UFCNet: Unsupervised Network based on Fourier transform and Convolutional attention for Oracle Character Recognition

Guoqi Liu[1,2,3], Yanan Zhou *[1,3], Yiping Yang[1,3], Linyuan Ru[1,2], Dong Liu[1,2,3], and Xueshan Li[2]

[1]College of Computer and Information Engineering, Henan Normal University, Henan, China
[2]Oracle Bone Intelligent Computing Laboratory, Henan Normal University, Henan, China
[3]Big Data Engineering Laboratory for Teaching Resource & Assessment of Education Quality, Henan, China

Abstract

Oracle bone script (OBS) is the earliest writing system in China, which is of great value in the improvement of archaeology and Chinese cultural history. However, there are some problems such as the lack of labels and the difficulty to distinguish the glyphs from the background of OBS, which makes the automatic recognition of OBS in the real world not achieve the satisfactory effect. In this paper, we propose a character recognition method based on an unsupervised domain adaptive network (UFCNet). Firstly, a convolutional attention fusion module (CAFM) is designed in the encoder to obtain more global features through multi-layer feature fusion. Second, we construct a Fourier transform (FT) module that focuses on the differences between glyphs and backgrounds. Finally, to further improve the network's ability to recognize character edges, we introduce a kernel norm-constrained loss function. Extensive experiments perform on the Oracle-241 dataset show that the proposed method is superior to other adaptive methods. The code will be available at https://github.com/zhouynan/UFCNet.

1 Introduction

The oracle bone inscriptions (OBIs) mainly refer to the OBIs of Yinxu, which are carved on tortoises in the Shang Dynasty. It is the earliest self-contained writing system in China,which is of great significance to the improvement of Chinese cultural history and the study of the formation and evolution of Chinese characters (Xie et al., 2020). The oracle bone character (OBC) image of rubbings is mainly the original image obtained by experts on the unearthed tortoise shell, animal bone, and other text carriers. As the oracle bones have been buried underground for a long time, they are badly damaged or contaminated, and there is serious noise (Huang et al., 2019), which makes it very challenging to recognize OBCs.

Early research methods mainly combine graph theory and topological properties. (Li and Zhou, 1996) proposed an OBIs recognition method based on graph theory. They abstracted oracle into an undirected graph composed of only points and lines, and extracted its topological features. (Li and Zhou, 1996) introduced the information of the adjacent points of the endpoint, and improved the recognition accuracy through the continuous recognition of multi-level feature coding. However, these methods cannot meet the real-world oracle recognition, which requires a lot of manpower and time.

To help with the excavation of new oracle bones and the identification of unseen characters, the advent of deep neural networks has a great impact on the recognition of oracle bone character (OBC) images. (Zhang et al., 2019) used CNNs to map character images into Euclidean space for classification by nearest neighbor rules. (Guo et al., 2015) utilized a low-level representation associated with Gabor and an intermediate representation associated with a sparse encoder and combines it with a CNN-based model. However, training a depth model requires a large number of labeled samples. (Wang et al., 2022) proposed an unsupervised structured Texture separation network (STSN) for Oracle identification and a dataset of 241 classes of Oracle-241 (Wang et al., 2022) for unsupervised identification. They took handprint characters transcribed by experts with high resolution and clean backgrounds as source domains. Accordingly, the original oracle character (scanned image) is taken as the target domains. They have achieved good results by finding a domain invariant feature space to align the distribution between two domains.

In this paper, we propose a network (UFCNet) combining Fourier transform and convolutional attention for oracle character recognition. The convolution attention fusion module (CAFM) combines deep and shallow features to obtain more global information and a better position location of char-

*Corresponding author: 2208283102@stu.htu.edu.cn

Proceedings of the 1st Workshop on Machine Learning for Ancient Languages (ML4AL 2024), pages 98–106
August 15, 2024 ©2024 Association for Computational Linguistics

acters. Additionally, we further design the Fourier transform (FT) module that converts the image from the spatial domain to the frequency domain, aiming to capture the edge details of the glyphs more efficiently and provide rich functionality for the CAFM. We utilize the FT module to capture the high-frequency information of character images and extract rich edge information. We also introduce a kernel norm-constrained loss function to improve the network's discriminative performance on edges. We conduct extensive experiments on the Oracle-241 dataset, and the results demonstrate that our network exhibits better recognition performance in the realm of unsupervised adaptation.

Our main contributions are summarized as follows:

- We deploy CAFM can better extract and fuse features at different levels, and establish a global relationship between multi-layer features.

- We design the FT module, the OBIs are converted to the frequency domain, which can extract the edge features, and provide more effective detail features for the CAFM.

- To validate the effectiveness of our method, we conduct extensive experiments on the Oracle-241 dataset and results demonstrate that UFCNet has better classification accuracy than the existing state-of-the-art (SOTA) unsupervised OBIs recognition method STSN.

2 Related work

2.1 Oracle character recognition

The recognition and deciphering of oracle characters is one of the major topics in the study of oracle bones. With the development of technology, many researchers have tried to recognize oracle characters by image processing. For example, by using non-directed graphs, DNA methods, and template matching (Lin et al., 2016). The earliest studies were (Zhou et al., 1995), (Li and Woo, 2000), (Gu, 2016) which considered oracle features as an undirected graph and used its topological properties as features for classification. (Li et al., 2011) proposed an algorithm based on graph isomorphism. They transformed inscriptions into labeled graphs and used an adjacency matrix of the labeled graphs to encode the inscriptions. (Lv et al., 2010) proposed a Fourier descriptor based on curvature histogram to identify OBIs. (Guo et al., 2015) regarded the oracle bone recognition problem as a sketch recognition task and constructed a hierarchical representation for it.

In addition, (Liu and Liu, 2017) extracted block histogram-based features and applied support vector machines (SVM) to recognize characters. (Gu et al., 2008) believed that the topological structure of OBIs was relatively stable, and calculated the fractal dimension of OBIs according to their fractal characteristics. However, most of these methods are complex large-scale systems composed of multi-layer features, so these methods mainly rely on artificial feature design, which is highly subjective. In particular, they are mostly suitable for small-scale datasets, not for large-scale dataset design and evaluation.

In recent years, convolutional neural networks (CNNs) have made great progress in some computer vision tasks and have been introduced into the recognition of oracle characters. (Huang et al., 2019) published a dataset of scanned oracle characters called OBC306 and proposed a CNN-based evaluation of this dataset as a benchmark, (Guo et al., 2015) aimed to use a CNN-based learning (Wang and Deng, 2018) model to represent oracle characters. They generated a dataset named Oracle-20K and trained and tested it with the proposed CNN. However, they did not discuss the real images of the OBIs and their features such as noise, fracture, and non-uniformity. (Zhang et al., 2019) proposed a deep metric learning-based nearest neighbor classification for oracle recognition and trained a DenseNet (Huang et al., 2017) with triplet state loss to classify manually printed and scanned dataset in a supervised manner. However, the difference in distribution between handprint and scanned characters is not taken into account.

2.2 Unsupervised domain adaptation

Cross-domain tasks are often encountered in computer vision and pattern recognition, there are two types of data, one with labeled information and the other without or little labeled information. To discard the target labeled data, unsupervised domain adaptation (UDA) was proposed in the literature (Wang and Deng, 2018) to solve the problem of domain drift between the labeled source domain and unlabeled target domain.

Popular UDA methods (Long et al., 2015), (Peng et al., 2019) align distributions by moment matching. For example, maximum mean difference

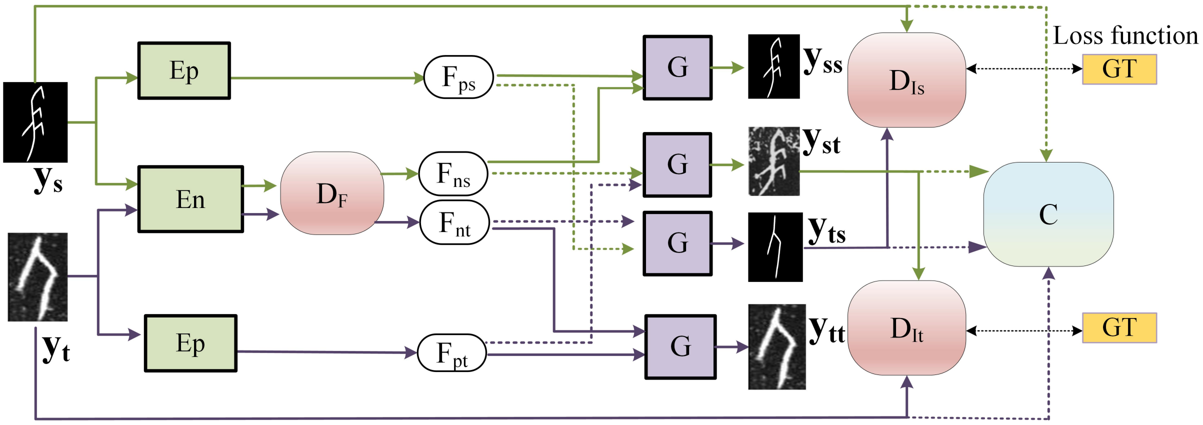

Figure 1: The overall structure of the proposed UFCNet includes a shared encoder E_n for extracting font, which uses RseNet-18 as the backbone network, and an independent encoder E_p. A generator G, a classifier and discriminators D_F and D_I for distinguishing features.

(MMD) (Long et al., 2015), (Chen et al., 2019) were used to reduce distribution mismatch. With a labeled source dataset and an unlabeled target dataset, their main goal is to train the recognition model on the source domain dataset so that it can be generalized to the target domain.

Another common approach to address unsupervised domain adaptation is through adversarial learning strategies (Yaroslav and Victor, 2015),(Eric et al., 2017) where the differences between domains are minimized by jointly training a network of recognizers and a network of domain discriminators. Adversarial learning (Yaroslav and Victor, 2015), (Long et al., 2018) were widely used for alignment of source and target domains. Domain adversarial neural networks (DANN)(Yaroslav et al., 2016) made it impossible for domain classifiers to predict the domain labels of features by the gradient inversion layer (GRL), making the distribution of features on two domains similar. Conditional adversarial domain adaptation (CDAN)(Long et al., 2018) built an adversarial adaptation model based on the discriminative information passed in the classifier prediction. In both methods, a subnetwork called a domain discriminator is used, trained to distinguish between source and target dataset and to learn depth features to confuse the discriminator in domain adversarial training.

If a model is trained directly in the source domain and applied to the target domain, the results are often poor because the feature distributions of the two may be somewhat different. (Wang et al., 2022) proposed the use of UDA to transfer knowledge from easily accessible handprint dataset to the scanned domain. They used a secure distributed alignment in the feature space associated with the structure (glyphs), which can mitigate the negative effects of severe noise and wear and tear. Second, with the idea of Generative Adversarial Networks (GANs), they designed a generator and duplex discriminator to realize the exchange of learned texture (background) information between any pair of images to transform the image. This approach successfully transfers the knowledge of handprint oracle character recognition to the scanned dataset and improves the recognition performance.

3 Methods

The UFCNet network proposed in this paper is shown in Figure 1. It adopts the unsupervised idea of STSN to transfer the knowledge of handprint oracle character recognition to scanned dataset. It consists of three encoders, one of which is a glyph-sharing encoder E_n for extracting both handprint and scanned characters. It is a ResNet-18 pre-trained on the ImageNet dataset as a structural encoder. The other two are independent encoders E_p used to extract the background features of handprint and scanned characters. Specifically, E_p consists of one convolution unit with a kernel size of 7x7 (convolution, BatchNormalization, and ReLU) and four convolution units with a kernel size of 3, CAFM and FT. The CAFM can cascade the high-level and low-level features of handprint and scanned images to obtain rich global features. The FT module can capture more edge features of characters by using the advantage of converting the image to the frequency domain. Alternatively, it includes a generator G, a feature-level discriminator

Generator(G)
Input:f_n, f_p
Deconv(k4n256s2), IN, Relu, ConvBlock(k3n128s1)
Deconv(k4n128s2), IN, Relu, ConvBlock(k3n64s1)
Deconv(k4n64s2), IN, Relu, ConvBlock(k3n32s1)
Deconv(k4n32s2), IN, Relu, ConvBlock(k3n32s1)1x2
Conv(k3n3s1)Tanh
Output: $y^{ss}/y^{st}/y^{ts}/y^{tt}$

Table 1: Network architecture of the generator is used for oracle characters recognition.

Discriminator(D_I)	Discriminator(D_F)
Conv(k6n64s2), IN, Relu(0.2)	Linear(1024), Relu
Conv(k6n128s2), IN, Relu(0.2)	Dropout(0.5)
Conv(k6n256s2), IN, Relu(0.2)	Linear(1024), Relu
Conv(k6n256s2), IN, Relu(0.2)	Dropout(0.5)
Linear(1)	Linear(1), Sigmoid
Output: Real/Fake	Output: Source/Target

Table 2: The discriminator is used for the network architecture Identify.

D_F, two image-level discriminators $\{D_{Is}, D_{It}\}$ and a classifier that is finally used to classify the recognized scanned characters. For the discriminators of images and features, the discriminative network structure uses in this paper is detailed in Table 1 and Table 2.

3.1 Convolutional attention fusion module

To get rich features, we try to fully mine the global and local information of the glyph to improve the dependency extraction of the glyph in the image. We pass the starting image $\chi \in R^{H\times W\times 3}$ through three multi-scale feature maps (i.e., S_1', S_2' and S_3') generated by serialized convolution blocks at different stages. Among these feature maps, S_1' and S_2' provide detailed information about the appearance of oracle characters, while S_3' provides high-level features. Specifically, we consider F as a convolutional unit containing 3×3 convolution, batch normalization (Sergey and Christian, 2015), and ReLU (Xavier et al., 2015). As shown in Figure 2. CAFM is divided into three parts.

Firstly, for the high-level feature S_3', we use an upsampling operation to make the highest-level feature maps S_3' and S_2' have the same size. In this paper, we use the convolutional operation units F1 and F2 with kernel size 3×3 to provide the required information for the network and filter out the unnecessary background texture noise, get the results S_{31} and S_{32}, multiply S_{31} with S_2', this can establish a global relationship between multi-layer features. And input the results obtained from the multiplication into the channel and spatial attention model (CSAM) to get C1. CSAM utilizes channel and spatial weighting on these basic features to better focus on interdependence between some features on channels and space to improve the sensitivity of the model to channels as well as spatial features. Denote the current process as Eq.1.

$$\begin{cases} S_{31} = F1\left[U\left(S_3'\right)\right] \\ S_{32} = F2\left[U\left(S_3'\right)\right] \\ S_{22} = F4\left[U\left(S_2'\right)\right] \\ C1 = CSM\left(S_{31} \times S_2'\right) \end{cases} \tag{1}$$

Secondly, for the features S_2' and S_1' in the lower two layers, we also use the same way of processing the higher-level features by performing convolutional upsampling operations on S_2' and S_3' respectively to reach the same size as S_1'. By multiplying the three features, we can build global features between multiple layers of features. The details of the low-level features are added to the high level after using convolutional attention to CSM to obtain C2. This process is denoted as Eq.2.

$$C2 = CSM\left\{F3\left[U\left(S_3'\right)\right] \times S_1'\right\} \tag{2}$$

Finally, we pass the feature through CSM and smoothly concatenate the resulting C1 with S_{32}, and the feature is mapped to two convolutional units (F5 and F6). Due to the potential loss of crucial detail information during the convolution process, and considering that C2 has already acquired rich local features following the CSM, we opt to integrate the output of the convolution unit with C2. This fusion strategy effectively harnesses some of the original structural information, enhancing the overall feature representation. Finally, we input the connected feature maps into F for dimensionality reduction to get the result T1, which is also the output of CAFM.

3.2 Fourier transform module

The discrete Fourier transform plays an important role in image processing and pattern recognition as an effective computational tool. Several studies (Justin et al., 2016) , (Leon A et al., 2015) have shown that higher feature layers are beneficial in maintaining structural information, while lower feature layers help to maintain what is associated with

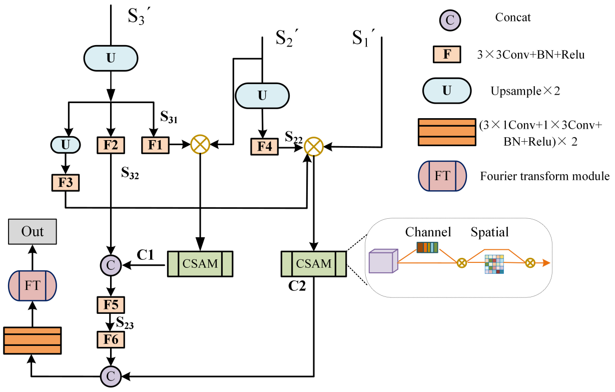

Figure 2: An architecture that passes shallow features into deep features and merges them.

texture. However, in the scanned dataset, it is difficult to distinguish the edge outline of the font because of the similarity between the characters and the background, which makes it difficult to identify the oracle characters accurately. Studies have shown that Fourier transform method can obtain high-frequency information of the object (the edge of the object). At the same time, compared with the spatial domain filtering with large number of cores, frequency domain filtering has obvious advantages. Therefore, we further consider to transfer the image recognition of text to the frequency domain for more detailed feature extraction.

In particular, high-pass filtering can make high-frequency components unimpeded, allowing only high-frequency features to be transmitted, and suppressing low frequencies. The high frequencies in the frequency domain correspond to the Outlines (edges) of the objects in the image. Therefore, FT combines with Gaussian filter is used to extract rich edge information of the oracle bone text image in the frequency domain, so that background pixels and text pixels can be effectively distinguished. The FT module is structured as shown in Figure 3.

It is worth noting that the global feature is obtained by aggregation at the bottom of the encoder. We transform global feature to a single-line greyscale image, performed a two-dimensional discrete FT, and obtained a frequency domain map.

After the discrete FT, it is transmitted to the center of the spectrum graph to obtain the low-frequency information. The number of frequencies of an image in the frequency domain corresponds to the number of pixels of that image in the time domain, indicating that the image has the same number of dimensions in the time and frequency domains. For an input grey-scale image of size

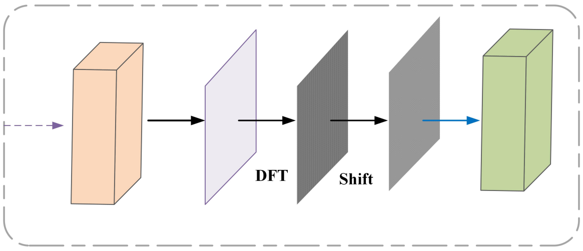

Figure 3: The structure of fourier transform module.

$H_1 \times W_1$, the two-dimensional DFT is expressed as Eq.3.

$$F\left(k,l\right)=\frac{1}{H_1W_1}\sum_{c=0}^{H_1-1}\sum_{d=0}^{W_1-1}f\left(c,d\right)e^{-j\frac{2\pi}{H_1}kc}e^{-j\frac{2\pi}{W_1}ld} \tag{3}$$

The discrete function is for the spatial domain image. We use a combination of Gaussian filter and Fourier transform to extract rich edge information in the frequency domain. Notably, we set the radius of the circular filter to 0.5, which can prevent the loss of details after image reconstruction. Where $F\left(0,0\right)$ shows the lowest frequency and $F\left(H_1-1,W_1-1\right)$ is the highest frequency. Then, the high-frequency portion is processed using the Fourier inverse transform to obtain high-frequency images to explicitly model the dependencies between channels. It can be written as Eq.4.

$$f\left(c,d\right)=\frac{1}{H_1W_1}\sum_{k=0}^{H_1-1}\sum_{l=0}^{W_1-1}F\left(k,l\right)e^{j\frac{2\pi}{H_1}kc}e^{j\frac{2\pi}{W_1}ld} \tag{4}$$

3.3 Loss function

To generate more realistic OBCs, the following perceptual loss (l_{pre}) (Wang et al., 2022) and reconstruction loss (l_{rec}) (Wang et al., 2022) are introduced in this paper to impose constraints on the structural similarity and texture similarity during image reconstruction. The first part, perceptual loss, constraints y_{st} to be similar to y_t in texture; it also requires y_{st} to be similar to y_s in structure. A similar constraint is imposed on the transformed image y_{ts}. The second part of the reconstruction loss ensures that the reconstructed images y_{ss} and y_{tt} should be the same as the original input images y_s and y_t. In addition, we apply the mean square loss (MSE) and the cross entropy loss function CrossEntorpyLoss.

In particular, we also propose a key loss function l_{bcem}, which is a loss function based on BCELoss.

Specifically, we introduce the nuclear norm constraint BNM (Leon A et al., 2015) to improve the edge discrimination ability of the network. In the case of insufficient labels, the performance of the network on the decision boundary will be degraded. To improve discriminability, we introduce nuclear norm maximization to improve target prediction ability. Experiments show that when the weighting factor is 0.5, BNM enables the network to obtain the optimal result for the discrimination of the target domain edge that lacks labels. So the total loss of our l_{bcem} is Eq.5.

$$l_{bcem} = l_{bce} - l_{BNM} \tag{5}$$

Thus, the overall loss in this paper is Eq.6.

$$l_{loss} = l_{mse} + l_{ce} + l_{pre} + l_{rec} + l_{bcem} \tag{6}$$

4 Experiment

4.1 Datasets

In this section, we use the Oracle dataset of Oracle-241 for character recognition, using our network to transfer knowledge from the handprint data to the scanned data. Oracle-241 contains 78,565 handprint and scanned characters in 241 categories. The handprint samples used for training and the unlabeled scan samples are 10861 and 50168, respectively. The dataset use for testing contains 3730 handprint data and 13806 scan data. As shown in Table 3.

4.2 Implementation details

The proposed method uses Pytorch as a framework and runs on a single NVIDIA GeForce GTX 3090Ti 24G GPU. We perform 150,000 iterations on data with a batch size of 16. For preprocessing, we randomly crop and flip the training samples, setting the weight decay and initial learning rate to 5e-4 and 2.5e-4, respectively. This paper follows standard protocols for unsupervised domain adaptation, e.g. (Yaroslav et al., 2016), (Long et al., 2018). Train with all marked source characters and all unmarked target characters. To quantitatively evaluate the recognition performance of UFCNet on handprint and scan datasets, classification accuracy is used as the evaluation metric in this paper, and the calculation method is as follows Eq.7.

$$ACC = \frac{TP + FN}{TP + TN + FP + FN} \tag{7}$$

Where TP and TN represent the number of pixels and background texture pixels of correctly identified oracle font structure, respectively. Similarly, FP represents a background pixel incorrectly identified as an oracle glyph structure, while FN represents an oracle glyph structure pixel incorrectly predicted as a background pixel.

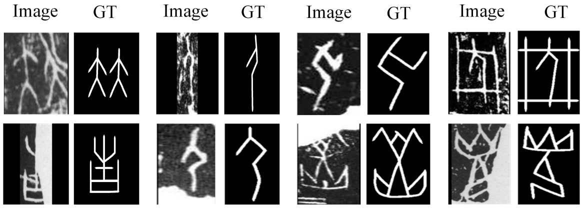

Figure 4: Eight images are misclassified with the "single-source" model, but our model classified them correctly.

	Train	Test
handprint	10861	3730
Scan	50168	13806

Table 3: Statistics from the ORACLE-241 dataset.

4.3 Comparative experiment

To demonstrate the effectiveness of our network, we compare the UFCNet with some of the methods used to identify (Huang et al., 2019). Since they only use handprint samples to train the network model, the model trained on the source domain has no adaptation, they are referred to as " single-source " models in this paper. In addition, we compare with other SOTA adaptive methods for image classification, such as CDAN, DANN, BSP (Chen et al., 2019), and GVB (Cui et al., 2020). All of these data are used with ResNet-18 as the backbone and experimented in the same environment to make a fair comparison.

Method	Accuracy (%)	
	Handprint	Scan
ResNet	94.9	2.9
CDAN	86.5	37.8
DANN	88.7	31.4
BSP	91.7	33.7
GVB	87.8	36.8
STSN	92.2	44.9
Ours	**94.7**	**56.5**

Table 4: Source and target Accuracy (MEAN %) on ORACLE-241 dataset is statistically compared with various state-of-the-art (SOTA) methods. The best numbers are represented in bold.

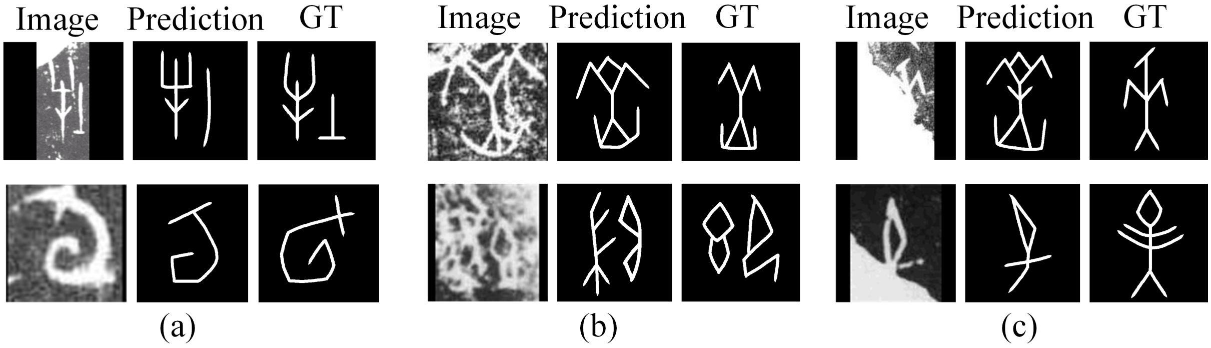

Figure 5: The example samples which are misclassified by UFCNet. For each set of characters, the left, middle, and right images represent the scan sample, model prediction, and ground-truth (GT), respectively. (a) represents characters that look similar, (b) denotes characters that contain heavy noise, and (c) is heavily polluted or occlusion characters.

We can see from the results in Table 4. Firstly, training on handprint dataset and testing on the same domain, the model trained and tested only on the source domain model can obtain higher accuracy. When directly apply to scanned dataset, the model's performance undergoes a marked degradation. Figure 4 shows some example images that are misclassified by the "single-source" model but correctly classified by our model. From these results, we can find that the "single-source" model has difficulty in identifying scanned images with severe noise and contaminated wear, while our model can successfully identify them. Our method in this paper transfers the knowledge from handprint dataset to scanned dataset by unsupervised domain transfer, and better results can be obtained on scanned dataset.

Secondly, we note that although the existing unsupervised domain adaptation methods can use domain invariant features to improve the performance of the target domain, this phenomenon illustrates the importance of mitigating domain transfer. However, if this method does not consider the texture feature information contained in the source domain and the target domain, it is still difficult to align the entire source domain and the target domain. In addition, the characteristics of having two domains meanwhile will also have a certain degree of negative impact on alignment, thus affecting the performance of the two domains. GVB uses a fully connected bridge to model domain-specific parts. Compared with the single domain method, the success rate of GVB for scanning sample recognition is 36.8%. However, the simple structure of the bridge makes it difficult to capture the characteristics of different fields very well.

Finally, DANN does not consider the relationship between samples and labels, but only directly connects samples and labels to form a higher-dimensional vector. This approach will hurt distinguishing the source domain and the target domain. Compared with DANN, CDAN has improved the scanned dataset by 6.4%. CDAN introduces sample weighting in the discriminator for both the source and target domains. As the classifier converges, the weight assigned to source domain samples will gradually approach unity, leading to equal weighting for source samples. Although BSP applies the singular value decomposition method to obtain the maximum k singular values of the source and target eigenmatrices, respectively. The BSP is utilized as the regularization term in these maximum k singular values. Nevertheless, due to the discrepancies between domains, the eigenvectors might not receive equal contributions from the source and target domains, potentially leading to distortions.

In particular, for the classical adaptive models CDAN and DANN, benefiting from the joint adaptation of STSN, pick-up entanglement and transformation and freedom from contamination by background textures during the adaptive process, our network model's improvements on top of them are more advantageous for the recognition and classification of scanned dataset. Inspired by the Fourier transform, detailed features of the character structure are extracted from a frequency domain perspective, especially the edge part of high frequency. In addition, a convolutional attention module is introduced to extract more comprehensive features at the encoder

However, due to the existence of some similar characters, the model classification fails. For ex-

Method	Accuracy (%)	
	Handprint	Scan
Baseline	92.2	44.9
Baseline+CAFM	93.2	50.7
Baseline+CAFM+FT	94.6	54.6
Baseline+CAFM+FT+bcem	94.7	56.5

Table 5: Statistical comparison of ablation experiments of two key components in UFCNet. CAFM stands for convolution attention fusion module. FT stands for Fourier transform module.

ample, the characteristics of prediction and ground-truth (GT) categories differ only in a few details. Secondly, as shown in Figure 5, severe noise, severe image degradation, even for humans, there are certain challenges.

4.4 Ablation experiments

To verify the experimental effectiveness of each block in our network, we conduct ablation experiments on UFCNet. The baseline network is a U-shaped codec structure where the private encoder consists of one convolution unit with a kernel size of 7x7 (convolution, BatchNormalization, and ReLU) and four convolution units with a kernel size of 3. After each convolution, the input feature is downsampled twice, the size of the feature map is reduced, and then it is re-amplified through the upsampling operation, which is used to transfer information between the encoder and the decoder, so as to retain more detailed information. Then the average pooling operation is performed to reduce the noise effect of irrelevant features. We add a convolutional attention module and a Fourier transform module to this and tested the baseline+CAFM and baseline+FT and loss function on dataset Oracle-241, respectively. All the ablation experiments are performed in the same computational environment. The test results are shown in Table 5.

Effectiveness of CAFM: Compared to the base network, the performance optimization of adding CAFM, especially in the classification accuracy of the scanned dataset, increased by 5.8%. This further indicates that adding the CAFM module to the base network can capture more global feature information, helping to locate the location of the object.

Effectiveness of FT: The addition of the FT module to the base network shows the superiority of our FT module by Table 5, especially the recognition accuracy for scanned dataset increases by 6.8%. In particular, the FT module can obtain more edge information when extracting high frequencies from images

Effectiveness of the loss function: We use the improved l_{bcem} function, and the results in Table 5 shows that our loss function can improve the discriminative property of the network for edges and can better extract the detailed features of oracle characters.

5 Conclusion

In this paper, we propose a new network UFCNet for the recognition of oracle character images. Different from the recognition method of OBCs based on CNNs, we use the Fourier transform to transfer the recognition of oracle character images from the image domain to the frequency domain and extract rich edge information. At the same time, we use the convolutional attention fusion module to fuse shallow features with deep features in multiple layers, which makes up for the important detailed features lost in the sampling process of the CNN. A large number of experiments show that our UFCNet has better recognition accuracy compared with SOTA methods. However, due to the serious incompleteness and blurring of OBCs, our network still needs to be further improved in recognition.

Acknowledgments

The work was supported by the National Natural Science Foundation of China (No.U1904123, No.61901160).

References

F Li and XL Zhou. 1996. The graph theory method of oracle bone inscriptions automatic recognition. *J. Electron*, 18(1):41–47.

F Li and X.L Zhou. 1996. Study on computer identification method of oracle. *J. Fudan Univ*, 481–486.

Jun Guo et al. 2015. Building hierarchical representations for oracle character and sketch recognition. *IEEE Transactions on Image Processing*, 25(1):104–118.

Yi-Kang Zhang et al. 2019. Oracle character recognition by nearest neighbor classification with deep metric learning. *2019 International Conference on Document Analysis and Recognition (ICDAR)*, 309–314.

Mei Wang et al. 2022. Unsupervised Structure-Texture Separation Network for Oracle Character Recognition. *IEEE Transactions on Image Processing*, 31:3137–3150.

Meng Lin et al. 2016. Recognition of oracular bone inscriptions using template matching. *International Journal of Computer Theory and Engineering*, 8(1):53.

Xin-Lun Zhou et al. 1995. A method of Jia Gu Wen recognition based on a two-level classification. *Proceedings of 3rd International Conference on Document Analysis and Recognition*, 2:833–836.

Feng Li and Peng-yung Woo. 2000. The coding principle and method for automatic recognition of Jia Gu Wen characters. *International Journal of Human-Computer Studies*, 53(2):289–299.

Qingsheng Li et al. 2011. Recognition of inscriptions on bones or tortoise shells based on graph isomorphism. *Jisuanji Gongcheng yu Yingyong(Computer Engineering and Applications)*, 47(8):112–114.

Shaotong Gu. 2016. Identification of oracle-bone script fonts based on topological registration. *Computer & Digital Engineering*, 10:029.

Xiaoqing Lv et al. 2010. A graphic-based method for Chinese Oracle-bone classification. *Journal of Beijing Information Science and Technology University*, 25(Z2): 92-96.

Yongge Liu and Guoying Liu. 2017. Oracle bone inscription recognition based on SVM. *Journal of Anyang Normal University*. 2:54–56.

Shuangping Huang et al. 2019. OBC306: A large-scale oracle bone character recognition dataset. *2019 International Conference on Document Analysis and Recognition (ICDAR)*, 681–688.

Gao Huang et al.(2017). Densely connected convolutional networks. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 4700–4708.

Mei Wang and Weihong Deng. 2018. Deep visual domain adaptation: A survey. *Neurocomputing*, 312:135–153.

Mingsheng Long et al.(2015). Learning transferable features with deep adaptation networks. *International conference on machine learning*, 97–105.

Xingchao Peng et al. 2019. Moment matching for multi-source domain adaptation. *Proceedings of the IEEE/CVF international conference on computer vision*, 1406–1415.

Yiming Chen et al. 2019. A graph embedding framework for maximum mean discrepancy-based domain adaptation algorithms. *IEEE Transactions on Image Processing*, 29:199–213.

Ganin Yaroslav and Lempitsky VVictor. 2015. Unsupervised domain adaptation by backpropagation. *International conference on machine learning*, 1180–1189.

Ganin Yaroslav et al. 2016. Domain-Adversarial Training of Neural Networks. *Journal of Machine Learning Research*, 17(1):2096-2030.

Tzeng Eric et al. 2017. Adversarial discriminative domain adaptation. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 7167–7176.

Mingsheng Long et al. 2018. Conditional adversarial domain adaptation. *Advances in neural information processing systems*, 31.

Ioffe Sergey and Szegedy Christian. 2015. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *International conference on machine learning*, 448–456.

Glorot Xavier et al. 2015. Deep sparse rectifier neural networks. *Proceedings of the fourteenth international conference on artificial intelligence and statistics*, 315–323.

Johnson Justin et al. 2016. Perceptual losses for real-time style transfer and super-resolution. *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings*, 694–711, Part II 14.

Gatys Leon A et al. 2015. A neural algorithm of artistic style. arXiv preprint arXiv:1508.06576.

Shuhao Cui et al. 2020. Gradually vanishing bridge for adversarial domain adaptation. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 12455–12464.

Shaotong Gu al. 2008. A study of oracle input encoding based on glyph topology. *Chinese Journal of Informatics*, 22(4):123–128.

Xinyang Chen et al. 2019. Transferability vs. discriminability: Batch spectral penalization for adversarial domain adaptation. *Proceedings of International conference on machine learning*, pp. 1081–1090.

Naihe Xie et al. 2020. Flourishing from Yin Ruins to the world—a review of "Busan, south Korea international symposium commemorating the 120th anniversary of oracle bone inscription discovery". *[J]. Guan zi journal*, pp. 125–128.