IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 3 February 2024, accepted 15 March 2024, date of publication 22 March 2024, date of current version 28 March 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3381039

RESEARCH ARTICLE

Incremental Recognition of Multi-Style Tibetan Character Based on Transfer Learning

GUANZHONG ZHAO[1], WEILAN WANG[1,2], XIAOJUAN WANG[1,2], XUN BAO[1], HUARUI LI[1], AND MEILING LIU[1]

[1]Key Laboratory of China's Ethnic Languages and Information Technology of Ministry of Education, Northwest Minzu University, Lanzhou 730030, China
[2]College of Mathematics and Computer Science, Northwest Minzu University, Lanzhou 730030, China

Corresponding author: Weilan Wang (wangweilan@xbmu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 62166036 and Grant 61772430, in part by Gansu Provincial Science and Technology Plan Natural Science Foundation under Grant 22JR5RA187 and Grant 23YFGA0072, and in part by the Fundamental Research Funds for the Central Universities Grant 31920220037 and Grant 31920220132.

ABSTRACT Tibetan script possesses a distinctive artistic form of writing, intricate glyph structures, and diverse stylistic variations. In the task of text recognition, effectively handling the recognition of Tibetan script with significantly different stylistic fonts remains a challenge. Existing research has made considerable progress in recognizing Tibetan script within a single style using techniques such as convolutional neural networks and convolutional recurrent neural networks. However, when dealing with multi-style Tibetan script recognition, the standard approach involves training models using a multi-label joint training method. This approach annotates the style and class of different font style samples and merges them into a single dataset for model training. Nevertheless, as the amount of data and performance requirements increase, this approach gradually faces issues such as decreasing accuracy, insufficient generalization capability, and poor adaptability to new style samples. In this paper, we propose a transfer learning-based method for incremental recognition of multi-style Tibetan script, referred to as ''multi-style Tibetan script incremental recognition.'' In the style recognition stage, we employ a convolutional neural network (CNN) to accurately differentiate between style categories. During the pre-training stage, we train a residual network on the Tibetan Uchen standard style and utilize it as the baseline model. In the multi-style Tibetan script recognition stage, we integrate transfer learning into the model training process to reduce the training time. These three stages collectively accomplish the task of multi-style Tibetan script incremental recognition. The experimental results demonstrate that our approach achieves a significant improvement in overall recognition accuracy, from 90.14% to 98.40%, when utilizing the TCDB and HUTD datasets compared to traditional multi-task recognition methods. This method exhibits high accuracy, strong generalization capability, and good adaptability to new style samples in multi-style Tibetan script character recognition. Furthermore, it can be applied to other tasks involving multi-style, multi-font, and multi-script recognition.

INDEX TERMS Tibetan recognition, multi-style recognition, residual networks, transfer learning, incremental recognition.

I. INTRODUCTION

With the rapid development of digital technology, there is a growing demand for the recognition and processing of diverse texts. Tibetan script, as a unique and complex writing system, plays a crucial role in the daily communication, cultural heritage, and academic research of the Tibetan people. Tibetan sentences are composed of syllables, with syllables separated by syllable boundaries. The structure of a Tibetan syllable is illustrated in Figure 1(a). A syllable can contain up to seven components, including prefix, base character, superscript, hyphen, super vowel (or lower vowel), first postposition, and second postposition. Furthermore, a syllable can have at most one vowel (either a super vowel

The associate editor coordinating the review of this manuscript and approving it for publication was Varuna De Silva.

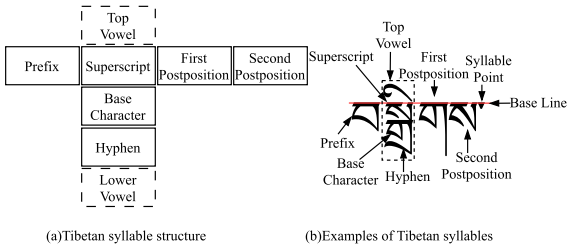


FIGURE 1. Tibetan syllable structure and examples.

or a sub vowel). Each vertical unit within a syllable is called a character ding, and prefix, first postposition, and second postposition are units consisting of character ding, while the character ding containing the base letter must have at least one base letter, and any other component can be omitted. This means that different character ding can have varying levels of superimposedlayers, as depicted by the dashed box in Figure 1(b), which represents a character ding with four layers of superimposition. In most studies, character ding are used as recognition units to investigate and understand the characteristics and styles of Tibetan character ding in a more detailed manner. Based on this background, this paper focuses on the research related to Tibetan character ding.

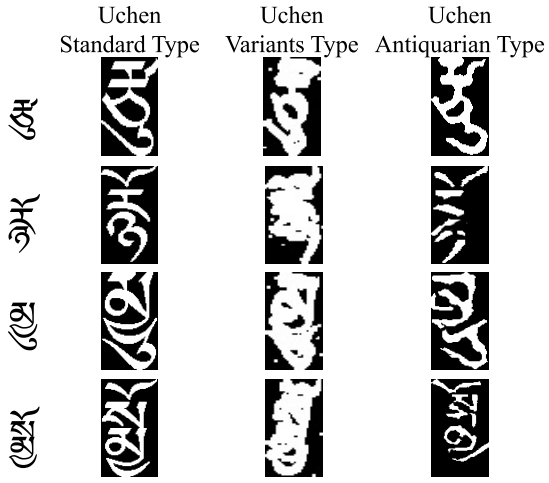


FIGURE 2. Tibetan characters in different writing styles.

Tibetan character recognition plays a crucial role in various applications such as document digitization, text mining, and language preservation. However, Tibetan writing styles exhibit rich diversity, with significant differences in stroke patterns, line thickness, and overall shapes. In this study, we focus on three distinct styles of Tibetan characters (Tibetan Uchen Standard, Variants, and Antiquarian) as shown in Figure 2, which clearly demonstrate their pronounced differences. For instance, the Uchen Standard characters exhibit an upright and square appearance, while the Uchen variants and antiquarian are relatively inclined, distorted, and blurry. Moreover, the antiquarian often suffer from stroke discontinuity and misalignment, posing significant challenges for accurate and robust character recognition. Consequently, developing an effective recognition system that can handle different styles remains a formidable task.

The main challenging issues in the existing Tibetan character ding recognition system are as follows:

Challenge 1: The model is capable of efficiently identifying standard-style characters, but its performance is compromised when it comes to recognizing characters with relatively messy handwriting and distorted shapes. Moreover, the difficulty in collecting adequate data samples further hampers the recognition accuracy.

Challenge 2: The model relies on training with specific-style character datasets, which limits its ability to effectively recognize characters of other styles, resulting in limited generalization capability.

Challenge 3: The model designed for recognizing characters of different styles in a multi-task mode effectively addresses the problem of recognizing characters of various styles. However, if there is a need to incorporate a new style character dataset, the entire model needs to be retrained, which significantly increases the time cost.

Challenge 4: When designing a recognition model for different style characters in a multi-task mode, it is necessary to ensure a one-to-one correspondence between the categories of characters in different styles. That is, the character categories in style A should be the same as those in style B, which limits the scalability of the model.

Upon analyzing the existing challenging problem, it has been observed that the recognition performance of standard Tibetan script characters is quite remarkable in current research. However, the recognition performance of characters with other stylistic variations is relatively poor. Furthermore, there are certain difficulties in collecting data samples for these characters. Therefore, we contemplate whether it is possible to leverage the efficiency of standard script recognition to assist in improving the recognition performance of characters with other stylistic variations. This entails establishing a correlation between two distinct domains of character recognition tasks. By adopting this approach, the issue of domain adaptation arises.

In the realm of domain adaptation, transfer learning, with its ability to effectively utilize knowledge from the source domain and address domain discrepancies, has emerged as an effective method for tackling domain adaptation problems. Specifically, it involves leveraging knowledge acquired from related tasks or domains to enhance performance on the target task, thereby mitigating the challenges posed by data scarcity and style diversity. This entails employing annotated data of existing styles to train the model and transferring the model's knowledge to the recognition of characters in new styles. Through the utilization of transfer learning methods, the task of recognizing characters with different stylistic variations based on standard script characters can be accomplished. However, this approach still falls short in addressing the recognition of characters with multiple styles. Our aim is to

achieve recognition of various stylistic variations of characters within a single model, i.e., incremental recognition.

When addressing the problem of incremental recognition, the strategy commonly employed is incremental learning. The primary rationale behind this lies in its ability to significantly reduce computational time costs, preserve existing knowledge, adapt flexibly to changes, and control forgetting. Specifically, based on the existing model, only the data pertaining to the newly added categories needs to be used for model updating and expansion, while the previously learned knowledge in the model remains unchanged. By adopting the incremental learning strategy, incremental recognition of newly introduced stylistic variations can be achieved, thereby accomplishing the objective of recognizing characters with multiple styles.

The aim of this study is to investigate incremental recognition of multi-stylistic Tibetan script characters based on transfer learning. The primary focus of the study is on Tibetan script characters, specifically, the Tibetan Uchen standard script (generated by rendering 19 Uchen Tibetan font files), Uchen variants script (generated by combining 12 Uchen Tibetan font files with extensive data augmentation techniques), and character images extracted from segmented Tibetan script passages in ancient Tibetan manuscripts (referred to as Tibetan antiquarian characters in this paper). Specifically, we address the following aspects:

- An effective transfer learning framework is proposed in this study. The framework involves pre-training the feature extractor on existing stylistic data to acquire a generalized representation of Tibetan script characters. This learned representation serves as the shared feature extractor within the model. To adapt to the new recognition task, fine-tuning techniques are applied to the classifier for new stylistic characters.
- By introducing the idea of incremental learning, the model makes it possible to gradually adapt to the recognition of new style character dings by sharing feature representations and task-related classifiers, without the need to re-train the whole model, which greatly saves time and cost.
- Among the recognition of the added style wordings, in order to verify the scalability of the model, the categories of wordings that not existed in the existing knowledge are joined to the added style wordings, and analyze whether the established knowledge has an effect on the recognition accuracy of the known categories versus the unknown categories.
- Evaluate and compare the performance of different classification algorithms in Tibetan character ding recognition to verify the effectiveness and performance advantages of our methods.

The principal innovations and contributions of this research are outlined below: 1) We have achieved significant recognition results in the field of Tibetan character recognition by pioneering the application of transfer learning and incremental identification methods. Transfer learning empowers the model to fully exploit existing relevant data and models, enabling the transfer and sharing of knowledge across diverse tasks. Incremental identification, on the other hand, allows us to continuously update and expand the model based on existing foundations, thereby adapting to new writing styles and font variations in recognition tasks. 2) Our proposed methodology exhibits remarkable robustness. Following training on the source domain, the model dynamically adapts to the target domain through specified updates, thus accommodating recognition tasks involving any language and font style. This dynamic update mechanism enables the model to flexibly adapt to varying data distributions and feature changes, thereby enhancing overall recognition performance. 3) We have significantly improved the accuracy of Tibetan character recognition across various styles and variations, providing a more reliable foundation for Tibetan language-related applications such as OCR systems and automatic translation. 4) Since our approach possesses exceptional generalization capabilities, extending beyond the study of recognition models to encompass recognition methods, our research holds valuable implications for multi-style recognition tasks involving other character systems.

This thesis is structured as follows: section II provides an overview of character recognition techniques, the application of transfer learning in character recognition and related work in incremental recognition. Section III describes the proposed approach in detail, including the transfer learning framework, feature extraction techniques and classification algorithms. Section IV describes the dataset, experimental setup, results and analysis. Finally, Section V summarizes the paper and discusses the limitations of the proposed method.

II. RELEATED WORK

A. CHARACTER RECOGNITION METHOD

Character recognition is a crucial component of Optical Character Recognition (OCR) technology. Its primary task is to take character images as input sources, undergo a series of computer processing, and output recognized encoded characters, thus enabling interaction between humans and computers. Traditional character recognition heavily relies on machine learning and image processing methods, where character recognition tasks are accomplished through image preprocessing, manual feature extraction, and simple classifiers in machine learning [1]. However, with the increasing amount of data and improved computational capabilities, the effectiveness of traditional methods has significantly diminished. Consequently, numerous deep learning techniques and algorithms have emerged, such as Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), Residual Connections, and Attention Mechanisms. Research has increasingly tended to utilize deep learning to solve the problem of character recognition.

Chinese character recognition has been improved in recent years. Among them, to solve the problem of accurate

recognition of simple numeric characters, Mursari et al. [2] proposed preprocessing operations such as grayscale, morphological processing, and denoising of the original numeric handwritten images before passing them into the recognition engine, which successfully improved the recognition performance, but it was limited to the simple numeric characters, and the recognition effect was not good if it was extended to other Latin or non- Latin texts. In the field of printed Chinese character recognition, similarly, Wang et al. [3] highlighted the character features of Chinese characters through pre-processing operations such as denoising, binarization, row and column segmentation, smoothing, and normalization, and completed the recognition of printed Chinese characters by using the vertical projection method to slice the characters and combining with the Tesseract-OCR recognition engine to shorten the recognition time of each Chinese character to 0.07 seconds. On this basis, Wang et al. [4] introduced phrase processing and proposed a hybrid recognition model based on Tesseract, KNN, and LSTM, i.e., the inputs and outputs of the three techniques are interconnected, and the outputs of one process are more suitable for the ideal input requirements of the next process, which solves the problem of low recognition accuracy when the images to be recognized are compressed, partially blurred, and smeared. However, all of them have simple and fixed text content, and cannot effectively recognize relatively complex character images. To solve the problem of multi-style Chinese character recognition, Wu et al. [5] proposed a CNN-based Chinese character radicals identification recognition, that the Chinese characters with the same radicals should have similar features, the text identifies 15 radicals and three font styles (italic, brush, and handwriting), and the recognition of Chinese characters with different styles has been accomplished by identifying the training of each style. However, there is an over-reliance on the head marking information and over- fitting due to too many differences in styles, so the robustness of the model is poor. In the field of a scene, Chinese character recognition, Yao et al. [6] proposed a target detection and invoice recognition method based on YOLOv3+CRNN, which realizes the fast recognition and processing of invoice content, but there are problems of excessive data storage and inability to solve the problem of effective recognition under high noise and pollution.

In contrast to Chinese character recognition, recognition studies of minor languages are relatively scarce, mainly due to the difficulty of data sample collection as well as the high complexity of characters. In the field of Newari language recognition, Bati et al. [7] constructed a Ranjana handwritten character dataset consisting of 62 characters and performed recognition tasks using models such as LeNet5, AlexNet, and ZFNET. However, due to the limited number of data samples and insufficient character categories, the recognition accuracy was not satisfactory. Similar to Newari, Tibetan belongs to a minority language as well. However, unlike Newari, Tibetan exhibits diverse styles, with significant differences between different styles of Tibetan characters. Moreover, collecting samples of complex-style characters is challenging. For example, Tibetan ancient manuscript characters require the collection of character samples through character segmentation methods in Tibetan ancient manuscript documents. However, these manuscript documents often exhibit complex phenomena such as stroke overlap, ligature, fragmentation, and erosion, making the document image character segmentation method highly demanding and technically challenging.

In the early research on Tibetan character recognition, Zhou et al. [8] proposed a geometric shape-based Tibetan character recognition method for simple characters. The characters were divided into multiple components, and then the structure and density features of each character were calculated. The components were recognized using template matching, and the recognition results of the components were combined to obtain the character recognition results. Wu et al. [9] proposed a Tibetan numeral recognition algorithm based on H-KNN, utilizing the nearest neighbor algorithm and Hilbert curve to recognize Tibetan numerals. Compared to traditional KNN algorithms, this method achieved significantly improved recognition accuracy. However, this method gradually loses effectiveness as the number of categories and image complexity increase.

For Tibetan handwritten recognition, Zhang et al. [10] proposed a lightweight neural network, ticonvb(HUTNet), based on Tibetan structure and internal connections between FLOPs and MAC, to perform handwritten Tibetan character recognition tasks. The model was compressed using pruning and distillation techniques to accelerate recognition speed. However, it exhibited relatively poor performance in recognizing similar and irregular characters. For complex images, such as those in Tibetan ancient manuscripts with severe problems of character ligature, fragmentation, and occlusion, Zhao [11] used the training method of BP neural networks to train on woodblock Uchigane-style Tibetan scripture characters, correcting the data and improving the recognition accuracy of woodblock Uchigane-style Tibetan scripture characters under certain conditions (slight interference). However, the recognition performance of woodblock Uchigane-style Tibetan scripture characters under severe interference (fragmentation, distortion, ligature) was not satisfactory.

In the field of multi-style Tibetan character recognition, Hou et al. [12] constructed a large- scale Uchigane-printed multi-font Tibetan text recognition dataset. Based on the DBNet segmentation network, they proposed a transcription dictionary strategy that uses 74 commonly used Tibetan character samples as the transcription dictionary for end-to-end text recognition. They selected ResNet34 as the backbone network to recognize various printed styles of Tibetan characters under the Uchigane style. Although this method achieved recognition of multi-styleTibetan characters, all

the styles involved belonged to the Uchigane printed style, resulting in poor generalization ability.

B. TRANSFER LEARNING IN CHARACTER RECOGNITION

The concept of transfer learning was formally introduced in 2010 by machine learning researcher Qiang [13] as a learning approach proposed to address domain adaptation and related issues. In current research, transfer learning has been widely applied to Chinese character recognition and recognition of text in languages with relatively simple character structures.

In the field of printed Chinese character recognition, Yan et al. [14] proposed a transfer learning-based convolutional neural network model for recognizing printed Chinese character fonts. They modified the Inception-v3 network and combined it with transfer learning methods to transfer a pre-trained model to the task of recognizing printed Chinese character fonts. However, due to the limitations of the model, retraining of the entire model is required when recognizing new fonts, resulting in poor scalability. Li et al. [15] developed an OCR system for recognizing printed Chinese characters in ID cards by improving GoogLeNet and fine-tuning. The system utilizes transfer learning with a large-scale dataset, ensuring a high recognition accuracy while maintaining a sufficient data scale. However, the system exhibits unsatisfactory performance in recognizing similar characters, and its development is constrained by fixed scenarios, limiting its applicability to diverse recognition scenarios.

In the domain of handwritten Chinese character recognition, Shi et al. [16] addressed the issue of character fusion in handwritten Chinese characters by proposing a morphology-based adaptive bounding box segmentation method. They combined the segmented handwritten Chinese characters with the ResNet101 model for transfer recognition, effectively resolving the problems of character fusion and low recognition accuracy in traditional handwritten Chinese character segmentation. However, this method imposes high requirements on the data, necessitating a large number of samples to support model training.

For the recognition of languages with simple character structures, Elaraby et al. [17] employed the pre-trained DarkNet-53 model for Braille recognition, successfully completing the Braille recognition task. However, due to limited data samples, severe overfitting occurred during the training process. Rizky et al. [18] achieved efficient recognition of digits and English characters by utilizing the VGG-16 model in conjunction with transfer learning methods. However, the preprocessing stage requires data augmentation operations on input images, and different augmentation methods and parameters have a significant impact on recognition results. Daood et al. [19] modified the MobileNetV2 model using transfer learning techniques to recognize handwritten Arabic digits and characters. Since the number of character categories was small and the sample size per category was large, the recognition accuracy was high, but it significantly increased the time cost. Maray et al. [20] proposed a novel SFODTL-AHCR model for recognizing handwritten Arabic characters. Specifically, the pre- trained DWNN model was fine-tuned using the SFO algorithm to adapt to the task of Arabic character recognition. This method requires extensive preprocessing work, which slows down the recognition efficiency.

In contrast to the recognition of languages with relatively simple character structures such as Chinese characters and characters, the application of transfer learning in complex scripts is relatively limited.In the field of woodblock character recognition, Priya et al. [21] proposed the SLOA-TL model and used transfer learning methods to recognize multiple Tamil character objects in a single image, improving the accuracy of character recognition in a short period of time. However, the recognition process requires high-quality input samples, which entails significant effort in image preprocessing.In the domain of handwritten character recognition, Pande et al. [22] transferred a pre-trained CNN model to the task of recognizing handwritten Devanagari characters, implementing a DHTR system for Devanagari character recognition. This approach addressed the issue of ambiguity in word recognition. However, the method involved extensive preprocessing and post-processing operations, significantly slowing down the recognition efficiency. Rasheed et al. [23] proposed an AlexNet model based on data augmentation and transfer learning for recognizing handwritten characters and digits in Urdu language. Compared to existing classification models, it demonstrated better recognition performance. However, the training process required a large amount of data samples, increasing the complexity of the data model.Madhu et al. [24] achieved Kannada character recognition in handwritten script using convolutional neural networks and transfer learning. It improved the recognition accuracy compared to existing models. However, various preprocessing operations such as image cropping, padding, denoising, and resizing were performed before extracting features with the convolutional neural network, which slowed down the recognition efficiency.

In existing exploratory research on transfer learning, significant attention has been given to the following directions: 1) the degree of fine-tuning of pre-trained models, 2) few-shot transfer learning, and 3) zero-shot transfer learning. To address the question of how much fine-tuning should be applied to the pre-trained convolutional neural network when training the target model, Goel et al. [25] employed three transfer learning strategies to demonstrate the impact of different fine-tuning strategies on model performance. They discovered that connecting two newly adapted fully connected layers to the final convolutional layer of the pre-trained model and fine-tuning only the fully connected layers improved model performance and significantly reduced training time.

To tackle the issue of model adaptability in few-shot transfer learning, Elaraby et al. [26] proposed a novel Siamese network that utilized transfer learning to construct

a pre-trained AlexNet model, replacing the original Siamese CNN network. They employed a contrastive loss instead of the traditional binary cross-entropy loss, resulting in higher recognition performance compared to traditional Siamese models, while also reducing training time.

In the domain of zero-shot transfer learning, Pham et al. [27] introduced a combination scaling method called BASIC, which aims to narrow the gap between different domains during the transfer process. This method does not require any labeled data and achieved a recognition accuracy of 85.7% on the ImageNet dataset. However, when the dissimilarity between the source and target domains is significant, the recognition performance is unsatisfactory, necessitating intervention from domain experts.

C. INCREMENTAL LEARNING IN CHARACTER RECOGNITION

Incremental recognition is a strategy or method for incrementally processing new data on the basis of an existing model to enable incremental updating and improvement of the model without re-processing the historical data. Incremental recognition is mainly implemented using incremental learning, which consists of three basic types: task incremental, domain incremental, and class incremental [28]. In traditional machine learning, when a model is used to process new data, it usually forgets the previously learned knowledge. This is due to the fact that traditional machine learning methods use only current data during training and lack a mechanism to retain previously learned knowledge. To address the knowledge- forgetting problem, researchers have proposed a series of storage-based approaches for mitigating catastrophic forgetting since the 1990s. French et al. [29] explored the causes, and consequences, and proposed a number of solutions, the predecessor of incremental learning.

In the field of class-incremental learning, Ao et al. [30] proposed a novel zero-shot handwriting recognition method called CMPL, which utilizes a specialized modality embedding network to classify handwritten data based on the prototype of printed characters. This method enables recognition of unknown classes. However, its accuracy exhibits a significant gap compared to known classes. Thus, effective training strategies need to be explored to enhance open space generalization.To address the catastrophic forgetting issue in class- incremental learning, Yao et al. [31] proposed an adaptive memory update mechanism and novel loss methods. Specifically, when the first forgetting occurs, they remedy it by exchanging more distinctive samples from the long-term memory established during the early training process. However, employing such an approach increases model complexity, and determining the conditional boundaries generated by forgetting becomes challenging.

In the field of few-shot incremental learning, Wang et al. [32] proposed a distance-based hybrid classifier for the FSCIL (Few-Shot Class-Incremental Learning) task. They utilized supervised information from both cosine space and Euclidean space to design an SCIL module that combines the knowledge learned in a single SAR ATR task, thereby avoiding catastrophic forgetting of old knowledge. However, their training strategy falls under the category of pseudo-incremental learning, and as the number of incremental classes increases, the recognition accuracy of the classifier gradually decreases. Similarly, Cui et al. [33] addressed the FSCIL task by proposing a semi-supervised learning framework called UaD-CE, which consists of CE (Confident Ensemble) and UaD (Uncertainty-aware Distillation) modules. They combined self-training with unlabeled samples and class-balanced training to mitigate overfitting and bias issues using the CE module. Additionally, they designed the UaD module with uncertainty-guided refinement and adaptive distillation to address catastrophic forgetting. However, finding a balance between old and new data becomes a challenge as the proportion of unlabeled samples needs to be balanced during each incremental learning process.To address the trade-off between old and new data in the incremental learning process, researchers have proposed a series of knowledge distillation-based methods. These include Adaptive Feature Integration [34], Ranking Loss Algorithm [35], PSHT Loss [36], and Pretrained Model Knowledge Distillation [37]. These methods have played a positive role in mitigating catastrophic forgetting and overfitting issues. However, they involve extensive computations on redundant data, which negatively impact model efficiency. To further investigate the stability issue during the incremental evolution process, Zhang et al. [38] introduced an Incremental Concept Tree (ICT) stability analysis method. This method visualizes the incremental learning process in the form of an ICT to study the stability of incremental learning. The ICT represents concepts and their relationships at different time points as a tree-like structure. It allows researchers to observe the changes in knowledge during incremental learning, including the introduction of new knowledge, the forgetting of old knowledge, and the variations in inter-knowledge relationships. This approach aids researchers in gaining a more comprehensive understanding of the dynamic nature of incremental learning and provides guidance for improving incremental learning algorithms and optimizing model performance.

III. METHODOLOGY

A. OVERARCHING FRAMEWORK

The general framework of the migration learning-based incremental recognition method for multi-style Tibetan character ding is illustrated in Figure 3. The framework contains a CNN Tibetan script style classifier, three ResNet50 Tibetan character ding image feature extractors, and three Tibetan character ding image classifiers. Among the Tibetan character ding image feature extractors, which are subdivided into Tibetan Uchen Variants feature extractors, Tibetan Uchen Standard feature extractors, and Tibetan Uchen Antiquarian character ding feature extractors, only the Tibetan Uchen Standard feature extractors are involved in the feature learning, while the Tibetan Uchen Variants feature extractors and the Tibetan

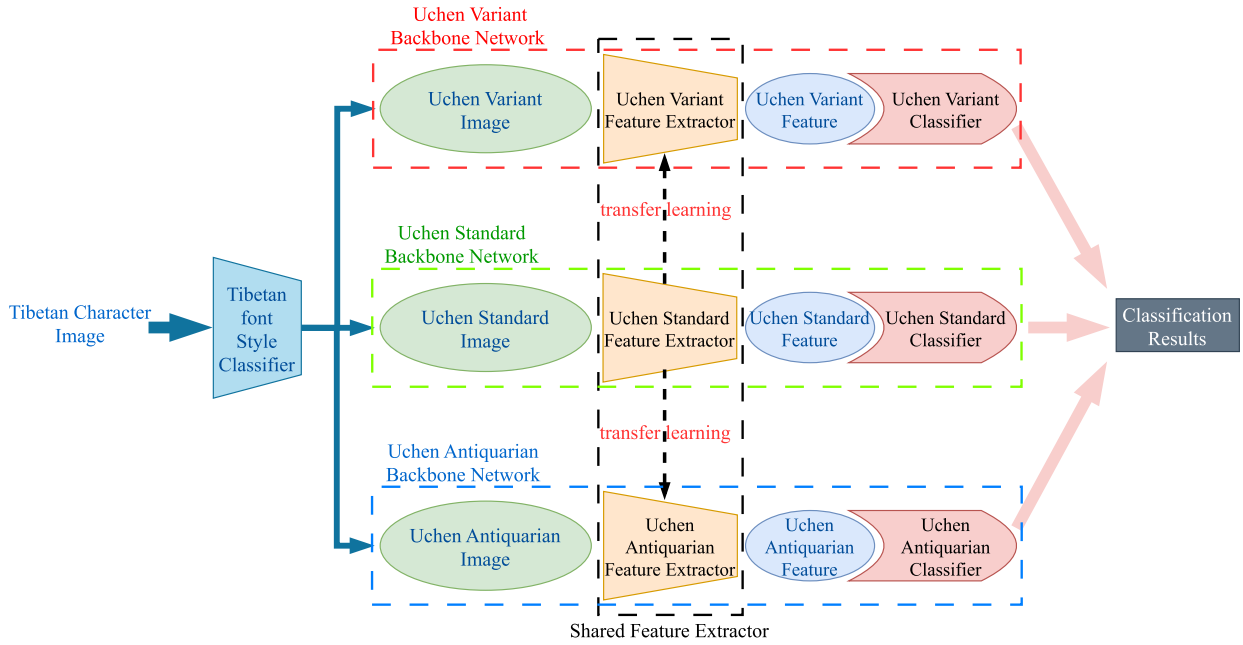


FIGURE 3. Framework of incremental recognition of multi-style Tibetan character ding based on transfer learning.

Uchen antique character ding feature extractors are based on the Tibetan Uchen Standard feature extractors, and utilize the migration Learning freezes the weight information of the backbone network, i.e., the feature extractors of Tibetan Uchen Variants, Tibetan Uchen Standard and Tibetan Uchen Antiquarian Character Ding are all the same, which is called the shared feature extractor. The realization process of this method is as follows:

- A Tibetan font style classifier is trained based on traditional CNN, which is used to categorize Tibetan Uchen Variants, Tibetan Uchen Standard and Tibetan Uchen Antiquarian character dings.
- To train a backbone network with good recognition accuracy on Tibetan Uchen Standard character images, which takes advantage of the normality of the Tibetan Uchen Standard character images.
- Utilizing the transfer learning method, the weights of the Tibetan Uchen Standard feature extractor are frozen and cloned into the Tibetan Uchen Variants feature extractor and the Tibetan Uchen Antiquarian character ding feature extractor.
- Based on different recognition tasks, to train their corresponding classifiers for incremental recognition of different styles of Tibetan character ding images.

The design of backbone network training and classifiers for Tibetan font style classifiers, Tibetan Uchen Standard, Variants, and Antiquarian character ding images are described in detail below.

B. TIBETAN FONT STYLE CLASSIFIER

To design the Tibetan font style classifier, considering the impact of the number of samples on the speed of computation, and hoping that the classifier will perform feature extraction automatically to strengthen its ability to expression and generalization of features, we choose VGG16 as the classifier in the task of recognizing different styles of Tibetan fonts, as shown in Figure 4.

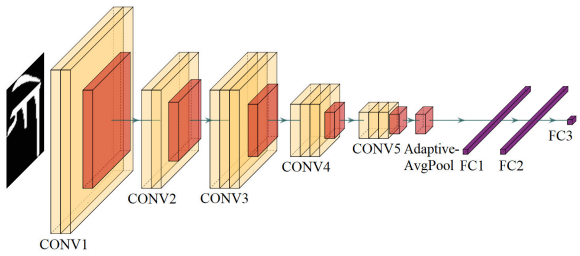


FIGURE 4. Tibetan font style classifier structure.

The input to the classifier is a three-channel image of size 224 × 224. The input image is passed through 13 convolutional layers, each of which is followed by a ReLu activation function and a 2 × 2 max pooling operation is performed periodically to reduce the spatial size of the feature map. The resulting feature map is then spread and fed into three fully connected layers for classification. The ReLu activation function is also applied after each fully connected layer. In addition, Dropout regularization is used to prevent overfitting. Finally, the output layer produces the output of the network, i.e., a probability distribution over the different categories, indicating the likelihood that the input image belongs to each category.

In the aforementioned process, Dropout regularization is applied to the fully connected layer to prevent overfitting in

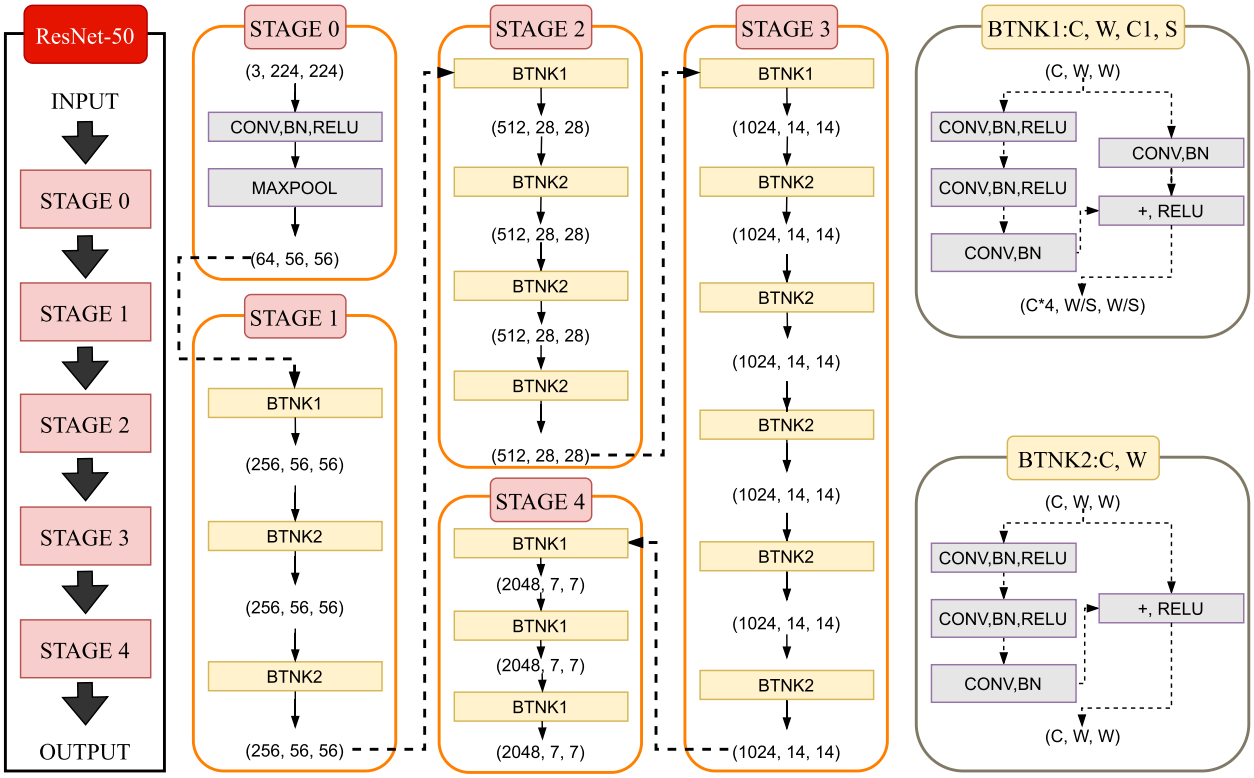


FIGURE 5. Structure of tibetan uchen standard feature extractor.

the neural network. Dropout regularization randomly "turns off" a portion of the neuron outputs, forcing different parts of the network to independently learn useful features. This reduces the complexity of the network and the dependency among parameters, thereby preventing overfitting. During the training process, for each output of the fully connected layer, a certain percentage (50% in this case) of the neuron outputs are randomly set to zero, effectively "turning off" those neurons to prevent overfitting, improve model robustness, and reduce co-adaptation among neurons.

C. TIBETAN UCHEN STANDARD BACKBONE NETWORK TRAINING

1) TIBETAN UCHEN STANDARD FEATURE EXTRACTOR

Considering the ease of accessibility and a large dataset of Tibetan Uchen standard images, as well as the strong regularity of the Tibetan Uchen standard, we employ the widely used and high- performance ResNet50 as the feature extractor for Tibetan Uchen standard image recognition, as shown in Figure 5. The input to the feature extractor is a preprocessed three-channel image with a size of 224 × 224. The preprocessing includes normalization and color channel standardization. The input image is first passed through an initial convolutional layer (STAGE0) for feature extraction. This layer consists of 64 convolutional kernels of size 7 × 7 and is followed by batch normalization (BN) and ReLU activation function. Subsequently, a stack of residual blocks, composed of four stages (STAGE1/2/3/4), is employed. Each stage contains multiple residual blocks. Each residual block consists of several convolutional layers and batch normalization layers (BN) and includes a skip connection to learn residual mapping. These residual blocks progressively learn more abstract and high-level feature representations. Following the residual blocks, the feature maps are downsampled through average pooling layers, gradually reducing the size of the feature maps. The final output of the network is the output feature representation. Each of the above residual blocks contains within it multiple convolutional and batch normalization layers for performing convolutional operations on the input feature maps and accelerating the training process.

Each residual structure has a jump connection as shown in Figure 6, also known as Identity Mapping. The jump connection adds the input feature map directly to the output of the residual block so that information is passed directly while avoiding information loss.

Suppose the input feature map size is $H \times W \times C$, in which is the height H, W is the width, and C is the number of channels. The input feature map is first subjected to a convolution operation of size 1×1 for dimensionality reduction to reduce the number of channels, and the output feature map size of this convolution layer is kept as $H \times W \times C/4$. Subsequently, a convolution operation of size 3×3 is subjected to feature extraction, and the output feature map size of this convolution layer is still $H \times W \times C/4$. Another

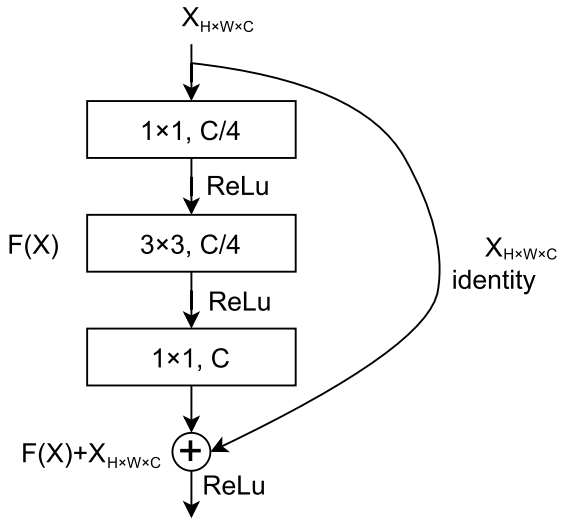


FIGURE 6. Residual structure diagram.

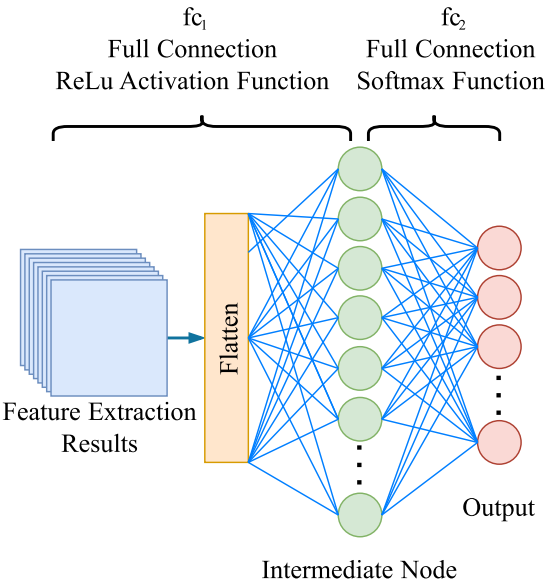


FIGURE 7. Structure of tibetan uchen standard classifiers.

convolution operation of size 1×1 is used for dimensionality uplift to restore the number of channels to C, and the output feature map size is $H \times W \times C$. Finally, the input feature map $X_{H\times W\times C}$ is directly added to the output feature map $F(X)$ of the third convolutional layer to form the output $Output_{block}$ of the residual structure, i.e., the final output result of the input original feature map after the residual structure can be formally expressed as:

$$Output_{block} = F(X) + X_{H\times W\times C} \tag{1}$$

where $Output_{block}$ denotes the output of the residual structure, $F(X)$ denotes the output feature map after three layers of convolution in the residual structure, and $X_{H\times W\times C}$ denotes the original input feature map.

2) TIBETAN UCHEN STANDARD CLASSIFIER

To design the classifier for the Tibetan Uchen Standard, considering that the feature extractor part utilizes the deep convolutional layer and residual block structure to extract the high-level feature representation from the Tibetan Uchen Standard image, if the complex neural network as the classifier for this recognition task based on this, it will introduce additional complexity and have some impact on the recognition effect. Therefore, we choose the simpler fully connected layer as the classifier for this recognition task, and the classifier structure is shown in Figure 7.

The feature map generated from the original input image after passing through the feature extractor is firstly converted into a one-dimensional vector after a spreading operation, which is used as input to the fully connected layer fc_1. At the same time, the ReLu function is utilized to increase the nonlinear relationship between the layers of the neural network. Subsequently, the results are computed by the fully connected layer fc_1 and passed to the fully connected layer fc_2. At the same time, the Softmax function is applied to transform the output results into probability distributions representing the probabilities of each category, so that the sum of the probabilities of all the categories equals to one. All the fully connected layers mentioned above are internally composed of a number of ''neurons'', and each ''neuron'' is a ''neuron'' that is a ''neuron'' in the neural network. All of the above fully connected layers are internally composed of multiple ''neurons'', each of which has a connection weight to each element of the input feature.

In this recognition task, the advantages of utilizing a fully connected layer as a classifier are as follows:

- The lower number of parameters makes the fully connected layer more efficient in the training process.
- The absence of complex local connectivity patterns makes the fully connected layer easier to train and optimize, reducing the risk of problems such as vanishing gradients or gradient explosion.
- High flexibility in adapting to different dimensions of input features and handling input features of arbitrary dimensions makes the fully connected layer more robust.
- Capable of feature integration and representation, the previous layer of features can be integrated and combined to provide a higher level of feature representation by learning the weighting parameters, which can enhance the classification performance of the model and capture a more discriminative feature representation.

D. TIBETAN UCHEN VARIANTS AND ANTIQUARIAN CHARACTER DING BACKBONE NETWORK TRAINING

In the conventional multi-task recognition paradigm, different styles of Tibetan character ding samples are typically integrated into the same feature space for recognition training. However, this approach requires a large amount of data samples to learn the differences between different style features, and it is difficult to collect a sufficient number of

samples from the Tibetan Uchen variants and antiquarian character ding image datasets. Additionally, retraining the entire model would consume a significant amount of time and resources. Taking into consideration that there are shared features and patterns among multiple fonts, such as the shape and structure of Tibetan script components, we propose a transfer learning strategy to train the backbone network for the Tibetan Uchen variants and antiquarian character ding. The specific architecture of the backbone network is illustrated in Figure 8. It is important to note that the training process for the backbone network of the Tibetan Uchen variants and antiquarian character ding is identical. The only modification required is in the design of the specific parameters of the classifier, which depends on the corresponding number of categories.

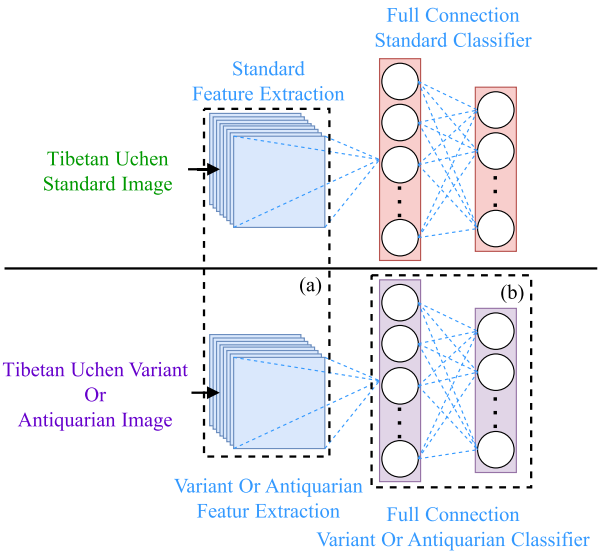


FIGURE 8. Structure of the tibetan uchen variants and the antiquarian character ding Backbone Networks.

In this method, the training process of the backbone network of Tibetan Uchen Variants and Antiquarian character ding is divided into two stages:

Stage 1. Freezing the feature extractor of the Tibetan Uchen standard, i.e., setting its weight parameter to be fixed, as the feature extractor of the Tibetan Uchen variants and the Antiquarian character ding by utilizing the transfer learning strategy.

Stage 2. Train a classifier in a new task using the fine-tuning method in transfer learning to improve the model's recognition performance on the new task.

1) TIBETAN UCHEN VARIANTS AND ANTIQUARIAN CHARACTER DING FEATURE EXTRACTOR

The feature extractors of the Tibetan Uchen variants and antiquarian character ding are exactly the same as those of the Tibetan Uchen standard. It mainly adopts the freezing operation in transfer learning to freeze the weights of the feature extractor of the Tibetan Uchen standard and migrate them to the recognition task of the Tibetan Uchen variants and the antiquarian character ding, as shown in part (a) in Figure 8.

After the original input images of different domains have gone through the feature extractors under their respective different tasks, we visualize the Tibetan Uchen standard, variants, and antiquarian character ding images using the inverse convolution method, as shown in Figure 9.

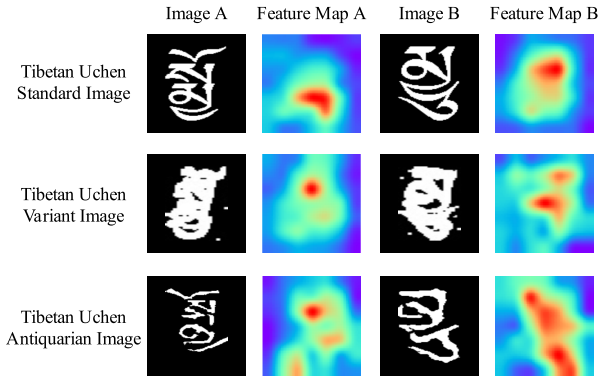


FIGURE 9. Visualization of the characteristics of different styles of Tibetan style.

For each dimension of the 2048-dimensional feature, we can obtain a feature map. To better observe the distribution of feature regions, we fuse the 2048 feature maps into a single feature map by summing them at each spatial position. It can be observed that whether it is Tibetan Uchen standard, Tibetan Uchen variants, or Tibetan Uchen antiquarian character ding, the learned features are concentrated in the text regions (rather than the background regions), reflecting the inherent structural characteristics of the text. Moreover, the feature maps of different character dings of the same style and different styles of the same character ding show significant variability, highlighting the effectiveness of using transfer learning for shared feature extraction.

2) TIBETAN UCHEN VARIANTS AND ANTIQUARIAN CHARACTER DING CLASSIFIER

To classify the Tibetan Uchen variants and antiquarian character ding in a new task, we employed the fine-tuning strategy in transfer learning, specifically targeting the pre-trained ResNet50 model on Tibetan Uchen standard images, as illustrated in Figure 8(b).

Initially, we loaded the pre-trained ResNet50 model and froze its feature extraction layers. This approach aimed to preserve the ResNet50's learned generic feature representations from a large-scale dataset and mitigate the risk of overfitting on the new task. The weights of the feature extraction layers were retained and set as non-trainable. Subsequently, we added a brand-new fully connected layer on top of the ResNet50 model, with the number of output nodes matching the categories of the Tibetan Uchen variants and antiquarian character ding. This fully connected layer

TABLE 1. Dataset details.

Dataset title	No. of class	No. of new class	No. of sample	Sample scope	Fonts include	Special handling
Uchen standard	500	None	88109	86-171	19	None
Uchen variants	584	169	23617	27-29	12	Expansion, corrosion, noise, etc.
Uchen antiquarian	610	110	21402 18	6-39	None	None

would be responsible for precise classification based on the extracted features.

During the fine-tuning process, we selected the cross-entropy loss function as the objective function, as shown in equation (2). The stochastic gradient descent (SGD) optimizer was employed to update the parameters of the fully connected layer. The annotated data of the Tibetan Uchen variants and antiquarian character ding were used to train the fully connected layer. Through backpropagation and optimization algorithms, the loss was minimized, and the parameters of the fully connected layer were updated.

$$Loss = -\frac{1}{N}\sum_{n=0}^{N-1}\sum_{c=0}^{C-1} y_{n,c}\log\left(p_{n,c}\right) \tag{2}$$

where $y_{n,c}$ is the labeling information of image n and $p_{n,c}$ is the cth element in the output result after the image has undergone feature extraction.

Finally, we further improve the performance of the classifier by evaluating the performance of the fine-tuned model on a test set and making the necessary adjustments and hyper-parameter tuning, and this iterative process ensures that we obtain the best model performance.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

The Tibetan character ding datasets used in this paper are TCDB and HUTD [39] provided by Northwest Minzu University, in which the TCDB dataset contains the standard and variant data of Tibetan Uchen, and HUTD is the data of Tibetan antiquarian character ding, as shown in Table 1. This dataset is divided into three parts in total, which are Tibetan Uchen standard dataset, Tibetan Uchen variants dataset and Tibetan Uchen antiquarian character ding dataset.

The font files used in this dataset are exclusively in the Uchen style, i.e., Uchen standard and Uchen variant script uses distinct font files. In order to assess the recognition capability of the proposed method on unknown categories and demonstrate its generalization and scalability, additional character classes that do not appear in the standard dataset were introduced into the variant and antiquarian character ding datasets. The number of samples per class in each dataset is illustrated in Figure 10.

All samples in this dataset are black-background white-character images with varying sizes. When inputting them into the corresponding network, size scaling operations need to be performed according to the network's input requirements. In the Tibetan Uchen standard recognition task, we utilized a Tibetan Uchen standard dataset consisting of 610 character classes to train the shared feature extractor

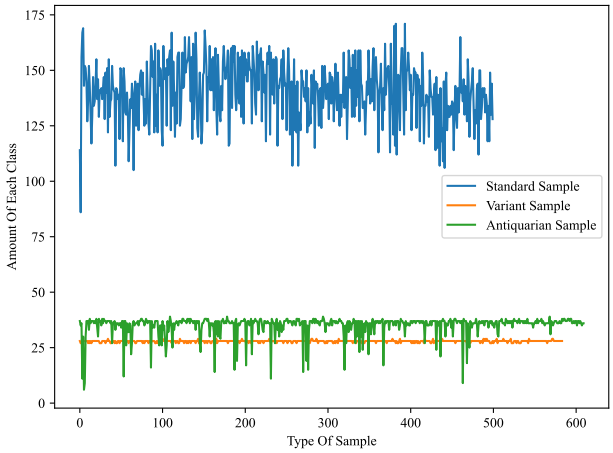


FIGURE 10. The distribution of sample counts for different styles of Tibetan.

(shared among standard, variants, and antiquarian character ding) and the standard classifier. For the Tibetan Uchen variants and antiquarian character ding recognition tasks, we employed the corresponding Tibetan Uchen variants dataset (584 character class) and Tibetan Uchen antiquarian character ding dataset (610 character class) to train the Tibetan Uchen variants and antiquarian character ding classifiers. As the feature extractor adopts a shared feature extraction mechanism, there is no need to retrain the feature extraction component. In the multi-style Tibetan script incremental recognition task, we utilized the Tibetan Uchen standard, variants, and antiquarian character ding datasets to train our Tibetan style classifier, aiming to achieve incremental recognition.

A. RECOGNITION OF TIBETAN UCHEN STANDARD BASED ON RESNET50

The ResNet50 model consists of convolutional layers, residual blocks, global average pooling layers, and fully connected layers. In this study, we constructed a ResNet50 model tailored for Tibetan Uchen standard recognition, which comprised 49 convolutional layers, 4 residual blocks, and 1 global average pooling layer. The specific recognition process is illustrated in Figure 11.

The input image is first passed through convolutional layers to extract its features. The extracted features are then propagated and enhanced through a series of residual blocks, which facilitate information transfer. After the last residual block, the multi-dimensional feature maps are

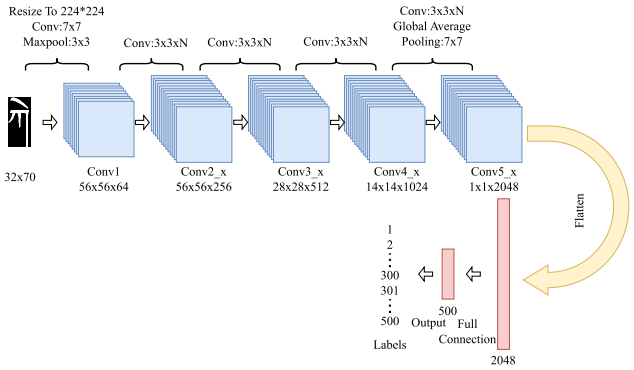


FIGURE 11. Architecture for Recognizing Tibetan Uchen Standard Based on ResNet50.

transformed into a one-dimensional feature vector using a global average pooling layer. This one-dimensional feature vector is then fed into a fully connected layer, where a non-linear transformation is applied using an activation function to obtain the final classification result.

To validate the effectiveness of the ResNet50-based Tibetan Uchen standard recognition model, traditional SVM linear classifier, FNN feedforward neural network, and CNN convolutional neural network are adopted as baseline models. To ensure experimental fairness, each model employs the same dataset.

The recognition accuracies of the four different Tibetan Uchen standard script recognition models are presented in Table 2.

TABLE 2. Comparison of the recognition accuracy of different models for the recognition of tibetan standard.

Methods and models	Recognition accuracy (%)
Support vector machine	78.25
Feedforward neural network	89.75
Visual geometry group-16	96.12
Residual network-50	**99.04**

As shown in Table 2, the recognition results of the ResNet50 residual network surpass those of the baseline models. Compared to the baseline models, the ResNet50 residual network achieves a significant improvement in recognition accuracy on the test set, increasing from 96.12% to 99.04%. This demonstrates the effectiveness and high accuracy of the ResNet50 residual network for Tibetan Uchen standard recognition.

Furthermore, to evaluate the performance of the ResNet50 residual network in Tibetan Uchen standard recognition and gain better understanding of the model's robustness and adaptability, we computed the top 1, 5, and 10 accuracy rates during recognition. These metrics measure the percentage of correctly predicted results when considering the top 1, 5, and 10 predictions according to their confidence, i.e., predictive reliability, and comparing them with the true labels. The specific recognition results are presented in Table 3.

TABLE 3. Recognition accuracy at different confidence levels.

Confidence level	Recognition accuracy (%)
Top 1	99.04
Top 5	99.23
Top 10	99.23

According to Table 3, it is evident that ResNet50 demonstrates a high level of accuracy in classifying the Tibetan Uchen Standard dataset. Specifically, we observed that the model achieved a remarkable performance in terms of Top-1 accuracy, reaching 99.04%. This implies that ResNet50 is capable of accurately matching the true label with a very high proportion in a single highest- confidence prediction result. Furthermore, we also discovered that ResNet50 achieved a Top-5 and Top- 10 accuracy of 99.23%, indicating that it is sufficient to correctly identify 99.23% of samples within the top five candidate predictions.

The aforementioned experiments clearly demonstrate the effectiveness of the ResNet50 model in the task of Tibetan Uchen Standard recognition, and its recognition accuracy is significantly remarkable.

B. RECOGNITION OF TIBETAN UCHEN VARIANTS AND ANTIQUARIAN CHARACTER DING BASED ON TRANSFER LEARNING

In this experimental section, a transfer learning approach is employed to accomplish the recognition task of Tibetan Uchen variants and antiquarian character ding. The feature extraction layer of the pre-trained ResNet50 network is frozen and utilized as the feature extractor for this specific recognition task. Simultaneously, two new classifiers, namely the Tibetan Uchen variant classifier and the antiquarian character ding classifier, are trained. The specific network architecture for this task is depicted in Figure 12.

After the input image undergoes classification filtering, it is passed through the corresponding feature extractor. The feature extractor is generated by employing the transfer learning method's freezing technique using the Uchen Standard feature extractor. Subsequently, the resulting feature extraction outcomes are fed into the respective fully connected layers for the classification task. Nonlinear transformations are applied using activation functions, ultimately yielding the classification results.

To validate the effectiveness of the transfer learning approach, this section adopts LeNet-5, VGG-16, and a single ResNet50 as baseline models. The recognition accuracy and training time per epoch under different models serve as evaluation metrics. To ensure experimental fairness, the same dataset is employed for each model.

The recognition accuracy of the three different recognition models is presented in Table 4.

As shown in Table 4, the recognition results based on the transfer learning method using ResNet50 outperform

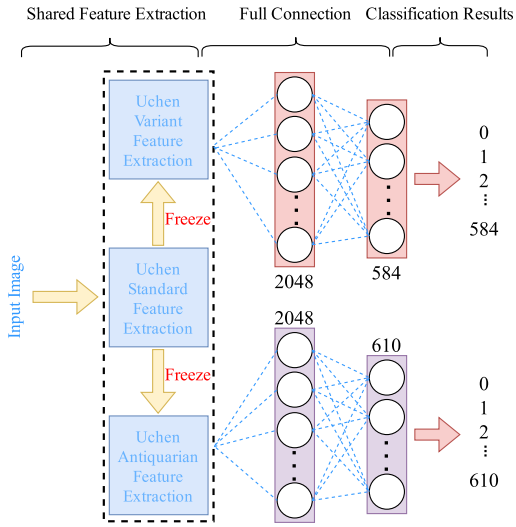


FIGURE 12. Architecture for recognizing tibetan uchen variants and antiquarian character ding.

TABLE 4. Accuracy of recognizing tibetan uchen variants and antiquarian character ding under different models.

Methods	Uchen variants		Uchen antiquarian	
	Accuracy(%)	Time(h)	Accuracy(%)	Time(h)
LeNet-5	54.35	0.19	51.89	0.27
VGG-16	91.56	0.85	90.67	0.98
Resnet50	96.33	0.45	93.24	0.52
Our method	**98.05**	**0.15**	**97.66**	**0.22**

the baseline model. Compared to the baseline model, the recognition accuracy of the ResNet50 transfer learning method on the test set has improved from 96.33% and 93.24% to 98.05% and 97.66%, respectively. Additionally, the training time per epoch has been reduced from 0.85 hours and 0.98 hours to 0.15 hours and 0.22 hours using the ResNet50 transfer learning method.

We observed that using our method, the accuracy rates have been increased by 1.72% and 4.42% compared to the standalone ResNet50 model. This improvement is attributed to the scarcity of samples for the Tibetan Uchen variants and the antiquarian character Ding, as well as the high number of categories. Traditional neural networks struggle to extract features effectively from limited samples, resulting in poor differentiation between different character ding classes. In contrast, the proposed method in this paper utilizes the feature extraction capability of the pre-trained feature extractor on the old task to extract features from unknown samples in the new task without specific weight training for the new recognition task.

Furthermore, it is noteworthy that our method significantly reduces the training time compared to the standalone ResNet50 network. The training time per epoch has decreased from 0.45 hours and 0.52 hours to 0.15 hours and 0.22 hours, respectively. This demonstrates that our method not only enhances the recognition accuracy of the new task based on existing knowledge but also substantially reduces the training duration, thereby reducing time costs.

Moreover, to avoid excessive reliance on the highest probability option, evaluate model uncertainty, and provide more information for subsequent processing and decision-making, we also calculated the Top1, Top5, and Top10 accuracy rates during recognition. The specific recognition results are presented in Table 5.

TABLE 5. Recognition accuracy of variants and antiquarian character ding at different confidence levels.

Confidence level	Accuracy of variants(%)	Accuracy of antiquarian(%)
Top 1	99.05	97.66
Top 5	98.35	98.77
Top 10	99.18	99.04

Based on the experimental results in Table 5, we can observe that regardless of whether it is the Tibetan Uchen variants or the antiquarian character ding, as the confidence level increases, the recognition accuracy also increases. This implies that for certain samples, the class with the highest probability may not completely match the true class, but the model is still able to find the correct class within the top 5 or top 10 options. This indicates the presence of similarities among samples from different classes, resulting in some ambiguity or uncertainty for certain samples, making it impossible for the model to provide a perfectly accurate class solely based on the highest probability. However, when the model provides candidate classes (Top 5/Top 10), more comprehensive predictive information can be obtained, highlighting the accuracy of the model.

Through the aforementioned experiments, the effectiveness of transfer learning is demonstrated. By utilizing existing models and knowledge, recognition performance can be improved under different stylistic variations of characters. Moreover, sensitivity to the quantity of newly added style data samples is low, as only a small number of samples are required to efficiently complete the recognition task. By freezing the weight parameters of the feature extraction layer that are pre-trained on other similar tasks and retraining (also known as fine-tuning) the classifier for the new task, training time is significantly reduced, thereby reducing training costs. Therefore, employing transfer learning for training new tasks is not only effective but also efficient to a certain extent.

C. INCREMENTAL RECOGNITION OF MULTI-STYLE TIBETAN

Based on the proposed method in this paper, incremental recognition of Tibetan characters in multiple styles can be achieved. The recognition task of Tibetan characters in different styles is divided into three parts:

Part 1: Transfer learning is employed, utilizing pre-existing models and knowledge from previous tasks to transfer to the

new task and accomplish recognition tasks under different font styles.

Part 2: By utilizing a shared feature extractor and combining it with a font style classifier, the incremental recognition task of Tibetan characters in multiple styles is accomplished.

Part 3: Building upon Part 2, recognition of unknown categories is performed.

In this section, the focus is primarily on conducting experiments for Part 2 and Part 3, which involve the incremental recognition of Tibetan characters in multiple styles. The experimental results are subsequently analyzed and explained.

1) MULTI-STYLE INCREMENTAL RECOGNITION

The flowchart for this section of the experiment is shown in Figure 13.

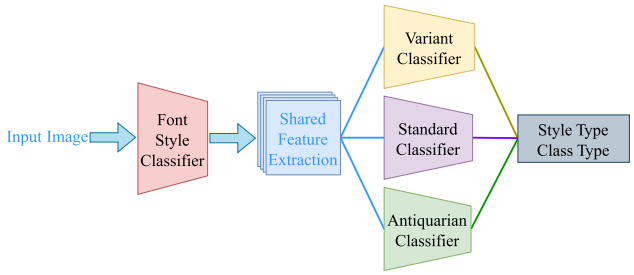


FIGURE 13. Architecture for incremental recognition of multi-style Tibetan.

After performing classification filtering, the resulting data undergoes feature extraction using the corresponding feature extractor. The feature extractor is generated by the Uchen standard feature extractor using the freezing method in transfer learning. Subsequently, the extracted features are fed into the corresponding fully connected layers for the classification task, undergoing nonlinear transformations through activation functions to ultimately obtain the classification results.

The input image is initially passed through a font-style classifier built using the VGG-16 network. Once the font style of the input image is classified, it is passed into the shared feature extractor for feature extraction. The resulting feature map is then passed into the classifier corresponding to the font style category for character classification. Finally, the output consists of the style category information and character category information of the input image.

To validate the effectiveness of this approach, a comparative experiment is set up using traditional training methods, which are currently the most commonly used methods in research. Specifically, multiple Tibetan character datasets with different font styles are merged into a single dataset. Each sample in this dataset has two labels: the font style label and the character category label. During the model training process, both tasks, i.e., font style classification and character category classification, are simultaneously performed. For the selection of the comparative experimental models, ResNet34 and MobileNetV3 are chosen as the baseline models. The recognition accuracy of each recognition method and model is shown in Table 6.

TABLE 6. Recognition accuracy of each recognition method and model.

Methods	Accuracy of font style(%)	Accuracy of character class(%)	Comprehensive accuracy(%)
MobileNetV3	96.63	91.31	80.03
ResNet34	98.12	97.11	90.14
Our method	**99.61**	**98.79**	**98.40**

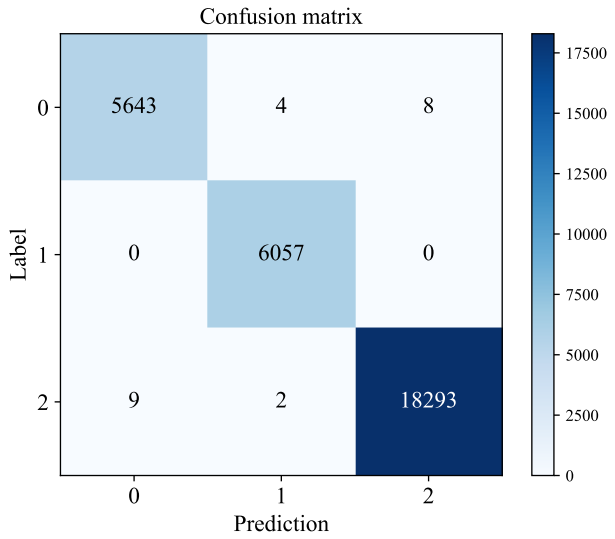
As shown in Table 6, it is evident that across various dimensions, including font style accuracy, character category accuracy, and comprehensive accuracy, the method proposed in this paper outperforms the baseline model. Specifically, the proposed method improves to 99.61%, 98.79%, and 98.40% for font style classification accuracy, character ding category classification accuracy, and overall accuracy, respectively. While the MobileNetV3 and ResNet34 models exhibit good recognition performance in font style and character category classification, they are prone to errors such as correctly classifying font styles but misclassifying character categories, or vice versa. As a result, their overall recognition accuracy is not high. In contrast, the proposed method leverages a font style classifier to accurately classify the style of an image and combines it with the corresponding character category classifier to achieve precise category classification. This approach not only achieves high recognition rates in the subtasks but also demonstrates excellent performance in the overall task.

To further illustrate the classification performance of our method, a confusion matrix is constructed using a randomly selected sample dataset of 30,000 samples. The confusion matrix, shown in Figure 14, visualizes the predictive performance and error patterns of our method.

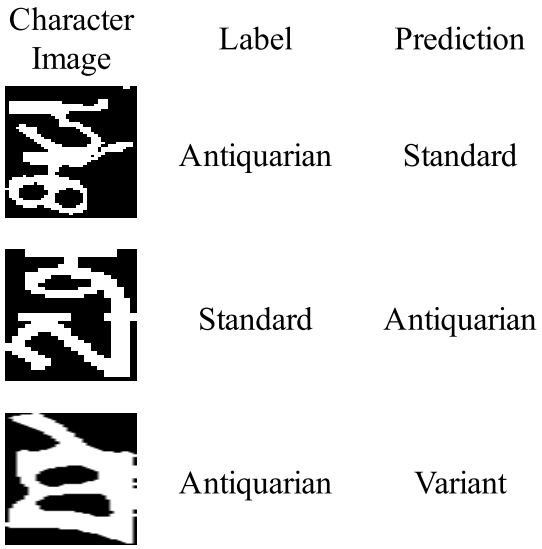
In Figure 14(a) of the predictions of tibetan style recognition, label 0 represents Tibetan Uchen Standard, label 1 represents Tibetan Uchen Variants, and label 2 represents Tibetan Uchen Antiquarian character ding. The confusion results are illustrated in Figure 14(b). From the figure, it can be observed that approximately 0.25% of Tibetan Uchen Antiquarian character ding samples were misclassified as Tibetan Uchen Standard, indicating a certain degree of similarity between these two font styles. Although the similarity is not high, it still has a certain impact on the recognition results. If the deep learning model can capture the differences between different styles during feature extraction, it would greatly enhance the overall recognition performance of the proposed method.

2) MULTI-CATEGORY INCREMENTAL RECOGNITION

In order to further investigate the ability of our method to recognize unknown classes, we expanded the Tibetan Uchen dataset by including 169 new classes in the Tibetan Uchen variant dataset and 110 new classes in the



(a) Confusion matrix of style recognition



(b) Example of obfuscated results

FIGURE 14. Predictions of tibetan style recognition.

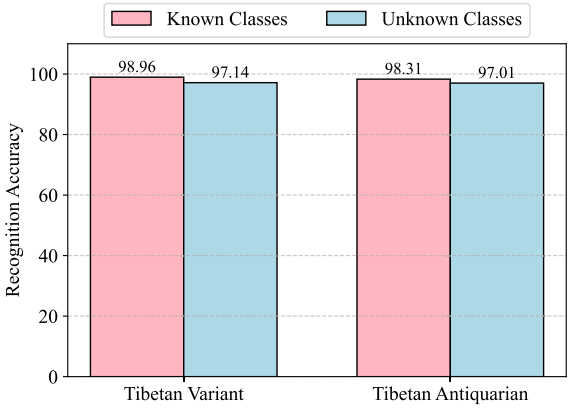


FIGURE 15. Recognition rate comparison between known and unknown categories.

Tibetan Uchen antiquarian character ding dataset. By utilizing the proposed transfer learning and incremental recognition methods in this paper, we successfully accomplished the recognition task of unknown classes in the new styles. The recognition results for the experiments in this section are displayed, as shown in Figure 15.

According to Figure 15, in the Tibetan variant dataset, the recognition rate for known classes reached 98.96%, and the recognition accuracy for unknown classes was 97.14%. In the antiquarian character ding dataset, the recognition accuracy for known classes was 98.31%, and for unknown classes, it was 97.01%. It can be observed that the recognition accuracy for unknown classes with the addition of new styles exceeded 97%, indicating that this method can still achieve a high level of recognition accuracy when faced with unseen classes. This demonstrates the successful transfer of existing knowledge of the model to new classes, effectively addressing the issue of imbalanced classes with new styles, and enabling efficient handling of recognition tasks for unknown classes.

V. CONCLUSION

We propose a transfer learning-based method for incremental recognition of multi-style Tibetan characters. The method consists of three components: 1) a Uchen Tibetan standard character recognition model based on residual networks; 2) a two-stage training strategy based on transfer learning; and 3) a multi-style incremental recognition approach. Our method utilizes transfer learning strategies to improve recognition accuracy when the sample size of newly added style characters is limited while reducing the time cost during the training process. By employing a shared feature extractor and task-specific classifiers, we address the challenges of multi-style and multi-class incremental recognition. Compared to traditional multi-task recognition approaches, our method exhibits stronger generalization and scalability, resulting in improved recognition accuracy and training efficiency.

However, there are some limitations that need to be addressed: 1) Due to the retention of a large number of redundant parameters in the transfer recognition training process, the model has a larger parameter size, which partially slows down the inference time; 2) In the incremental recognition process, joint training is employed, which achieves better recognition performance but increases the training cost to some extent by training on all known data; 3) This study uses samples of different styles of characters under the Uchen typeface, and in our future research, we will expand it to other Tibetan fonts, such as handwriting and Ume.

Although our method has only been applied and validated in the task of multi-style Uchen Tibetan character recognition, it can provide effective solutions for scenarios facing the following challenges: 1) excessive variation between styles;

2) limited sample size of newly added styles; 3) recognition of the different languages. This method can be applied in the following domains: a) multi-language recognition systems; b) few-shot recognition; and c) single-image, multi-text recognition.

ACKNOWLEDGMENT

The authors thank all the anonymous reviewers for their insightful comments.

REFERENCES

[1] T. Kumar, C. Khosla, and K. Vashistha, "Character recognition techniques using machine learning: A comprehensive study," in *Proc. Int. Conf. Appl. Intell. Sustain. Comput. (ICAISC)*, Jun. 2023, pp. 1–6.
[2] L. R. Mursari and A. Wibowo, "The effectiveness of image preprocessing on digital handwritten scripts recognition with the implementation of OCR tesseract," *Comput. Eng. Appl. J.*, vol. 10, no. 3, pp. 177–186, Oct. 2021.
[3] X. Wang, H. Du, and X. Wen, "Research on segmentation and recognition of printed Chinese characters," *J. Phys., Conf. Ser.*, vol. 1237, no. 2, 2019, Art. no. 022011.
[4] B. Wang, Y. W. Ma, and H. T. Hu, "Hybrid model for Chinese character recognition based on tesseract-OCR," *Int. J. Internet Protocol Technol.*, vol. 13, no. 2, pp. 102–108, 2020.
[5] Y. T. Wu, E. Fujiwara, and C. K. Suzuki, "Image-based radical identification in Chinese characters," *Appl. Sci.*, vol. 13, no. 4, p. 2163, Feb. 2023.
[6] X. Yao, H. Sun, S. Li, and W. Lu, "Invoice detection and recognition system based on deep learning," *Secur. Commun. Netw.*, vol. 2022, pp. 1–10, Jan. 2022.
[7] J. Bati and P. R. Dawadi, "Ranjana script handwritten character recognition using CNN," *JOIV, Int. J. Informat. Visualizat.*, vol. 7, no. 3, pp. 984–990, Sep. 2023.
[8] W. Zhou, L. Chen, and Z. Zeng, "Tibetan character recognition based on geometric shape analysis," *Comput. Eng. Apps.*, pp. 201–205, 2012.
[9] Y. Wu, "Research on the Tibetan characters recognition based on H-KNN," *Mod. Inf. Technol.*, pp. 92–94, 2022.
[10] G. Zhang, W. Wang, C. Zhang, P. Zhao, and M. Zhang, "HUTNet: An efficient convolutional neural network for handwritten Uchen Tibetan character recognition," *Big Data*, vol. 11, no. 5, pp. 387–398, Oct. 2023.
[11] D. Zhao, "Research on wooden blocked Tibetan character recognition based on BP network," *Microprocess*, pp. 35–38, 2012.
[12] Y. Hou, D. Gao, and H. Gao, "Wujin printing multi-fonts text detection and recognition in Tibetan," *Comput. Eng. Des.*, pp. 1058–1065, 2023.
[13] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010.
[14] F. Yan, "Study on the font type recognition model of printed Chinese character using convolutional neural network based on transfer learning," *Print. Digital. Media. Technol. Study.*, pp. 36–45, 2021.
[15] Y.-Q. Li, H.-S. Chang, and D.-T. Lin, "Large-scale printed Chinese character recognition for ID cards using deep learning and few samples transfer learning," *Appl. Sci.*, vol. 12, no. 2, p. 907, Jan. 2022.
[16] P. Shi, Y. Lou, and R. Xia, "Handwritten Chinese character recognition based on morphology and transfer learning," in *Proc. Int. Conf. Intell. Perception Comput. Vis. (CIPCV)*, May 2023, pp. 47–51.
[17] N. Elaraby, S. Barakat, and A. Rezk, "A generalized ensemble approach based on transfer learning for Braille character recognition," *Inf. Process. Manag.*, vol. 61, no. 1, Jan. 2024, Art. no. 103545.
[18] A. F. Rizky, N. Yudistira, and E. Santoso, "Text recognition on images using pre-trained CNN," 2023, *arXiv:2302.05105*.
[19] A. Daood, A. L. I. Al-Saegh, and A. F. Mahmood, "Handwriting detection and recognition of Arabic numbers and characters using deep learning methods," *J. Eng. Sci. Technol.*, vol. 18, pp. 1587–1598, Jun. 2023.
[20] M. Maray, B. B. Al-Onazi, J. S. Alzahrani, S. M. Alshahrani, N. Alotaibi, S. Alazwari, M. Othman, and M. A. Hamza, "Sailfish optimizer with deep transfer learning-enabled Arabic handwriting character recognition," *Comput., Mater. Continua*, vol. 74, no. 3, pp. 1–16, 2023.
[21] S. Karthikeyan, I. Jaganathan, K. S. Rameshbabu, A. Abraham, L. A. Gabralla, R. Sivaraj, and S. M. Nandhagopal, "Self-adaptive hybridized lion optimization algorithm with transfer learning for ancient Tamil character recognition in stone inscriptions," *IEEE Access*, vol. 11, pp. 39621–39634, 2023.
[22] S. D. Pande, P. P. Jadhav, R. Joshi, A. D. Sawant, V. Muddebihalkar, S. Rathod, M. N. Gurav, and S. Das, "Digitization of handwritten devanagari text using CNN transfer learning—A better customer service support," *Neurosci. Informat.*, vol. 2, no. 3, Sep. 2022, Art. no. 100016.
[23] A. Rasheed, N. Ali, B. Zafar, A. Shabbir, M. Sajid, and M. T. Mahmood, "Handwritten Urdu characters and digits recognition using transfer learning and augmentation with AlexNet," *IEEE Access*, vol. 10, pp. 102629–102645, 2022.
[24] S. B. Madhu, C. V. Aravinda, and M. S. Sannidhan, "Handwritten Kannada character recognition using convolutional neural networks and transfer learning," *J. Phys., Conf.*, vol. 2571, no. 1, Oct. 2023, Art. no. 012012.
[25] P. Goel and A. Ganatra, "Handwritten Gujarati numerals classification based on deep convolution neural networks using transfer learning scenarios," *IEEE Access*, vol. 11, pp. 20202–20215, 2023.
[26] N. Elaraby, S. Barakat, and A. Rezk, "A novel Siamese network for few/zero-shot handwritten character recognition tasks," *Comput., Mater. Continua*, vol. 74, no. 1, pp. 1837–1854, 2023.
[27] H. Pham, Z. Dai, G. Ghiasi, K. Kawaguchi, H. Liu, A. W. Yu, J. Yu, Y.-T. Chen, M.-T. Luong, Y. Wu, M. Tan, and Q. V. Le, "Combined scaling for zero-shot transfer learning," *Neurocomputing*, vol. 555, Oct. 2023, Art. no. 126658.
[28] G. M. van de Ven, T. Tuytelaars, and A. S. Tolias, "Three types of incremental learning," *Nature Mach. Intell.*, vol. 4, no. 12, pp. 1185–1197, Dec. 2022.
[29] R. French, "Catastrophic forgetting in connectionist networks," *Trends Cognit. Sci.*, vol. 3, no. 4, pp. 128–135, Apr. 1999.
[30] X. Ao, X.-Y. Zhang, and C.-L. Liu, "Cross-modal prototype learning for zero-shot handwritten character recognition," *Pattern Recognit.*, vol. 131, Nov. 2022, Art. no. 108859.
[31] X. Yao, X. Wang, Y. Liu, and W. Zhu, "Continual recognition with adaptive memory update," *ACM Trans. Multimedia Comput., Commun., Appl.*, vol. 19, no. 3s, pp. 1–15, Oct. 2023.
[32] L. Wang, X. Yang, H. Tan, X. Bai, and F. Zhou, "Few-shot class-incremental SAR target recognition based on hierarchical embedding and incremental evolutionary network," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5204111.
[33] Y. Cui, W. Deng, H. Chen, and L. Liu, "Uncertainty-aware distillation for semi-supervised few-shot class-incremental learning," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, May 31, 2024, doi: 10.1109/TNNLS.2023.3277018.
[34] M. Kang, J. Park, and B. Han, "Class-incremental learning by knowledge distillation with adaptive feature consolidation," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.* Jun. 2022, pp. 16071–16080.
[35] Y. Liu, X. Hong, X. Tao, S. Dong, J. Shi, and Y. Gong, "Model behavior preserving for class-incremental learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 10, pp. 7529–7540, Oct. 2023.
[36] Z. Ji, Z. Hou, X. Liu, Y. Pang, and X. Li, "Memorizing complementation network for few-shot class-incremental learning," *IEEE Trans. Image Process.*, vol. 32, pp. 937–948, 2023.
[37] S. Tian, W. Li, X. Ning, H. Ran, H. Qin, and P. Tiwari, "Continuous transfer of neural network representational similarity for incremental learning," *Neurocomputing*, vol. 545, Aug. 2023, Art. no. 126300.
[38] T. Zhang, M. Rong, H. Shan, and M. Liu, "Stability analysis of incremental concept tree for concept cognitive learning," *Int. J. Mach. Learn. Cybern.*, vol. 13, no. 1, pp. 11–28, Jan. 2022.
[39] C. Zhang and W. Wang, "Character segmentation for historical uchen Tibetan document based on structure attributes," *Laser. Optoelectron.*, pp. 260–275, 2021.

GUANZHONG ZHAO received the B.S. degree in software engineering from Northwest Minzu University, Lanzhou, China, in 2022, where he is currently pursuing the M.S. degree with the Key Laboratory of China's Ethnic Languages and Information Technology of Ministry of Education. His research interests include pattern recognition, incremental learning, and computer vision.

WEILAN WANG received the B.S. degree in mathematics from Northwest Normal University, Lanzhou, China, in 1983. She was a Visiting Scholar with Sun Yat-sen University, Guangzhou, China, in 1987. From 2001 to 2002, she was a Visiting Scholar with Tsinghua University, Beijing, China. From 2006 to 2007, she was a Visiting Scholar with Indiana University, Bloomington, IN, USA. She is currently a Professor and a Doctoral Supervisor with the Key Laboratory of China's Ethnic Languages and Information Technology of Ministry of Education, Northwest Minzu University, Lanzhou. She has published more than 70 papers in major journals and international conferences in image processing field. Her research interests include image processing, pattern recognition, Tibetan information processing, and computer vision.

XIAOJUAN WANG received the Ph.D. degree from Northwest Minzu University, Lanzhou, Gansu, China. She is currently a Lecturer with the College of Mathematics and Computer Science, Northwest Minzu University. Her research interests include image processing and pattern recognition.

XUN BAO was born in Guang'an, China, in 2000. He received the B.S. degree in computer science and technology from Sichuan University of Science and Engineering, Yibin, China, in 2022. He is currently pursuing the M.S. degree with the Key Laboratory of China's Ethnic Languages and Information Technology of Ministry of Education, Northwest Minzu University, Lanzhou, China. His research interests include pattern recognition, few-shot learning, and contrastive learning.

HUARUI LI was born in Tianshui, China, in 2000. She received the B.S. degree in computer science and technology from Baoji University of Arts and Sciences, Baoji, China, in 2022. She is currently pursuing the M.S. degree with the Key Laboratory of China's Ethnic Languages and Information Technology of Ministry of Education, Northwest Minzu University, Lanzhou, China. Her research interests include pattern recognition and image processing.

MEILING LIU was born in Jiamusi, China, in 2001. She received the B.S. degree in computer science and technology from Jiamusi University, Jiamusi, in 2022. She is currently pursuing the M.S. degree with the Key Laboratory of China's Ethnic Languages and Information Technology of Ministry of Education, Northwest Minzu University, Lanzhou, China. Her research interests include pattern recognition and small-sample detection.