Cross-modal Prototype Learning for Zero-shot Handwriting Recognition

Xiang Ao[1,2], Xu-Yao Zhang[1,2], Hong-Ming Yang[1,2], Fei Yin[1,2], Cheng-Lin Liu[1,2,3]
[1]*National Laboratory of Pattern Recognition, Institute of Automation of Chinese Academy of Sciences, 95 Zhongguancun East Road, Beijing 100190, P.R. China*
[2]*University of Chinese Academy of Sciences, Beijing, P.R. China*
[3]*CAS Center for Excellence in Brain Science and Intelligence Technology, Beijing 100049, P.R. China*
Email: aoxiang2017@ia.ac.cn, {xyz, hongming.yang, fyin, liucl}@nlpr.ia.ac.cn

***Abstract*—In contrast to machine recognizers that rely on training with large handwriting data, humans can recognize handwriting accurately on learning from few samples, and can even generalize to handwritten characters from printed samples. Simulating this ability in machine recognition is important to alleviate the burden of labeling large handwriting data, especially for large category set as in Chinese text. In this paper, inspired by human learning, we propose a cross-modal prototype learning (CMPL) method for zero-shot online handwritten character recognition: for unseen categories, handwritten characters can be recognized without learning from handwritten samples, but instead from printed characters. Particularly, the printed characters (one for each class) are embedded into a convolutional neural network (CNN) feature space to obtain prototypes representing each class, while the online handwriting trajectories are embedded with a recurrent neural network (RNN). Via cross-modal joint learning, handwritten characters can be recognized according to the printed prototypes. For unseen categories, handwritten characters can be recognized by only feeding a printed sample per category. Experiments on a benchmark Chinese handwriting database have shown the effectiveness and potential of the proposed method for zero-shot handwriting recognition.**

***Keywords*-printed character; handwritten character; cross-modal; prototype learning; zero-shot**

I. Introduction

Handwritten Chinese character recognition (HCCR) has been studied for more than fifty years [1] and is widely used to evaluate different pattern recognition techniques. The accuracies on both online and offline HCCR have been constantly improved in recent years [2]. The first reason comes from improved architectures and training strategies in deep neural networks. Another important reason is due to the big data collection for covering all character classes and different handwriting styles. Training with big data is essential for the success of deep learning. However, this is significantly different from the human learning process which relies only on small and even incomplete data. Humans are good at recognizing handwritten characters by reading text books containing only regular and fixed-shape printed characters. Even for characters that we never saw them in handwritten format, we can still recognize them, due to the pre-learned knowledge in matching printed and handwritten characters. This is actually a kind of learning-to-learn or meta-learning ability in our human brain for generalizing to new situations that never shown in training stage.

Under special circumstances, there exists no data but semantic descriptions in some classes, which is known as zero-shot learning [3]. In this problem, the main challenge is that the model should be generalized to identify novel object categories which are unseen in training [4]. To achieve zero-shot HCCR and motivated by the cooperated learning between printed and handwritten characters in human brain, we use two modalities in a joint learning process: the printed character image and the online handwriting trajectory. These two modalities are hard to fuse at raw data level due to heterogenous data format. Therefore, we propose a new model called cross-modal prototype learning (CMPL) to fuse two modalities in a deep neural network transformed semantic space. Particularly, for each character class, there is a single fixed printed character image, and these images are transformed with a convolutional neural network (CNN) to get class-specific prototypes representing different classes. The online handwriting trajectory of each character sample is transformed by a recurrent neural network (RNN) to extract feature representation. To make joint training, the handwritten samples in RNN transformed space are classified with the nearest prototype rule by defining the prototypes as the printed characters in CNN transformed space. A multi-class classification loss function is then defined to learn these two networks simultaneously.

By learning with both printed and handwritten characters, CMPL can be applied for zero-shot handwriting recognition. In the training stage, the samples of handwriting data are not needed to cover all classes. It is possible to train on only a subset of the character classes, and then generalize to other unseen character classes, because the prototypes are not explicitly learned but implicitly produced by applying the CNN on printed character images. For a 3755-class HCCR problem, we show that by training with only 500 classes, CMPL can achieve near 50% accuracy on all 3755 classes, although more than 85% (3255/3755) of the characters are unseen in training stage. With more classes involved in training, the performance can be further improved: training with 1000, 1500, 2000 classes leads to nearly 75%, 85%, 90% test accuracies on 3755 classes, showing the potential

2379-2140/19/$31.00 ©2019 IEEE
DOI 10.1109/ICDAR.2019.00100

of CMPL for zero-shot handwriting recognition.

Previous studies have shown that the printed information is helpful for handwriting analysis. For example, the adversarial feature learning [5] seeks a feature representation that is indistinguishable between printed and handwritten characters, thus making the learned features to be writer-independent and more close to standard printed characters for improving performance. Similar idea is also used for mathematical expression recognition [6] where printed templates are used to guide the recognition of handwritten expressions in an adversarial manner.

The cross-modal learning between printed and handwritten characters is also closely related to the template matching [7] approaches which are widely used in structural pattern recognition. We find a recent work similar to our approach named deep template matching [8] which adopts a deep Siamese neural network to match offline handwritten characters and printed character templates. However, we consider a much more difficult task of cross-modal learning between printed image and online handwriting trajectory, and the final decision-making is a multi-class nearest prototype classification rather than a binary matching. Besides handwriting recognition, the proposed CMPL can hopefully be extended to other pattern recognition tasks.

II. Related Works

1) Handwritten Chinese Character Recognition: Traditional approaches for handwritten Chinese character recognition (HCCR) usually contain multiple stages [9]. Recently, a dominant trend is using end-to-end learning ability of deep neural networks to learn the feature representations and classifiers simultaneously from raw data. The convolutional neural network is widely used for offline HCCR [2] due to its powerful ability in dealing with image-like data, while the recurrent neural network is shown to be very effective for online HCCR [10] owing to its efficiency and generality in feature extraction for sequential data. Improvements on HCCR are gradually reported like faster and more compact models [11], higher accuracies [5], and so on.

2) Few-shot and Zero-shot Learning: The success of deep learning relies heavily on large-enough training data. Contrarily, humans are good at few-shot and even zero-shot learning. The key insight for few-shot learning [12] is that the categories we have already learned can give us information that helps us to learn new categories with fewer examples. In case of zero-shot learning [3] where there is no training example for novel categories, some side information like textual description or attribute definitions is needed to transfer knowledge from known category to unseen category. In this paper, we use the printed character images as an efficient and effective side information for zero-shot HCCR.

3) Multi-modal Learning: Learning from multiple related modalities is an important direction in machine learning. One perspective on multi-modal learning is the fusion of multiple modalities [13] for better decision-making. Another important trend is the cross-modal learning [14] for capturing the relationship between different modalities such as cross-modal translation (like text-to-speech generation), cross-modal alignment (like attention mechanism in image captioning), cross-modal retrieval (like using texts to search images), and so on. Our work is also a kind of cross-modal learning by using one modality as prototype to classify another modality.

4) Open-set Recognition: Most pattern classifiers are based on the closed-set assumption, i.e., there is a predefined fixed number of classes, for example, the widely-used softmax in deep learning. To solve this problem, the open-set recognition [15] is widely-studied in literature with many representative methods like sparse representation based open set recognition [16], probability open-set models [17], open-set deep neural networks [18], and so on. Our proposed method is also a kind of open-set recognition, because the prototypes used for defining different classes can be added incrementally during the test stage to enlarge the category set.

5) Prototype Learning: Nearest neighbor model is a well-known and widely-used classifier for solving pattern recognition tasks, by searching nearest training samples for decision-making. A further improvement is the nearest prototype classifier also known as learning vector quantization [19] which avoids saving all training samples but learns and maintains only several prototypes of each class for classification. Prototype learning has been widely investigated and is efficient and effective for handwritten character recognition [20]. Recently, a new trend is to combine the nearest prototype classifier with deep neural networks to improve the robustness [21] and the few-shot learning ability [22]. Our work is also a kind of deep prototype learning with emphasis on cross-modal prototype learning.

III. Cross-modal Prototype Learning

The whole framework of the proposed cross-modal prototype learning (CMPL) is illustrated in Fig. 1, which contains two embedding networks $\varphi(\cdot)$ and $\pi(\cdot)$ for the printed character image and the online handwriting trajectory. These two networks are jointly learned to minimize the loss function of a nearest prototype classifier.

A. Printed Character Embedding

As shown in the top of Fig. 1, for each character class, a single and fixed printed character image with resolution 32×32 is used for calculating the prototype representation for this class. The data preprocessing is implemented by projecting the value of pixel that ranges from 0 to 255 linearly into the interval of $[-0.5, 0.5]$. After that, a convolutional neural network $\varphi(\cdot)$ is used to extract features from these printed images. Considering the modality of printed character image is simple without large variations, we use a

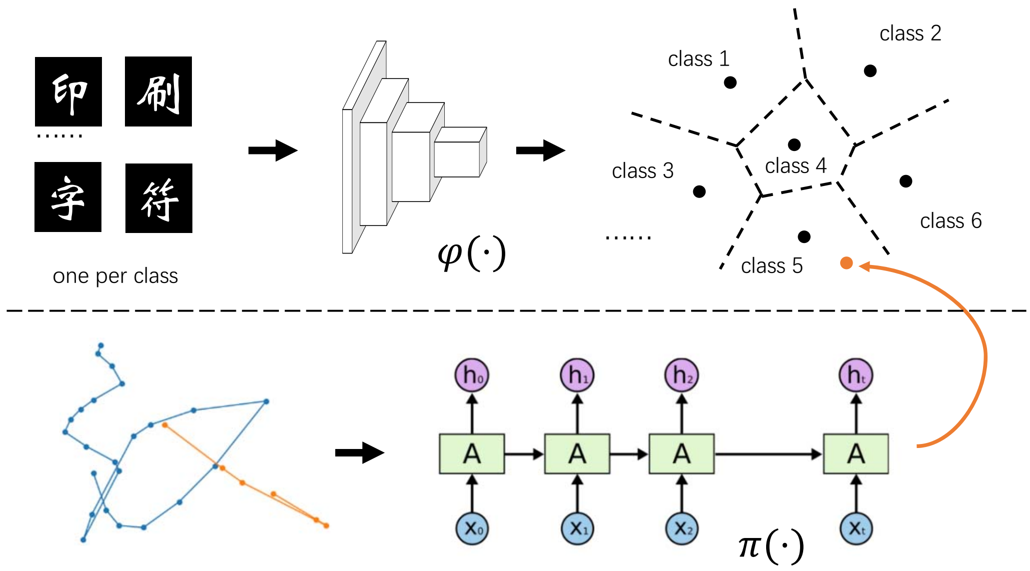

Figure 1. An illustration of cross-modal prototype learning. Top: the modality for printed character images. Bottom: the modality for online handwriting trajectory. These two modalities are learned jointly by classifying the handwritten samples using nearest prototypes from the printed characters. Once trained, it can well generalize to novel character class which has no handwriting sample, by using only a printed character as prototype for this class.

small convolutional neural network as $\varphi(\cdot)$, which contains a stack of 5 convolutional layers with 3×3 kernel. The number of channels in each convolutional layer is enlarged gradually from 50 in layer-1 to 400 in layer-5. The max-pooling layer is also used to reduce the size of feature maps for enlarging the receptive fields. At last, the obtained feature maps are flattened to a vector with dimensionality 400 and then fed into a fully-connected layer to get a compact 200 dimensional feature vector as the final representation. The batch normalization is also applied to accelerate the training process.

Suppose there are C classes, and for each class we have a printed character image. Then the image set can be denoted as $\{I_1, I_2, \ldots, I_C\}$. With the embedding network $\varphi(\cdot)$, the prototype for class i is:

$$\varphi(I_i), i = 1, 2, \ldots, C. \tag{1}$$

Different from previous prototype learning [20], [21] where the prototype is totally learned from data, the prototype here is implicitly learned as a mapping from printed character.

B. Online Handwriting Trajectory Embedding

As shown in the bottom of Fig. 1, each online handwritten character is actually represented as a sequential trajectory and a recurrent neural network (RNN) $\pi(\cdot)$ is used to embed it. We follow the method proposed in [10] by firstly removing redundant points and applying coordinate normalization to the trajectory, for reducing the difference in number of sampling points and the variation in size of the coordinates for different characters. After the preprocessing, a deep bidirectional RNN is used to integrate the information from both the past and the future of the trajectory. As suggested by [10], the gated recurrent unit (GRU) is used in the RNN other than the long-short term memory (LSTM). Two recurrent layers are used with 100 neurons in layer-1 and 500 neurons in layer-2. After that, all the hidden states in layer-2 of the RNN are collected and passed through a mean pooling layer to output a fixed-length feature vector with dimensionality 500, which is further fed into a fully-connected layer to produce a 200 dimensional vector as the final representation. The batch normalization is applied after the fully-connected layer to speedup convergence.

C. Cross-modal Nearest Prototype Classifier

With the above two embedding networks $\varphi(\cdot)$ and $\pi(\cdot)$, both the printed character image and online handwriting trajectory are mapped into the same dimensional space. We use x to represent a general online handwritten character and $y \in \{1, 2, \ldots, C\}$ to denote its label. The embedding of this character is $\pi(x)$, and we can therefore define the distance of this sample to each class as the Euclidean distance between $\pi(x)$ and class-specific prototype:

$$d(x, i) = \|\pi(x) - \varphi(I_i)\|_2^2, i = 1, 2, \ldots, C. \tag{2}$$

We can further transform the distance to probability as:

$$P(i|x) = \frac{e^{-\beta d(x,i)}}{\sum_{j=1}^{C} e^{-\beta d(x,j)}}, \tag{3}$$

where β is a hyper-parameter controlling the hardness of probability assignment. With this, the cross entropy loss can be defined as:

$$\text{Loss}(x, y) = -\log P(y|x). \tag{4}$$

Prototypes used to define the loss will be all updated during optimization, but with an emphasis on the genuine prototype and the competitive prototype (nearest prototype from false classes) as shown in Fig. 1. However, directly optimizing the cross entropy loss will usually leads to overfitting, therefore, we also apply a regularization [21]:

$$\text{Reg}(x, y) = \left\|\pi(x) - \varphi(I_y)\right\|_2^2. \tag{5}$$

This is to constrain each handwritten sample to be close to the printed prototype from the same class, for making the features of samples within the same class more compact. The whole CMPL is optimized as

$$\min_{\varphi,\pi} \sum_{(x,y)} \left\{\text{Loss}(x, y) + \lambda \text{Reg}(x, y)\right\}, \tag{6}$$

where λ is a hyper-parameter to balance the loss and regularization.

D. Zero-shot Handwriting Recognition

Although we call our model cross-modal prototype learning, actually, no explicit prototype is learned, and only two embedding networks $\varphi(\cdot)$ and $\pi(\cdot)$ are optimized as shown in Eq. (6). In testing stage, the decision can be made as:

$$x \in \arg\min_{I \in \mathcal{P}} \left\|\pi(x) - \varphi(I)\right\|_2^2, \tag{7}$$

where $\mathcal{P}$ is a gather of printed character images, which is no longer restricted to $\{I_1, I_2, \ldots, I_C\}$ used in training. In other words, we can actually add more printed character images into $\mathcal{P}$ (or delete some images from it) without need to re-train the model. This is a very important property, because in some cases, there is no handwritten data for some particular characters, and it is also hard and expensive to collect large number of samples covering all categories. However, the printed character images are easy to obtain. Therefore, this kind of cross-modal class incremental learning is a good solution for zero-shot handwriting recognition.

IV. Experiments

A. Datasets

We use a benchmark online handwritten Chinese character recognition dataset for evaluating the effectiveness of our method. The database used for training is the CASIA database [23] including OLHWDB1.0 and OLHWDB1.1. The database used for testing is from the ICDAR-2013 competition [24] of online Chinese handwriting recognition. There are totally 2,693,183 samples for training and 224,590 samples for testing. The number of character class is 3,755 (level-1 set of GB2312-80). For printed character images, we use 3755 images (one per class) generated by font of Microsoft XinWei, where each contains a standard white character in the middle of image with black background.

B. Implementation Details

During training, the sequential dropout [10] is applied by randomly removing each straight line of the trajectory with probability 0.3. The optimization algorithm of Adam is used in our experiment. In each step, a mini-batch of 1000 online handwritten samples and all the printed character images from seen classes are fed to the RNN and CNN respectively. The initial learning rate of RNN is same as that of CNN, which is set to be 0.001, and it drops by $\times 0.3$ when the accuracy on the training set stops improving. We implemented experiments under the framework of Tensorflow using 4 NVIDIA Titan X 12G GPUs.

To evaluate the performance of CMPL, we only use part of all the 3755 classes in training, and leave other unseen classes for evaluating the performance of zero-shot learning. In our experiments, the character classes are sorted in the order of GB2312-80 level-1 set, and we simply select the first N classes as *seen* and the remaining 3755-N classes as *unseen*. Different numbers of N are evaluated in our experiments: 500, 1000, 1500, 2000, 2500.

For the convenience of following description, we use "A/B" here to indicate the situation where test classes are "A" and chosen prototypes are "B". For example, when $N = 500$, "Seen/All" means 500 classes tested with all 3755 prototypes while "Unseen/Unseen" refers to the remaining 3255 classes tested with 3255 prototypes.

C. Closed-set Performance

We first evaluate the performance on the selected N seen classes. Since they are used in both training and testing, this is the traditional closed-set recognition. The results are shown in the "Seen/Seen" column of Table I. It is shown that, all performances are very high (more than 98%), indicating that using the printed character modality as prototypes will not reduce the performance of closed-set recognition, although more constraint is actually adopted on the prototypes. Another important result is shown on the "Seen/All" column which means: although only N classes are used in training, we evaluate them with 3755 prototypes from all classes. It is shown that: even training with only 500 classes, the "Seen/All" accuracy is still more than 96%. This demonstrates that: the prototypes in seen classes can be effectively separated from newly added unseen prototypes and the distributions of handwriting data in seen classes are intra-class compact.

D. Open-set Performance

A main advantage of the proposed CMPL is that it can be extended to recognize new unseen classes. For the 3755-N classes (unseen in training), the "Unseen/Unseen" column in Table I means evaluating them with prototypes only from unseen classes, while "Unseen/All" means evaluating with prototypes from all classes. First of all, the performance on unseen classes are much lower than the seen classes,

Table I
ZERO SHOT LEARNING PERFORMANCE OF THE PROPOSED METHOD WITH DIFFERENT NUMBER OF TRAINING CLASSES.

Prototypes		Seen	Unseen	All				Common
Testing Classes		Seen	Unseen	Seen	Unseen	All	Common	Common
N Training Classes	500	0.9902	0.4468	0.9633	0.4159	0.4887	0.4167	0.5862
	1000	0.9885	0.7101	0.9671	0.6699	0.7489	0.6668	0.7938
	1500	0.9852	0.8049	0.9671	0.7553	0.8399	0.7520	0.8528
	2000	0.9844	0.8673	0.9695	0.8112	0.8955	0.8067	0.8888
	2500	0.9822	0.9074	0.9721	0.8407	0.9282	0.8407	0.9077

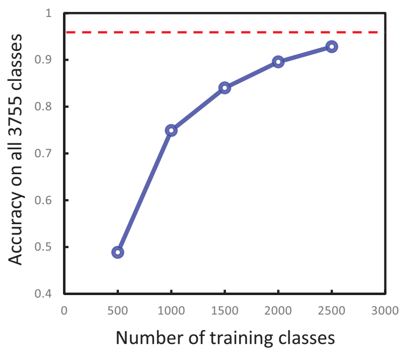

Figure 2. The accuracies on all 3755 classes.

indicating that zero-shot learning of new categories is a difficult task due to the lack of real handwriting samples. However, when N is increased (more classes are used in training), the performance on unseen classes is improved significantly, and the gap between the accuracies on seen and unseen classes is gradually reduced. Moreover, the difference between "Unseen/Unseen" and "Unseen/All" is not significant, indicating that the confusion of unseen classes is from themselves rather than the seen classes.

We also give the performance of "All/All" in Table I which indicates evaluating all classes w.r.t. all prototypes, and its changing trend w.r.t. N in Fig. 2. Note that the upper bound of CMPL should be the performance trained with all the 3755 classes which is 97.76% as reported by [10]. As shown by Fig. 2, with more classes used for training, the performance is gradually increased to approach the upper bound, although much less classes are used in training compared with 3755 classes, indicating the effectiveness of CMPL for zero-shot learning. At last, we also show the performance of "Common/Common" and "Common/All" on the common (last) 1255 unseen classes no matter $N = 500, 1000, 1500, 2000, 2500$. This is to fairly compare the performance on unseen classes w.r.t. N by using the same testing set. The performance trend on "Common" is coincide with "Unseen" which again verifies our discussions.

E. Error Analysis

To know how wrong predictions are produced by CMPL in open-set recognition, we select and exhibit a few common failure samples appeared in two situations ("Unseen/All" and "Unseen/Unseen"). As shown in Fig. 3, although being wrongly classified, in most case, the correct labels can be found in the top-3 predictions, and the mistakes are usually caused by the confusion between similar characters with subtle difference in shape. Evaluation under "Unseen/All" will cause more mistakes since more prototypes (classes) are used, however, the difference is not significant compared with "Unseen/Unseen".

F. Transfer Learning from Seen to Unseen Characters

In CMPL, we have two modules: CNN embedding for printed character and RNN embedding for online trajectory, which are trained on a subset of the categories. Once trained on the first 2500 classes, they can be extended to the unseen 1255 classes as discussed above. However, if we fine-tune the networks on these 1255 classes, the performance would be further improved. In this subsection we consider training the RNN+Prototype network (similar to [21]) for the 1255 unseen classes in two situations: totally from scratch, and with knowledge transferred from CMPL by copying the RNN and CNN-embedded prototypes (from unseen printed characters) as initialization. We also use a small percentage of samples to evaluate training with small sample sizes. The results are listed in Table. II. It is demonstrated that the knowledge learned from seen classes in CMPL is helpful and transferable to unseen classes, which can consistently and significantly improve performance under different sampling rates.

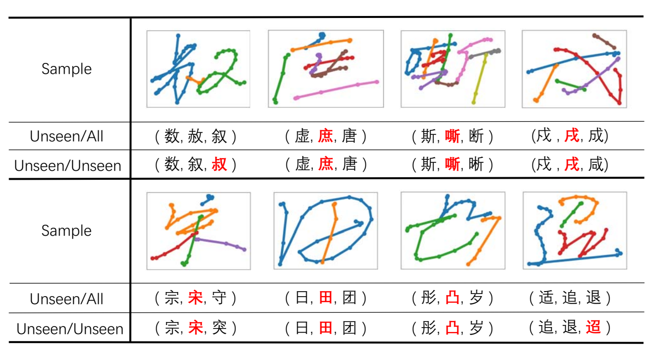

Figure 3. Top-3 predictions where the correct one is in red color.

Table II
PERFORMANCE ON THE UNSEEN 1255 CLASSES.

Knowledge Transfer?	Percentage of training samples			
	10%	20%	30%	100%
yes	0.9809	0.9829	0.9847	0.9872
no	0.9685	0.9768	0.9801	0.9857

V. CONCLUSION AND FUTURE WORK

In this paper, we propose a new method named cross-modal prototype learning (CMPL) for zero-shot handwriting recognition. The main idea is using printed character as prototype to classify handwriting data with modality-specific embedding networks. Experimental results have verified the effectiveness of the proposed method in generalizing to novel unseen classes. However, the accuracies on unseen classes still have large distances compared with the accuracies on seen classes, indicating that more efforts could be paid on finding effective training strategies to improve open space generalization, since the regularization now is only defined on seen classes. Moreover, besides printed character and online trajectory, our future work will also take offline handwritten character into consideration for joint learning with three modalities to further improve performance.

ACKNOWLEDGMENT

This work is supported by the National Natural Science Foundation of China (NSFC) Grants 61721004, 61633021, 61836014, the Beijing Science and Technology Program Grant Z181100008918010.

REFERENCES

[1] F. Kimura, K. Takashina, S. Tsuruoka, and Y. Miyake, "Modified quadratic discriminant functions and the application to Chinese character recognition," *IEEE Trans. Pattern Analysis and Machine Intelligence*, no. 1, pp. 149–153, 1987.

[2] X.-Y. Zhang, Y. Bengio, and C.-L. Liu, "Online and offline handwritten Chinese character recognition: A comprehensive study and new benchmark," *Pattern Recognition*, vol. 61, pp. 348–360, 2017.

[3] C. Lampert, H. Nickisch, and S. Harmeling, "Attribute-based classification for zero-shot visual object categorization," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 36, no. 3, pp. 453–465, 2014.

[4] Y. Fu, T. Xiang, Y.-G. Jiang, X. Xue, L. Sigal, and S. Gong, "Recent advances in zero-shot recognition," *arXiv:1710.04837*, 2017.

[5] Y. Zhang, S. Liang, S. Nie, W. Liu, and S. Peng, "Robust offline handwritten character recognition through exploring writer-independent features under the guidance of printed data," *Pattern Recognition Letters*, vol. 106, pp. 20–26, 2018.

[6] J.-W. Wu, F. Yin, Y.-M. Zhang, X.-Y. Zhang, and C.-L. Liu, "Image-to-markup generation via paired adversarial learning," in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, 2018, pp. 18–34.

[7] X.-H. Xiao and R.-W. Dai, "On-line handwritten Chinese character recognition directed by components with dynamic templates," in *Int'l Conf. Computer Processing of Oriental Languages*, 1997, pp. 89–94.

[8] Z. Li, M. Jin, Q. Wu, and H. Lu, "Deep template matching for offline handwritten Chinese character recognition," *arXiv:1811.06347*, 2018.

[9] R. Dai, C. Liu, and B. Xiao, "Chinese character recognition: history, status and prospects," *Frontiers of Computer Science in China*, vol. 1, no. 2, pp. 126–136, 2007.

[10] X.-Y. Zhang, F. Yin, Y.-M. Zhang, C.-L. Liu, and Y. Bengio, "Drawing and recognizing Chinese characters with recurrent neural network," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 40, no. 4, pp. 849–862, 2018.

[11] X. Xiao, L. Jin, Y. Yang, W. Yang, J. Sun, and T. Chang, "Building fast and compact convolutional neural networks for offline handwritten Chinese character recognition," *Pattern Recognition*, vol. 72, pp. 72–81, 2017.

[12] L. Fei-Fei, R. Fergus, and P. Perona, "One-shot learning of object categories," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 28, no. 4, pp. 594–611, 2006.

[13] D. Ramachandram and G. Taylor, "Deep multimodal learning: A survey on recent advances and trends," *IEEE Signal Processing Magazine*, vol. 34, no. 6, pp. 96–108, 2017.

[14] T. Baltrušaitis, C. Ahuja, and L.-P. Morency, "Multimodal machine learning: A survey and taxonomy," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 41, no. 2, pp. 423–443, 2019.

[15] W. Scheirer, A. Rocha, A. Sapkota, and T. Boult, "Toward open set recognition," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 35, no. 7, pp. 1757–1772, 2013.

[16] H. Zhang and V. Patel, "Sparse representation-based open set recognition," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 39, no. 8, pp. 1690–1696, 2017.

[17] W. Scheirer, L. Jain, and T. Boult, "Probability models for open set recognition," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 36, no. 11, pp. 2317–2324, 2014.

[18] A. Bendale and T. Boult, "Towards open set deep networks," in *IEEE Conf. Computer Vision and Pattern Recognition*, 2016, pp. 1563–1572.

[19] A. Sato and K. Yamada, "Generalized learning vector quantization," in *Advances in Neural Information Processing Systems*, 1996, pp. 423–429.

[20] C.-L. Liu and M. Nakagawa, "Evaluation of prototype learning algorithms for nearest-neighbor classifier in application to handwritten character recognition," *Pattern Recognition*, vol. 34, no. 3, pp. 601–615, 2001.

[21] H.-M. Yang, X.-Y. Zhang, F. Yin, and C.-L. Liu, "Robust classification with convolutional prototype learning," in *IEEE Conf. Computer Vision and Pattern Recognition*, 2018, pp. 3474–3482.

[22] J. Snell, K. Swersky, and R. Zemel, "Prototypical networks for few-shot learning," in *Advances in Neural Information Processing Systems*, 2017.

[23] C.-L. Liu, F. Yin, D.-H. Wang, and Q.-F. Wang, "CASIA online and offline Chinese handwriting databases," *Proc. Int'l Conf. Document Analysis and Recognition (ICDAR)*, 2011.

[24] F. Yin, Q.-F. Wang, X.-Y. Zhang, and C.-L. Liu, "ICDAR 2013 Chinese handwriting recognition competition," *Proc. Int'l Conf. Document Analysis and Recognition (ICDAR)*, 2013.