving this problem.

III. PROPOSED METHOD

Inspired by multi-task learning [25], [26], we present a new Siamese recognition network which is trained with margin Contrastive loss and Softmax loss. First, we train the network only with Softmax loss. When the model has the basic ability to extract features, we add Contrastive loss to train the model with two loss functions at the same time until the model converges. In order to improve the generalization of the model and cope with the few sample learning, we apply the Automatic Background Generation (ABG) unit to the training process. The key point of ABG is that we only need to feed each subnetwork with one original image, and do not need to generate a fixed number of images in advance. ABG would generate countless training data online. Different from the traditional way of feeding data, based on the central constraint, we fix the input data of channels as default center, then randomly select the other input data for another channel. In this way, our model could learn features invariant to the backgrounds.

A. SIAMESE-MTL ARCHITECTURE

Traditional learning algorithm is the single-task learning mode. For the complex learning tasks, researchers usually divide them into multiple unrelated tasks, then these tasks learn from each loss. Multi-task learning (MTL) can help learn the shared representations from multiple tasks. These shared representations have strong ability of adapting to multiple different but related goals, and enable the main task to gain better generalization ability [27]. Furthermore, in the case of few samples, the generalization ability of multi-task learning is obviously better than that of single-task learning. Multi-task learning can make better use of limited data and the training time is relatively short.

![FIGURE 2. Configurations of the Siamese-MTL (ours).](image-url)

Original Siamese network architecture ties two identical neural networks via a loss function at the last stage of the network. The model with this architecture does not need the class of the processed input, only need to know that two inputs are from the same class or different class. Seal images have the characteristic of high similarity, especially in our CSRS, it is usually need to deal with large-scale datasets. Using the traditional architecture makes training difficult. To this end, we develop a new Siamese network architecture as shown in Figure 2. The input images to our model have a fixed dimension of 128 × 128 × 3, where the third dimension represent RGB color channels. During training, a pair of images passes through the stacks of convolutional layers and max-pooling layers simultaneously. Then they are flattened to produce a pair of 1024-dimensional feature vectors that capture the very high level features from each image. At last layer, Softmax functions are used to estimate the probability distributions of each class. Moreover, feature vector 1 and feature vector 2 are feed into Contrastive loss function for the similarity measure learning. Our model can not only learn a similarity metric from data, but also increase feature distances between different classes at the same time.

In our model, convolutional layers are the core layers for feature learning and extraction. Each convolutional layer produces a feature map by convolving its input with a set of convolutional kernels. Max-pooling layers serve as a method of non-linear down-sampling, by providing a summary of the outputs of a set of neighboring elements in the corresponding feature maps [28]. It enable the combination of low-level features into high-level features.
Training Siamese-MTL can be roughly divided into two stages. Firstly, we only use Softmax loss to train the model and let network gain the basic feature extraction ability. Next, we add Contrastive loss and train the model with two loss function at same time until the model converge.

Softmax layer is used for multi-category classification [29]. In Softmax, score function gives a specific probability based on the final score, the sum of all categories’ probability is 1. The function form is shown in Equation 1.

\[
f(z) = \frac{e^{z_i}}{\sum_k e^{z_k}}
\]

where the input of Softmax is a vector \( z \), \( i \) represents the category index, and the total number of categories is \( k \), then the result is mapped from the exponent domain to the probability. Finally, the loss function can be regarded as the entropy of two probabilities, as shown in Equation 2.

\[
H(p, q) = -\sum_x p(x) \log(q(x))
\]

where \( p \) represents the probability of true classification, and \( q \) represents the probability of the predicted classification. The purpose of the loss function is to measure the size of the error between the true classification result and the predicted classification result, and then optimize and modify it based on this value.

A Siamese network learns the embedding by minimizing the contrastive loss [30]. In this paper, Contrastive loss is defined over the distance between the features from the two identical inner neural networks. Let \( x_1 \) and \( x_2 \) be the two inputs with the same dimensions, and two output features from the Siamese-MT are \( G(x_1) \) and \( G(x_2) \). The squared Euclidean distance can be denoted as

\[
D(x_1, x_2) = \|G(x_1) - G(x_2)\|_2^2
\]

We define the binary label \( y \in \{0, 1\} \) to be 0 meaning the inputs are not from the same category or 1 otherwise. Thus the Contrastive loss function designed for the architecture is defined as

\[
L = \frac{1}{2} y D + \frac{1}{2} (1 - y) \max(0, m - D)^2
\]

In the training phase, the loss of \( L \) will be minimized to encourage samples with the same identity be close to each other while the ones with different identities be pushed away from each other. \( m \) is the target margin between the embedded vectors having different identities [25].

As mentioned above, the total loss of the model has differences at different training stages. Generally, it can be divided into two stages as

\[
\begin{align*}
\text{Stage 1: } & \quad \text{loss}_{\text{total}} = \text{loss}_{\text{softmax}}^{\text{max}-1} + \text{loss}_{\text{softmax}}^{\text{max}-2} \\
\text{Stage 2: } & \quad \text{loss}_{\text{total}} = \text{loss}_{\text{con}} + \text{loss}_{\text{softmax}}^{\text{max}-1} + \text{loss}_{\text{softmax}}^{\text{max}-2}
\end{align*}
\]

where \( \text{loss}_{\text{softmax}}^{\text{max}-1} \) and \( \text{loss}_{\text{softmax}}^{\text{max}-2} \) represent the Softmax losses of the two subnetworks, respectively. \( \text{loss}_{\text{con}} \) represents the Contrastive loss of the Siamese network.

B. AUTOMATIC BACKGROUND GENERATION (ABG)
Seals usually exist in the traditional Chinese painting and calligraphy, which can be located via the general detection algorithms, then be recognized by classifiers. However, the detected seal images from the artwork usually have complex backgrounds, which will severely affect the performance of the recognition. On the other hand, in our system, each seal is treated as a separate class. There are large numbers of classes but only one standard seal image in the training set is available for each class. Based on above factors, we propose an Automatic Background Generation (ABG) algorithm to enlarge the diversity of training samples.

1) BACKGROUND DATABASE
We establish a background database which consists of 1,700 images with calligraphy and painting-like backgrounds. The examples of the background images are shown in Figure 3. While training, we only need to feed each subnetwork with one original image, then choose whether to use ABG. If using ABG, one seal image can automatically generate numerous new images with different backgrounds in each iteration. On the contrary, if not using ABG, original seal image is used as the training sample in this iteration. Based on above operation, we add the data augmentation operation to each training sample, which greatly enhance the diversity of the training data and prevent the problem of overfitting.

2) AUTOMATIC BACKGROUND GENERATION
Here we discuss how to generate different backgrounds automatically. With a selected original seal image, we would randomly select a background image from the background database. Next, we will determine the size of the background image and convert it to the same size as the seal image. The resize operation can be divided into two parts. If the background image is large enough, we only need to crop an area same as the size of the original image. Otherwise, we would calculate a resize ratio to resize the background image which adapts the same size of original image.

3) SEAL AND BACKGROUND IMAGES COMBINATION
The next step is the combination operation. Seal image has such characteristics as illustrated in figure 4 (a), we call them...
the carving techniques of Yin and Yang. First, we convert the original image to a binary grayscale one and then calculate a threshold as to determine the region to be replaced by the generated backgrounds. This region determination is implemented on the grayscale image, we define the region A where the pixel values are above the threshold. Conversely, region B’s pixel values are less than the threshold. With the locations of region A and B, we apply the combination on the original seal image. This operation could be divided into two parts. 1) Adding background pixels directly to the region A. 2) Using a hyper parameter $\alpha$ to adjust the combination ratio between the pixels belong to foreground and background, respectively. The process could be simply described as Equation 6.

\[
\begin{align*}
\text{RegionA}: & \quad \text{pixel}_c = \text{pixel}_b \\
\text{RegionB}: & \quad \text{pixel}_c = \alpha\text{pixel}_o + (1-\alpha)\text{pixel}_b
\end{align*}
\]

where $\alpha \in (0.4, 0.8]$, $\text{pixel}_c$ represents the pixel of combined image, $\text{pixel}_b$ represents the pixel of background image, $\text{pixel}_o$ represents the pixel of original image. The advantage of above operation is that lets combined images get close to the seal images taken directly from painting or calligraphy. The examples of combined images are shown in Figure 4(b) and (c) respectively.

**C. MODEL OPTIMIZATION**

In this work, two types of image data were involved for model training, i.e., the Original Seal Image (OSI), and the generated Seal Image with Backgrounds (SIB). In our Siamese-MTL, the input images are with a fixed dimension of $128 \times 128 \times 3$. The model parameters are learnt by Adam Optimizer [31] in order to minimize the Softmax loss and Contrastive loss. We set learning rate to 0.001 and the batch size to 128. To further reduce the potential of overfitting, a drop-out [32] with the ratio of 50% is applied at the second last fully connected dense layers.

In order to follow the idea of central constraint, we fix the input data of channel A and randomly select the input samples for channel B. The original input data of two channels maybe same or different, the probability of such a selection is kept at 50%. For channel B, we select whether to use the ABG method for the input data, the probability is set to 0%, 50% and 100% respectively. If using ABG, the training sample is SIB in this iteration, otherwise is OSI. In the standard training process, the probability of using ABG for channel A is set to 100%, for channel B is set to 50%. As shown in Figure 5, unlike the way of feeding data randomly, our method makes the distribution of training data more evenly, which is not relevant to the amounts of samples. In this way, data distributions are centered on the original image, encouraging samples with the same identity be close to each other while the ones with different identities be pushed away from each other.

At the end of each stage, we save the training model, and fine-tune the saved model in the next stage directly. The whole training process can be divided into four stages.

1) The model is difficult to converge when using $\text{loss}_{\text{con}} + \text{loss}_{\text{softmax}}$ directly. So, firstly, we pre-trained the Siamese-MTL model only using Softmax loss, as to let model obtain the basic extraction capability. The most important one is that activated the nodes of ‘fc10’ layer should be added from 1,000 to 4,000, then to 10,000 gradually. In the first stage, we did not add the Contrastive loss and data augmentation. The probability coefficient of using ABG for channel A and B was set to 1, the training samples of channel A and B were both SIB. The patience for early stopping was 3,500 batch sizes with at least 98% classification accuracy.

2) In the second stage, we add the Contrastive loss and data augmentation into the training process. Same as stage-1,
the training samples of channel A and B were both SIB. When the classification accuracy reached around 90% and the Contrastive loss was less than 2.0, we stopped the training. In the next stage, the value of Contrastive loss would keep below 2.0.

3) Different from stage 1 and 2, the probability coefficient of using ABG for channel B was set to 0.5. Hence, in this stage, the training sample of A was SIB, while channel B was OSI or SIB at each iteration. We stop training the model until the classification accuracy was around 96%.

4) In the later stage of training, it was difficult to improve the accuracy. Based on stage 3, we dropped the learning rate to 0.0001 and continued to train the model until the classification accuracy achieves roughly 99%. So far the training process was over.

With the method of above step-by-step training, the model can adapt to the characteristics of seal data gradually. When using the data pattern of SIB+OSI, the model mainly learn the original features of the seal image. On the other hand, when using the data pattern of SIB+SIB, the model can learn the features independent to the background. The core idea of our method is to force the feature representations to be close for the positive pairs while be far away for the negative pairs.

IV. EXPERIMENTS
A. DATASETS
In order to validate the effectiveness of the proposed method, we established a standard dataset, including 15,000 Chinese seal images. Among them, 10,000 seals were taken as training set samples, 5,000 as test set samples. In order to assist the ABG algorithm, we also established a background dataset which consists of 1,700 background images. Figure 6 shows the examples of the Chinese seals.

B. MODEL ANALYSIS
In this subsection, we evaluate several variants of the proposed method to validate the impact of multi-task learning and ABG unit. Using the Keras framework with Tensorflow as a backend, we run all the experiments on the Nvidia GTX 1080Ti GPU platform.

Similarity measurement is one of the key issues in our work. In recognition phase, seal images are represented as features in the database. Once the features are extracted from the indexed images, the recognition becomes the measurement of similarity between the features. Many similarity measurements exist. We evaluate the effects of different distance measurements on accuracy and results are shown in Table 1. The distance measurements include Cosine distance, \(X^2\) statistics [33], Euclidean distance and Manhattan distance [34]. The \(X^2\) statistics is defined as

\[
d_{X^2}(Q, T) = \sum_{i=0}^{N-1} \frac{(Q_i - m_i)^2}{m_i} \cdot (7)
\]

where \(Q = \{Q_0, Q_1, \ldots, Q_{N-1}\}\) and \(T = \{T_0, T_1, \ldots, T_{N-1}\}\) are the query and target feature vectors respectively, \(m_i = \frac{Q_i + T_i}{2}\).

As shown in Table 1, results show that cosine distance outperforms other distance measure in terms of accuracy.
In the later experiment subsection, we use cosine distance as our similarity measurement.

In test stage, we propose to use accuracy, macro-precision, macro-recall and macro-F1 score to evaluate different models. Macro-averaged result is a measure of effectiveness on the large classes in a test collection [35]. After using ABG unit, the number of samples on the test dataset will be changed to 50,000. All models are tested on the same test set.

1) EVALUATE THE IMPACTS OF MULTI-TASK LEARNING

To validate the effectiveness of multitask-learning, we use three different losses to train models, as shown in Table 2. Compared with the other two models, Model-3 has the highest accuracy. Using two loss at the same time makes it possible to simultaneously increase the distance between classes and learn the similarity metric. Samples of the same class are close to each other while the ones of different classes are pushed away from each other. As for Model-2, even the distance between different classes is clear, but for many similar seals, it is difficult to identify them only with single criteria.

In the training process, we find that the model dose not converge at all when using Contrastive loss directly. Hence, we made some adjustments for the training of Model-1. Firstly, we only use Softmax loss to train the Model-1 until the classification accuracy achieves 99%. Next, we remove Softmax loss and only use Contrastive loss for fine-tuning until Contrastive loss is less than 2.0. According to the results of the experiment, Model-1 performs badly with the final accuracy reaching only 50.01%. Without the role of the classifier, the analytical power of the model on the data became worse. We think that the model may not be able to cope with large-scale data when only using contrastive loss.

In conclusion, in order to get better recognition performance, the Softmax loss and Contrastive loss are indispensable. Multi-task learning is the core idea of the proposed Siamese-MTL. Classifier serve as the basis of the model which make the distance between classes more clear and simplify the measurement problem, while the similarity metric acts directly on the recognition process.

2) EVALUATE THE IMPACTS OF ABG UNIT

As have mentioned in Section III.C, in the standard training process, we always adopt the ABG for channel A. As for the other channel B, we set a probability coefficient for choosing whether to use the ABG unit. In this section, we set up three comparative experiments to validate the impact of ABG. The probability coefficients of using ABG for channel A are set to 1, 1 and 0, while the coefficients of channel B are set to 0.5, 1 and 0, corresponding to Model-3, Model-4 and Model-5, respectively.

As shown in Table 3, Model-4 performs not well with only 93.23% accuracy. Compared with the standard model Model-3, the accuracy of the Model-4 is lower roughly 2%. Using the data pattern of SIB+SIB, Model-4 only learns very few original features of the seals, which leads to a decrease in accuracy. The data pattern of Model-5 is OSI+OSI. Although this model can achieve a great fitting degree in the training process, the actual testing accuracy is very low. Contrary to Model-4, Model-5 only learns the original features of seals which are useless for real application. Thus the generalization of Model-5 is poor, whose accuracy is only 23.34%. With the data pattern of SIB+(SIB, OSI), we could get a model with good generalization and high accuracy. The learnt features are help for recognition.
In brief, ABG is an efficient online data generation algorithm and plays an important role in the training process. In addition, there is no need to generate a large number of training samples in advance, ABG unit could help save a lot of memory space.

3) EVALUATE THE IMPACTS OF DIFFERENT WAYS OF FEEDING DATA
To evaluate the effects of different ways of feeding data, we test two different input modes to train the models, as shown in Table 4. Refer to Figure 5, using random selection mode, the training samples are not evenly distributed. Negative pairs occupy large proportions of the training samples, while small amounts are positive pairs. In fact, at the beginning of training, we train a classifier with different seals being treated as different classes. The process of training is to increase the distances between different classes. In the mature classification network, the distances between negative pairs already becomes large enough. With the contrastive loss, we further decrease the distance between the positive pairs. As for the random input mode, there is only a process in which samples with different identities are pushed away from each other, while no process in which samples with the same identities get close to each other. This is the reason for decreasing in accuracy of the Model-6.

In this paper, we adopted the input mode based on the Central constraint to uniform the data distribution and make the model training more reasonable. The data distribution is completely independent of the number of training samples and only relates to the chosen proportion, as mentioned in Section III.C.

C. DISCUSSION
The general idea of the Siamese network is to predict the distance between two inputting images. The image pairs with a distance which is lower than the empirically determined threshold will be considered as matched pairs. The core content of the proposed Chinese seals retrieval system (CSRS) can be divided into three parts, i.e., a new Siamese network with multi-task learning, an online data generation algorithm called ABG and a new training method based on the Central constraint. We will discuss them respectively.

1) MULTI-TASK LEARNING
In subsection IV.B.1), we have compared models with three different loss functions. The multi-task learning mechanism makes it possible to increase the distance between dissimilar images and learn a similar metric for similar images at the same time. Enhance the performance of the target task by learning multiple tasks could help make full use of the data characteristics between multiple tasks, providing more accurate knowledge for the target task [36].

2) ONLINE DATA GENERATION ALGORITHM
Refer to section IV.B.2), proposed ABG algorithm can not only solve the over-fitting problem, but also enhance the generalization of the model. The difficulty of seal recognition is that the same seal may have many different forms, such as different thickness of strokes and writing styles. Moreover, same seal may exist in different art works with different background patterns. The proposed ABG algorithm can solve above problems.

3) CENTRAL CONSTRAINT BASED TRAINING
In our network architecture, the distances of feature representations should be independent to any particular targets. To this end, the network should have seen training samples as many as possible. Furthermore, the training dataset should be various enough to cover a good amount of semantics and not focus on any particular type of objects. The distribution of training samples is an important part of the training process. Thus the central constraint based training is reasonable to make the model have better performance. The experimental results in section IV.B.3) also demonstrate this point.

V. CONCLUSION
In this paper, we have proposed a new method named CSRS for seal image recognition. The CSRS mainly contains three units, i.e., a new Siamese network with multi-task learning, an online data generation algorithm called ABG and a new training method based on the Central constraint. The new Siamese Network with Multi-task Learning (Siamese-MTL) can effectively solve the similarity measurement problem and improve the generalization of the recognition model. The Automatic Background Generation (ABG) could generate numerous seal images with different backgrounds for effective training. Whereas the new training method is based on a central constraint. In order to validate the effectiveness of the proposed method, we have established two large scale seals image databases, including 15,000 Chinese seal images and 1,700 background images, respectively. We have evaluated our method and compared with the variant methods on these datasets, achieving satisfied performance. Extensive experimental results indicate that our proposed method is effective and has a great potential for the practical application in Chinese seals recognition. In the future, we plan to...
use the separable convolution structure [37] to construct a lightweight Siamese network to speed up this system.

REFERENCES
[1] L. Tan and M.-L. Hou, “A study on the application of SAM classification algorithm in seal of calligraphy and painting based on hyperspectral technology,” in Proc. 4th Int. Workshop Earth Observ. Remote Sens. Appl. (EORSA), Jul. 2016, pp. 415–418.
[2] H. Bao, D. Xu, and S. Feng, “An effective method to detect seal images from traditional Chinese paintings,” in Proc. Int. Conf. Wireless Commun. Signal Process., Nov. 2009, pp. 1–4.
[3] H. Leung, “Analysis of traditional Chinese seals and synthesis of personalized seals,” in Proc. IEEE Int. Conf. Multimedia Expo (ICME), Jun. 2004, pp. 1283–1286.
[4] C.-L. Su, “Chinese-seal-print recognition by color image dilating, extraction, and gray scale image geometry comparison,” J. Intell. Robotic Syst., vol. 44, no. 4, pp. 349–359, 2005.
[5] C.-L. Su, “Edge distance and gray level extraction and orientation invariant transform for Chinese seal recognition,” Appl. Math. Comput., vol. 193, no. 2, pp. 325–334, 2007.
[6] B. Yu, S. Wan, J. Liu, and Q. Wang, “Classification and recognition of seal image based on Krnovský-RBF,” Image Process. Multimedia Technol., vol. 33, no. 6, pp. 44–47, 2014.
[7] H. Yu, P. D. Reiner, T. Xie, T. Bartczak, and B. M. Wilamowski, “An incremental design of radial basis function networks,” IEEE Trans. Neural Netw. Learn. Syst., vol. 25, no. 10, pp. 1793–1803, Oct. 2014.
[8] J. Lee, S. G. Kong, Y.-S. Lee, K.-W. Moon, O.-Y. Jeon, J.-H. Han, B.-W. Lee, and J.-S. Seo, “Forged seal detection based on the seal overlay metric,” Forensic Sci. Int., vol. 214, nos. 1–3, pp. 200–206, 2012.
[9] Y. Sun, Y. Chen, X. Wang, and X. Tang, “Deep learning face representation by joint identification-verification,” in Proc. 27th Int. Conf. Neural Inf. Process. Syst., Dec. 2014, pp. 1988–1996.
[10] C. Ding and D. Tao, “Trunk-bank ensemble convolutional neural networks for video-based face recognition,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 40, no. 4, pp. 1002–1014, Apr. 2018.
[11] X. Li, Q. Liu, N. Fan, Z. He, and H. Wang, “Hierarchical spatial-aware siamese network for thermal infrared object tracking,” Knowl.-Based Syst., vol. 166, pp. 71–81, Feb. 2019.
[12] M. S. Hanif, “Patch match networks: Improved two-channel and Siamese networks for image patch matching,” Pattern Recognit. Lett., vol. 120, pp. 54–61, Apr. 2019.
[13] H. Kashiani and S. B. Shokouhi, “Visual object tracking based on adaptive Siamese and motion estimation network,” Image Vis. Comput., vols. 83–84, pp. 17–28, Mar./Apr. 2019. Accessed: Feb. 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0262885719300113
[14] J. Wang, Z. Fang, N. Lang, H. Yuan, M.-Y. Su, and P. Baldi, “A multi-resolution approach for spinal metastasis detection using deep Siamese neural networks,” Comput. Biol. Med., vol. 84, pp. 137–146, May 2017.
[15] M. Sabri and T. Kurita, “Facial expression intensity estimation using Siamese and triplet networks,” Neurocomputing, vol. 313, pp. 143–154, Nov. 2018.
[16] C. Jiang, J. Xiao, Y. Xie, T. Tillo, and K. Huang, “Siamese network ensemble for visual tracking,” Neurocomputing, vol. 275, pp. 2892–2903, Jan. 2018.
[17] Z. Han, N. Weiping, Y. Weidong, W. Junzheng, B. Hui, and X. Deliang, “Visual tracking using Siamese convolutional neural network with region proposal and domain specific updating,” Neurocomputing, vol. 275, pp. 2645–2655, Jan. 2018.
[18] A. Y. Ichida, P. Meneguzzi, and D. D. Ruiz, “Measuring semantic similarity between sentences using a siamese neural network,” in Proc. Int. Joint Conf. Neural Netw. (IJCNN), Jul. 2018, pp. 1–7.
[19] C.-I. Wang and G. Tzanetakis, “Singing style investigation by residual neural networks,” in Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP), Apr. 2018, pp. 116–120.
[20] U. Shaham and R. R. Lederman, “Learning by coincidence: Siamese networks and common variable learning,” Pattern Recognit., vol. 74, pp. 52–63, Feb. 2018.
[21] Z. Liu, A. McClung, W. F. H. Yeung, Y. Y. Chung, and S. M. Zandavi, “Top-down person re-identification with Siamese convolutional neural networks,” in Proc. Int. Joint Conf. Neural Netw. (IJCNN), Jul. 2018, pp. 1–8.
[22] W.-S. Zheng, S. Gong, and T. Xiang, “Re-identification by relative distance comparison,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 35, no. 3, pp. 653–668, Mar. 2013.
[23] R. R. Varior, M. Haloi, and G. Wang, “Gated siamese convolutional neural network architecture for human re-identification,” in Proc. Eur. Conf. Comput. Vis. Cham, Switzerland: Springer, 2016, pp. 791–808.
[24] T. Matsukawa, T. Okabe, E. Suzuki, and Y. Sato, “Hierarchical Gaussian descriptor for person re-identification,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2016, pp. 1363–1372.
[25] Y. Chen, D. Zhaoa, L. Lv, and Q. Zhang, “Multi-task learning for dangerous object detection in autonomous driving,” Inf. Sci., vol. 432, pp. 559–571, Mar. 2018.
[26] Y. Tian, H. Wang, and X. Wang, “Object localization via evaluation multi-task learning,” Neurocomputing, vol. 253, pp. 34–41, Aug. 2017.
[27] Y. Wang, L. Tang, and Y. Liu, “Vehicle multi-attribute recognition based on multi-task convolutional neural networks,” Comput. Eng. Appl., vol. 54, no. 8, pp. 21–27, 2018.
[28] B. Graham, “Fractional max-pooling,” Dec. 2014. arXiv:1412.6071. [Online]. Available: https://arxiv.org/abs/1412.6071
[29] B. Yuan, “Efficient hardware architecture of softmax layer in deep neural network,” in Proc. 29th IEEE Int. Syst.–Chip Conf. (SOCC), Sep. 2016, pp. 323–326.
[30] R. Hadsell, S. Chopra, and Y. LeCun, “Dimensionality reduction by learning an invariant mapping,” in Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Jun. 2006, pp. 1735–1742.
[31] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” Dec. 2014. arXiv:1412.6980. [Online]. Available: https://arxiv.org/abs/1412.6980
[32] P. Baldi and S. Sadowski, “The dropout learning algorithm,” Artif. Intell., vol. 210, pp. 78–122, May 2014.
[33] D. Zhang and G. Lu, “Evaluation of similarity measurement for image retrieval,” in Proc. Int. Conf. Neural Netw. Signal Process., Dec. 2003, pp. 928–931.
[34] Y. Zhang, Y. D. Liu, and Z. Ji, “Vector similarity measurement method,” Tech. Acoust., vol. 28, no. 4, pp. 532–536, 2009.
[35] V. V. Asch, “Macro- and micro-averaged evaluation measures [BASIC DRAFT],” Tech. Rep., Sep. 2013.
[36] J. Zheng, S. Guo, L. Gao, and N. Zhao, “Microblog users big-five personality predict. based multi-task learning,” J. Univ. Chin. Acad. Sci., vol. 35, no. 4, pp. 550–560, 2018.
[37] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, H. Adam, “Mobilenets: Efficient convolutional neural networks for mobile vision applications,” Apr. 2017, arXiv:1704.04861. [Online]. Available: https://arxiv.org/abs/1704.04861

ZHENYU WANG received the B.Sc. degree from the National Defense University of Science and Technology, China, in 1997, the M.Sc. degree from Xi’an Jiaotong University, China, in 2003, and the Ph.D. degree from the National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, China, in 2007. He is currently an Associate Professor with the School of Control and Computer Engineering, North China Electric Power University, Beijing, China. His research interests include pattern recognition, computer vision, and machine learning.

JIE LIAN received the B.S. degree from the College of Information Engineering, Southwest University of Science and Technology, Mianyang, China, in 2017. He is currently pursuing the M.S. degree with the School of Control and Computer Engineering, North China Electric Power University, Beijing, China. His research interests include deep learning, pattern recognition, and computer vision.
CHUNFENG SONG received the B.Sc. degree from the Qilu University of Technology, China, in 2012, and the M.Sc. degree from North China Electric Power University, China, in 2016. He is currently pursuing the Ph.D. degree with the National Laboratory of Pattern Recognition (NLPR), Institute of Automation, Chinese Academy of Sciences (CASIA), Beijing, China. His research interests include image segmentation, data clustering, person recognition, and deep learning.

WEI ZHENG received the B.S. degree from the School of Automation and Electronic Engineering, Qingdao University of Science and Technology, Qingdao, China, in 2017. He is currently pursuing the M.S. degree with the School of Control and Computer Engineering, North China Electric Power University, Beijing, China. His research interests include deep learning, pattern recognition, and computer vision.

SHAOULONG YUE received the B.Sc. degree from the Shandong University of Technology, China, in 2018. He is currently pursuing the master’s degree with North China Electric Power University, China. His research interests include pattern recognition, computer vision, and machine learning.

WEI ZHENG

SHAOULONG YUE

SENRONG JI received the B.Sc. degree from Qufu Normal University, China, in 2018. He is currently pursuing the master’s degree with North China Electric Power University, China. His research interests include pattern recognition, computer vision, and deep learning.

SENRONG JI

* * *