Information Sciences 482 (2019) 156–170

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

Semi-supervised feature learning for improving writer identification

Shiming Chen [a], Yisong Wang [a,b,*], Chin-Teng Lin [d], Weiping Ding [c], Zehong Cao [d,e,*]

[a] *School of Computer Science and Technology, Guizhou University, Guizhou, China*
[b] *Key Laboratory of Intelligent Medical Image Analysis and Precise Diagnosis of Guizhou Province, Guizhou, China*
[c] *School of Computer Science and Technology, Nantong University, Jiangsu, China*
[d] *Centre for Artificial Intelligence, Faculty of Engineering and IT, University of Technology Sydney, NSW, Australia*
[e] *Discipline of ICT, School of Technology, Environments and Design, University of Tasmania, Hobart, Australia*

A R T I C L E I N F O

Article history:
Received 6 October 2018
Revised 7 January 2019
Accepted 9 January 2019
Available online 10 January 2019

Keywords:
Semi-supervised feature learning
Feature extraction
Regularization
CNN
Writer identification.

A B S T R A C T

Data augmentation is typically used by supervised feature learning approaches for offline writer identification, but such approaches require a mass of additional training data and potentially lead to overfitting errors. In this study, a semi-supervised feature learning pipeline is proposed to improve the performance of writer identification by training with extra unlabeled data and the original labeled data simultaneously. Specifically, we propose a weighted label smoothing regularization (WLSR) method for data augmentation, which assigns a weighted uniform label distribution to the extra unlabeled data. The WLSR method regularizes the convolutional neural network (CNN) baseline to allow more discriminative features to be learned to represent the properties of different writing styles. The experimental results on well-known benchmark datasets (ICDAR2013 and CVL) showed that our proposed semi-supervised feature learning approach significantly improves the baseline measurement and perform competitively with existing writer identification approaches. Our findings provide new insights into offline writer identification.

1. Introduction

Handwritten texts, speech, fingerprints, and faces are often applied in physiological biometric identifiers. Handwritten text plays an especially important role for forensics and security in proving a person's authenticity. Research into writer identification, such as historical document analysis for the mass-digitization process of historical documents [24,29,47] through machine learning, has received renewed interest in recent years; unfortunately, this process requires considerable time and detection costs. Therefore, many researchers have proposed state-of-the-art pattern recognition approaches to automatically recognize writing styles [1,7,11,30,43].

The aim of writer identification is to search and recognize texts written by the same writer in a query database. Writer identification has been investigated for different handwritten scripts, such as English [4,39], Chinese [18,19,46], Arabic [1], Indic [30], Persian [20] and Latin scripts [9]. This task generally presents substantial challenges because it requires the documents to be sorted according to high similarity (e.g., the distance between feature vectors). Writer identification can

* Corresponding author.
E-mail addresses: gchenshiming@gmail.com (S. Chen), yswang@gzu.edu.cn (Y. Wang), zhcaoctu@gmail.com (Z. Cao).

https://doi.org/10.1016/j.ins.2019.01.024
0020-0255/

be classified as online writer identification and offline writer identification according to the handwritten document acquisition method. The latter approach can be further categorized into allograph-based and textual-based methods. Textural-based methods compute global statistics directly from handwritten documents (pages) [3,14,32]; for example, the angles of stroke directions, the width of the ink trace, and the histograms of local binary patterns (LBP) and local ternary patterns (LTP) have been used for writer identification purposes. Allograph-based methods rely on local descriptors computed from small patches (allographs), and a global document descriptor is statistically calculated using the local descriptors of a single document [7,8,18]. These two methods can be further combined to form a discriminative global feature [4,17,46]. The semi-supervised feature learning pipeline for offline writer identification proposed in this work is based on allographs.

Although writer identification has achieved excellent performance on some benchmark datasets, there are considerable challenges in real-world applications. First, the use of different pens, the physical condition of the writer, the presence of distractions (such as multitasking and noise), and the changes in writing style with age are key factors resulting in the unsatisfactory performance of writer identification. Second, the writers of the training set are different than those of the test set, and every writer only contributes a few handwritten text images in the typically used benchmark datasets. Third, the number of handwritten documents in benchmark datasets is highly insufficient for convolutional neural network (CNN) model training; therefore, training a reliable CNN model using limited data is a challenge. Moreover, almost all published methods are based on supervised learning, which cannot achieve landmark results due to the limited amount of labeled data present in the benchmarks. Some researchers utilize different data augmentation methods to address these problems. However, these data augmentation methods that are used in writer identification easily lead to model overfitting and require a considerable amount of extra data. To overcome the aforementioned challenges and then tightly integrate with writer identification in practice, we propose a novel insight for writer identification.

CNNs are a well-known deep learning architecture inspired by the natural visual perception mechanism of living creatures. CNNs have been widely used and have achieved superior performance in the fields of image classification, object recognition and object detection and tracking [16,41] due to their powerful ability to learn deep features. The recent progress in writer identification is attributed mainly to advancements in CNNs based on supervised [6–8,11,17,43,47] and unsupervised feature learning [9]. The features extracted from CNNs perform better than handcrafted features as discriminative characteristics. For example, Xing and Qiao [47] designed a multistream CNN structure for writer identification and achieved a high identification accuracy on the IAM [31] and HWDB [27] datasets using a small amount of handwritten documents. In [8], Christlein proposed using activation features from CNNs as local descriptors for writer identification and improved the identification performance on the ICDAR2013 dataset. Eldan and Shamir [10] showed that a deeper network would learn a more discriminative representation but will need more resources to train. Therefore, we recommend that a tradeoff and a deep residual neural network with 50 layers (ResNet-50) could be applied in our work.

Semi-supervised feature learning significantly outperforms supervised feature learning when annotated data are limited in the training set, e.g., weakly labeled or unlabeled data are available [21,45]. Specifically, semi-supervised feature learning saves time and reduces the cost needed for annotating data when the volume of clean labeled data is limited. Some recent studies investigated a semi-supervised feature learning pipeline by combining unsupervised feature learning with supervised feature learning [38,44] to assign an original or new label to unlabeled data [25,34]. Motivated by these previous studies, we attempt to use a modified semi-supervised feature learning method by assigning a weighted uniform label distribution to extra unlabeled data (extra data) according to the original labeled data (real data). We believe that the proposed approach has the potential to regularize the baseline for improving identification performance.

Therefore, we proposed a semi-supervised feature learning method that leverages a deep CNN and weighted label smoothing regularization (WLSR) to construct a powerful model that learns discriminative representations for offline writer identification. Specifically, we first preprocess the original labeled data and the extra unlabeled data. Then, these original labeled data and extra unlabeled data are fed into a deep residual neural network (ResNet) [16] simultaneously. Furthermore, the WLSR method regularizes the learning process by integrating the unlabeled data, which can reduce the risk of overfitting and direct the model to learn more effective and discriminative features. Finally, the local features of every test handwritten document are extracted and encoded as a global feature vector for identification.

To summarize, this study makes the following contributions:

A. This study is a pioneering work that uses a semi-supervised feature learning pipeline to integrate extra unlabeled images and original labeled images into the ResNet model for writer identification.

B. The WLSR method of semi-supervised feature learning is used to regularize the identification model with unlabeled data. We thoroughly evaluate its availability on public datasets.

C. Our results show that the proposed semi-supervised feature learning model shows consistent improvement over the deep residual neural network baseline and achieves better performance than that of existing approaches on benchmark datasets.

The remainder of this paper is organized as follows. Section 2 provides an overview of the related works in the field of writer identification. The semi-supervised feature learning pipeline is presented in Section 3. The performance and evaluation are given in Section 4. Section 5 presents the discussion. Section 6 provides a summary and the outlook for future research.

2. Related work

Data augmentation [6–8,11,43,47] and semi-supervised learning [2,36,37] are widely used for classification and identification when the amount of annotated data in a training set is limited. In this section, we review related work on writer identification that implemented different data augmentation approaches or semi-supervised learning methods to address cutting-edge challenges.

2.1. Data augmentation

In this section, we review related work on writer identification that considered different data augmentation approaches to address cutting-edge challenges. Some researchers implemented data augmentation in intrasets [8,11,43,47], but this method easily led to model overfitting. Two recent studies added extra labeled data to the original data to enlarge the training set; however, this method required a vast amount of additional data to improve the identification results [6,7].

Fiel and Sablatnig [11] used a series of image preprocessing methods (binarization, text line segmentation, and sliding window) and generated a discriminative feature via CaffeNet for each 56×56 image patch. Because CNNs have to be trained on a large amount of data to achieve good results, he cut line images into patches using a sliding window model with a step size of 20 pixels and rotated each patch of the sliding window from -25 to $+25$ degrees using a step size of 5 degrees. Thus, the new training set consisted of more than 2,300,000 image patches, which artificially enlarged the original training set. His proposed algorithm achieved good performance on the ICDAR2011 [12] and CVL [24] datasets, but this algorithm failed to improve the performance on the ICDAR2013 [29] dataset. Furthermore, the CNN was trained on word images of the IAM dataset and the features of the CVL dataset extracted from the pretrained CNN. It suggested that the IAM and CVL datasets share a similar sample space. In [43], Tang introduced a new method for offline writer identification using a CNN and a joint Bayesian approach to contend with insufficient benchmark datasets for CNN model training. Tang also used words segmented from handwritten documents as elements to permute the texts to generate a significant number of images, which were subsequently converted to form handwritten pages. In addition, all the reconstructed handwritten pages were split into some nonoverlapping patches for training. In [47], Xing introduced a data augmentation method to enhance the performance of the proposed DeepWriter. However, these data augmentation methods only enlarged the dataset in the area of the intraset, and existing models did not consider dealing with the generated data, leading to an overfitting situation and limitations of feature learning in CNNs.

In [6], Christlein created a combined dataset (MERGED) consisting of 559 scribes with four documents per writer, resulting in 2236 documents from the ICDAR2013 and CVL datasets. Thereby, the training set was enlarged, and the outcomes on the MERGED datasets slightly differ from the image vocabularies that can be calculated from the ICDAR2013 experimental set or the CVL dataset. Furthermore, Christlein et al. [7] showed that the identification rate on the CVL test set could be improved by adding additional datasets (ICDAR2011 and IAM [31]) into the CVL training set. Although existing data augmentation approaches have the capability to improve the identification performance using the extra data, we can imagine that it requires a large amount of extra labeled data. In practice, however, we do not have access to collect a large number of samples for writer identification.

In contrast to the aforementioned works, we employed a semi-supervised feature learning pipeline that allows the addition of data without labels. We assumed that the semi-supervised feature learning approach could effectively avoid overfitting and require less additional data to improve the feature learning ability of the baseline.

2.2. Semi-supervised learning

Semi-supervised learning significantly outperform supervised learning for writer identification in small datasets. In [36,37], Porwal proposed a structural-correspondence-learning-based semi-supervised learning approach for writer identification. This semi-supervised learning method improved the performance of a writer classifier with two distinct handcrafted features (gradient, structural and concavity features (GSC), and contour angle features). The method first generates the GSC and contour angle features for the labeled data and the contour angle features for the unlabeled data. The GSC features are then used for training a support vector machine (SVM) for auxiliary label generation of the unlabeled data features, and all contour angle features are used for training the classifiers of all subtasks. In fact, the semi-supervised learning in Porwal's work makes direct use of the handcrafted features and requires that the number of unlabeled data is same as the that of the testing data. Additionally, Porwalâs method is desirable in that the auxiliary tasks are related to each other, so a common structure can be retrieved.

3. Semi-supervised feature learning pipeline

As shown in Fig. 1, our proposed semi-supervised feature learning pipeline consists of three parts. **A**. Preprocessing: For the ICDAR2013 dataset, the handwritten documents are segmented into line images by a line segmentation method [40], and then the line images are split up using a sliding window approach without overlapping. For the IAM and CVL datasets, we normalize the word images already provided. **B**. Semi-supervised feature learning: During training, the original labeled data (real data) and extra unlabeled data (extra data) are shuffled and simultaneously fed into the ResNet-50 baseline, which is

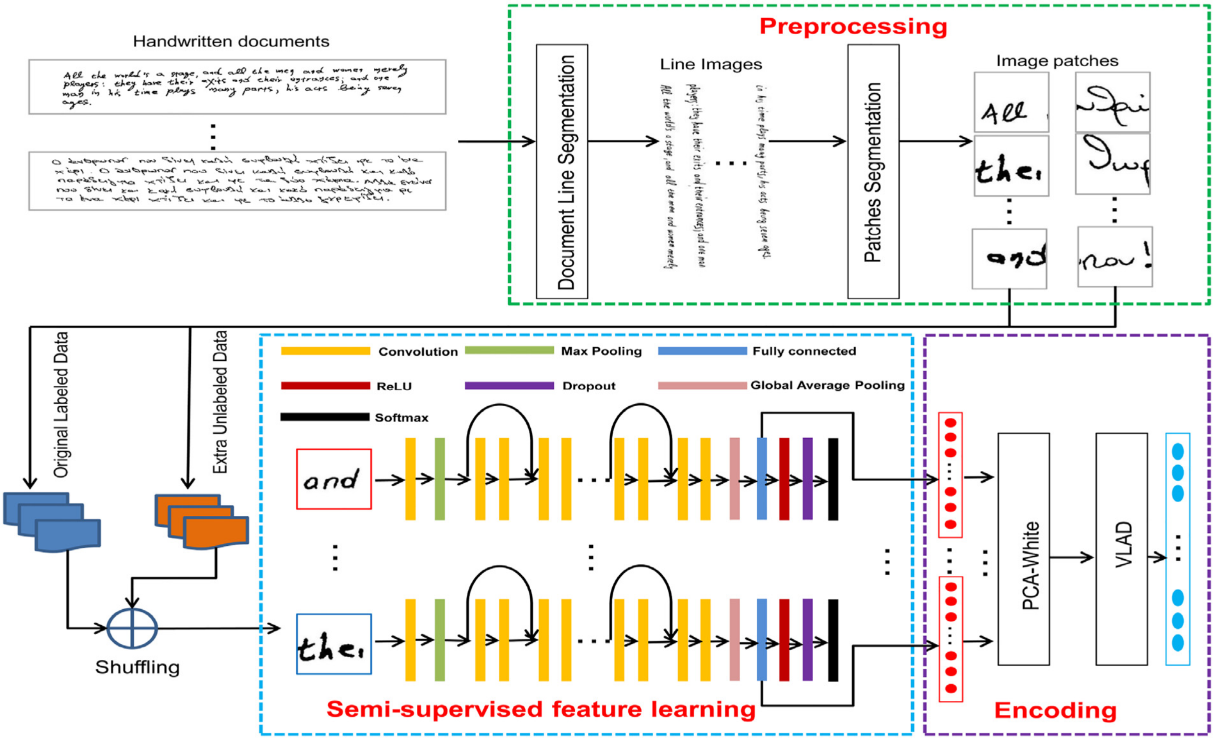


Fig. 1. The pipeline of semi-supervised feature learning, which consists of three parts: preprocessing (green dotted box), semi-supervised feature learning (blue dotted box) and encoding (purple dotted box). During training, the original labeled data and extra unlabeled data are shuffled and fed into the semi-supervised feature learning network. For testing, the local features (red rectangles with solid edge in encoding part) of testing handwritten documents are extracted from the fully connected layer of the pretrained model, and then all the local features of one handwritten test document are encoded into a global feature vector (blue rectangles with solid edge in encoding part). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

regularized by WLSR. Furthermore, the trained model is used for extracting local features of testing handwritten documents. Specifically, all local features of handwritten test documents are extracted from the fully connected layer, and thus, all layers after the fully connected layer can be discarded. **C**. Encoding: We reduce the dimensions of local features with PCA-White [23], and then the vector of locally aggregated descriptors (VLAD) [22] is used to encode the local features of every test document as a global feature vector, which is used for writer identification with the nearest neighbor approach. All of the parts will be concretely introduced in the following.

3.1. Preprocessing

First, a binarization is implemented for all handwritten pages with the Otsu [33] method. Second, the handwritten pages have to be segmented. Because the CVL dataset [24] and IAM [31] dataset already provide a segmentation of the words, these images are directly used for training and evaluating after normalization, as shown in Fig. 3. For the ICADR2013 competition on the Writer Identification dataset [29], the handwritten documents are segmented into lines with the method proposed by Srinivasan and Srihari [40]. The line segmentation method is based on a statistical approach that segments the text lines exactly. In addition, we normalize the line images to a height of 256 pixels and maintain their aspect ratio. Finally, all text lines are cut into patches with a size of 256×256 without overlap using the sliding window approach. Some line images and patches of the ICDAR2013 dataset are shown in Figs. 2 and 3, respectively. Furthermore, we remove noise patches (e.g., blank patches) to avoid adverse effects.

3.2. Semi-supervised feature learning

In this section, we thoroughly introduce the process of the proposed semi-supervised feature learning. Semi-supervised feature learning is based on a baseline (ResNet-50) and WLSR method. The baseline serves as an identification model, and the local features of testing handwritten pages are extracted from the fully connected layer of the baseline during testing. WLSR regularizes the baseline and directs the model to learn more discriminative features.

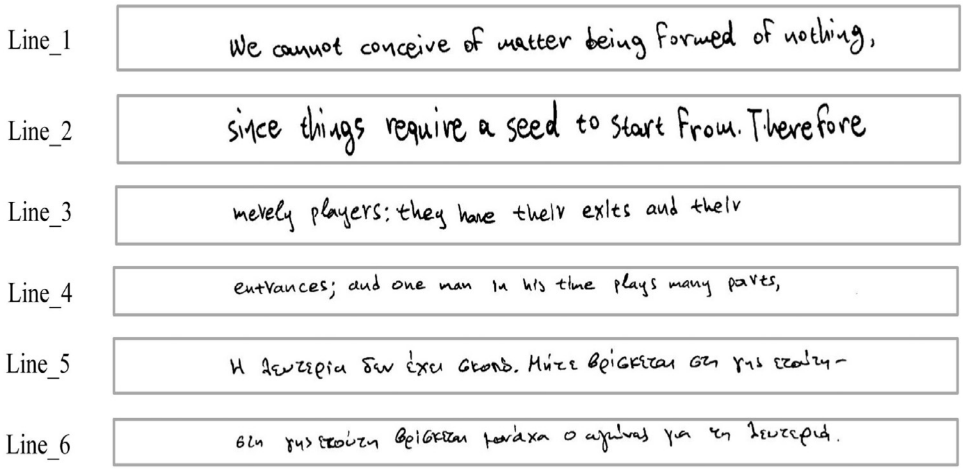


Fig. 2. Part of the line images of the ICDAR2013 dataset are segmented by the proposed line segmentation approach and are normalized with their original aspect ratio.

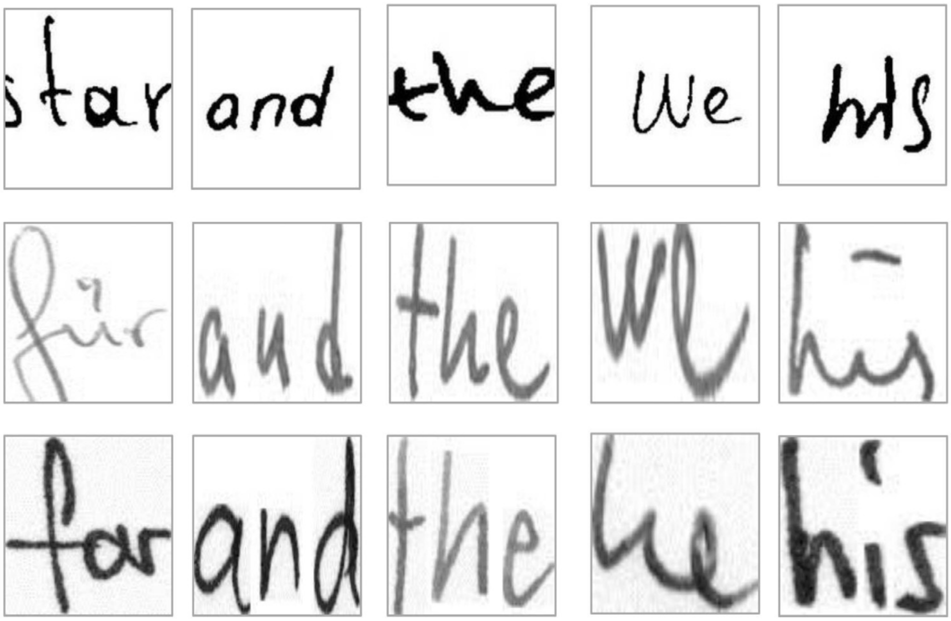
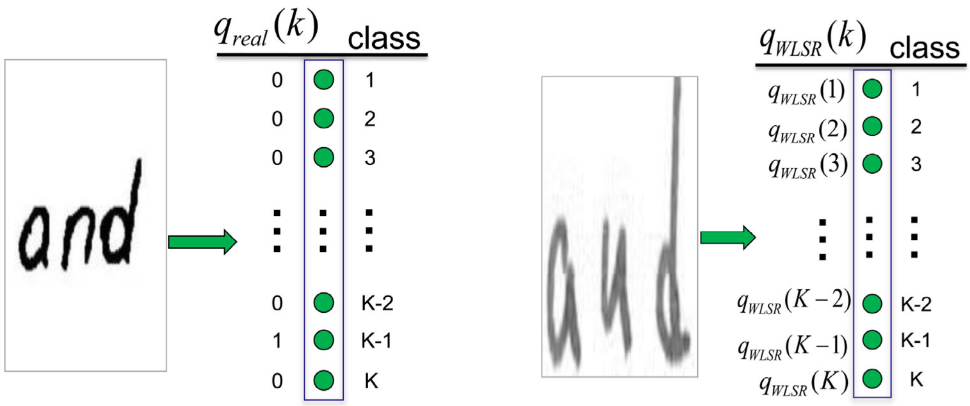


Fig. 3. Part of the patches extracted from the ICDAR2013 dataset (top row), word images provided by the CVL dataset (middle row) and word images provided by the IAM dataset (bottom row), where all have been preprocessed. The patches of the ICDAR2013 dataset are normalized to 256×256. Each word image with size $x \times y$ in the CVL and IAM datasets is normalized to an image of size $256 \times m$ or $m \times 256$ such that $\frac{x}{y} = \frac{256}{m}$ or $\frac{x}{y} = \frac{m}{256}$.

3.2.1. CNN baseline

He et al. [16] first proposed ResNet for image classification and object recognition and achieved impressive results; ResNet has since been widely applied in other tasks due to its strong feature learning ability. In this work, ResNet-50 is used as a baseline because it learns discriminative representations without consuming too much of the time and computational budgets in writer identification. A ResNet consists of residual units that have two branches. One branch has several convolutional layers and learns the features of the input, and the other bypasses the other branch and forwards the result of the previous layer. These units help the CNN model preserve the identity and maintain a deeper structure. Following the conventional fine-tuning strategy, we use a model pretrained on ImageNet. To avoid model overfitting and to learn more discriminative features, we add a rectified linear unit (ReLU) layer [13] and replace the original pooling layer with a global average pooling layer [26] before the fully connected layer. In addition, we modify the last layer to have K neurons to predict the K classes, where K is the number of classes in the original training data. The extra data are mixed with the original data as the input of the CNN. That is, the original labeled training data and the extra unlabeled data are shuffled and simultaneously trained.



(a) Label distribution of real data (b) Label distribution of extra data

Fig. 4. The label distributions of the real data and extra data used in our proposed semi-supervised feature learning pipeline. The cross-entropy loss combines them and will be simultaneously optimized (Eq. (8)). (a) The label distribution of real data (Eq. (2)) is a one-hot distribution, which shows that the original cross-entropy loss only takes the ground-truth term into account (Eq. (3)). (b) We propose the virtual weighted uniform label distribution for the extra data (Eq. (6)), which is assumed to not belong to any predefined training classes. All extra data will result in an incorrect prediction, and thus, the network will be penalized.

After training, the local features of all test handwritten documents are extracted from the fully connected layer. Additional implementation details are provided in Section 4.3.

Based on prior studies [15,16,28], the time complexity and space complexity of a CNN model are $FLOPs \sim O(\sum_{l=1}^{D} M_l^2 \cdot K_l^2 \cdot C_{l-1} \cdot C_l)$ and $Parameters \sim O(\sum_{l=1}^{D} K_l^2 \cdot C_{l-1} \cdot C_l + \sum_{l=1}^{D} M_l^2 \cdot C_l)$, respectively, where l is the index of a convolutional layer in the model, D is the number of convolutional layers (also known as the "depth"), C_l is the number of filters in the l-th layer (also known as the "width"), K_l is the spatial size of the filter in the l-th layer, and M_l is the spatial size of the output feature map in the l-th layer. Our proposed algorithm has 3.8 billion FLOPs (multiply-adds) and 27.6 million parameters. The running time is 3.16 ± 0.04 seconds for one batch of 128 patches during training, while the running time is 0.0476 ± 0.0016 seconds for the feature extraction of one patch during testing.

3.2.2. Weighted label smoothing regularization method

Label smoothing regularization (LSR) was first used for fully supervised learning in the 1980s and was recently proposed to regularize the classifier layer by estimating the marginalized effect of label dropout during training [42]. In the person reidentification task, Zheng et al. [48] extended LSR to label smoothing regularization for outliers (LSRO), which leveraged unsupervised data generated by GAN and set the virtual label distribution to be uniform over all classes, effectively regularizing the baseline model and achieving better retrieval performance than the baseline. In this work, we propose the WLSR method to regularize the CNN baseline with the extra unlabeled data for offline writer identification. WLSR sets the virtual label distribution to be a weighted uniform distribution over all classes, which effectively regularizes the baseline according to the original training data distribution. For instance, if the original training set has a large number of common features that do not benefit writer identification (e.g., some ink traces and scribe width), the identification model may be misdirected to take these common features as a discriminative representation, which limits the discriminative ability of the model. However, if we add these common features of extra unlabeled data into the model for training, the classifier will make an incorrect prediction toward the labeled words, and thus, the classifier will be penalized. Moreover, the regularization ability of WLSR is decided by the similarity of the sample space between the original labeled data and the extra unlabeled data. If the extra unlabeled data are located nearer the original training data in the sample space, the regularization ability of WLSR will be more effective. Otherwise, the performance of WLSR will be undesirable.

WLSR is proposed to be used with cross-entropy loss. Formally, let $k \in \{1, 2, \ldots, K\}$ be the original training data class and N be the numbers of the original training data. The cross-entropy loss is shown in Eq. (1).

$$l = -\sum_{k=1}^{K} log(p(k))q(k), \tag{1}$$

where $p(k) \in [0, 1]$ is the predicted probability of training data belonging to class k, which is derived from the softmax function that normalizes the output of the previous CNN layer, and $q(k)$ is the ground-truth distribution. Let y be the ground-truth class label. A pair (x_i, y_i) is called the original training example, and $i \in \{1, 2, \ldots, N\}$.

For the original labeled data of the training set, its ground-truth distribution $q_{real}(k)$ is shown in Fig. 4(a). It can be formulated as:

$$q_{real}(k) = \begin{cases} 0, & k \neq y; \\ 1, & k = y. \end{cases} \tag{2}$$

Combining Eqs. (1) and (2), the cross-entropy loss of real data $loss_{real}$ can be rewritten as:

$$loss_{real} = -log(p(y)). \tag{3}$$

From Eq. (3), it is clear that minimizing $loss_{real}$ is equivalent to maximizing the predicted probability of the ground-truth class.

However, LSR was proposed to take the distribution of non-ground-truth classes into consideration [42]. LSR discouraged the network from being confident toward its prediction. Formally, its label distribution $q_{LSR}(k)$ is formulated as:

$$q_{LSR}(k) = \begin{cases} \frac{\varepsilon}{K}, & k \neq y; \\ 1 - \varepsilon + \frac{\varepsilon}{K}, & k = y. \end{cases} \tag{4}$$

where $\varepsilon \in [0, 1]$ is a smoothing parameter. Intuitively, if ε is too large, the network may fail to predict the ground-truth label. Considering Eqs. (1) and (4), the cross-entropy loss is written as:

$$loss_{LSR} = -(1 - \varepsilon)log(p(y)) - \frac{\varepsilon}{K}\sum_{k=1}^{K} log(p(k)). \tag{5}$$

Thus, $loss_{LSR}$ not only takes the ground-truth class into account but also pays attention to other classes, which effectively avoids network overfitting.

We extend LSR from the supervised domain to the semi-supervised domain and propose weighted label smoothing (WLSR) to train the extra unlabeled data. Specifically, we set the virtual label distribution as a weighted uniform distribution over all classes for the extra unlabeled data according to the real data distribution, as shown in Fig. 4(b). Thus, the label distribution of the extra data $q_{WLSR}(k)$ can be formulated as:

$$q_{WLSR}(k) = \frac{\sum_{n=1}^{N} I(y_n = k)}{N}. \tag{6}$$

Thus, by combining Eqs. (1) and (6), the cross-entropy loss of the extra data $loss_{extra}$ can be written as:

$$loss_{extra} = -\sum_{k=1}^{K} log(p(k)) \frac{\sum_{n=1}^{N} I(y_n = k)}{N}, \tag{7}$$

where $I(y_n = k)$ is an indicator function. The proposed semi-supervised feature learning pipeline shuffles and simultaneously trains the real data and the extra data. Combining Eqs. (3) and (7), we can rewrite the cross-entropy loss of semi-supervised feature learning $loss_{WLSR}$ as:

$$\begin{aligned} loss_{WLSR} &= -(1 - Z) \cdot loss_{real} - Z \cdot loss_{extra} \\ &= -(1 - Z) \cdot \log(p(y)) - Z \cdot \sum_{k=1}^{K} log(p(k)) \frac{\sum_{n=1}^{N} I(y_n=k)}{N}, \end{aligned} \tag{8}$$

where Z is an indicator. For the extra data, $Z = 1$. For the original training data, $Z = 0$. Therefore, the proposed semi-supervised feature learning method has two types of loss: one for real images and the other for extra images.

We visualize the intermediate feature maps of the two pretrained models to identify the differences between the baseline (ResNet-50) and the proposed semi-supervised feature learning pipeline (baseline + WLSR). We take some patches of the ICDAR2013 test set for testing. The selected patches belong to various handwritten documents on which the baseline performs poorly, while the desired results are obtained with the semi-supervised feature learning model. For each patch, its activation is obtained from the intermediate layer "res4fx" of the network, the size of which is 14 Ã 14. Then, we visualize the sum of several activation maps. As shown in Fig. 5, the baseline network and the proposed semi-supervised feature learning network activate different patterns in the content of patches. Specifically, the activation maps of the semi-supervised feature learning more accurately and clearly exhibit the contents of the test patches than do the activation maps extracted from the baseline. That is, the representations of the semi-supervised feature learning model are more discriminative, which is why the proposed semi-supervised feature learning produces better results than the baseline.

3.3. Encoding

The all-local descriptors were extracted from the pretrained model during testing. We need to aggregate them to encode a global feature vector for each test document. First, we reduce the dimensionality of the local descriptors with PCA-White, which has been shown to effectively reduce the identification time and improve the identification performance [7,9]. In

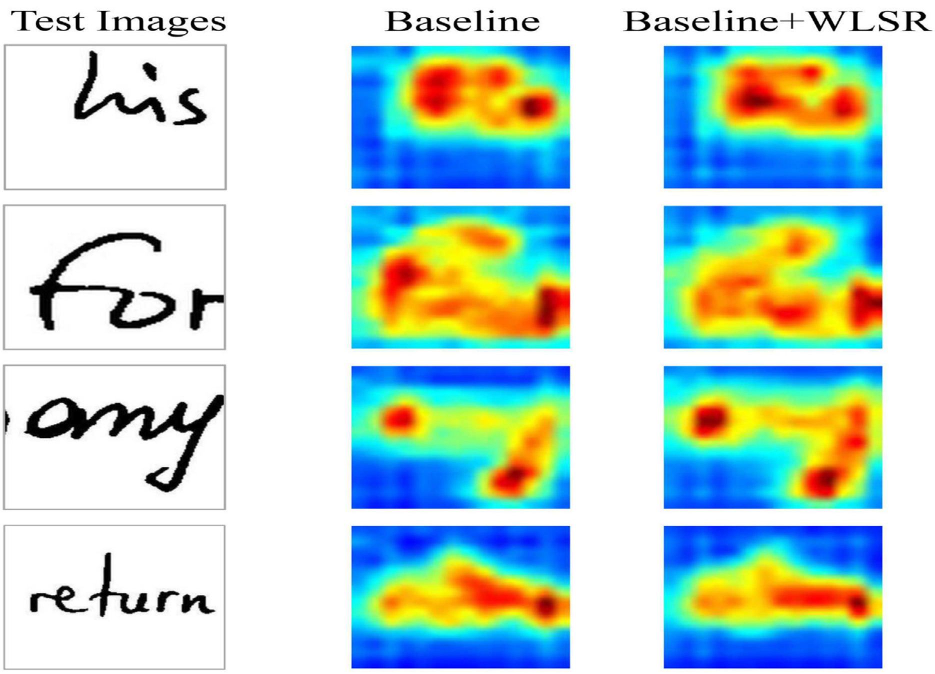


Fig. 5. Visualization of the activation maps of the test patches of the ICDAR2013 test set in the baseline (ResNet-50) and the proposed semi-supervised feature learning model (baseline + WLSR). The baseline and the proposed semi-supervised feature learning network activate different patterns of the patch content. The activation maps of the semi-supervised feature learning network more accurately and clearly represent the contents of the test patches than do the activation maps extracted from the baseline.

addition, we encode the all-local descriptors of each test page as the global feature vector with VLAD, which encodes the first-order statistics by aggregating the residuals of local features to their corresponding nearest cluster centroid. VLAD is a standard encoding method that has been widely used in writer identification [9] and other information retrieval tasks [5,35]. Formally, a codebook $D = \{c_1, c_2, \ldots, c_k\}$ is first computed by k-means with k centroids, and all S local features $f_S \in R^m$ of each test handwritten image are assigned to their nearest cluster centroid. Then, all the residuals between the cluster centroid and the assigned local features are accumulated for each cluster:

$$v_k = \sum_{f_S : NN(f_S) = c_k} (f_S - c_k), \tag{9}$$

where $NN(f_S)$ refers to the nearest neighbor of f_S in dictionary D. All v_k are concatenated as a global feature vector of one handwritten page:

$$v = (v_1^T, v_2^T, \ldots, v_K^T)^T. \tag{10}$$

Thus, the global feature of each test document will eventually be km-dimensional.

4. Evaluation

In the following sections, we describe the datasets and evaluation metrics that we used for evaluating our proposed method. Then, we verify that WLSR has the potential to regularize the baseline for improving identification performance. Furthermore, we show the impacts of using various dimensions of local features, different numbers of extra unlabeled data during training and different centroids of k-means during encoding. Finally, we compare our method to other methods for writer identification.

4.1. Datasets

There are three different benchmark datasets used for evaluation: The ICDAR2013 dataset[1] [29], the CVL dataset[2] [24] and the IAM dataset[3] [31]. All these datasets are publicly available and have been used in many recent publications

[1] http://rrc.cvc.uab.es/.
[2] https://cvl.tuwien.ac.at/category/research/cvl-database/.
[3] http://www.fki.inf.unibe.ch/databases/iam-handwriting-database/.

[6,7,11,32,43,47]. Notably, Fiel and Sablatnig [11] achieved good performance by training a network on the IAM dataset and evaluating the network on the CVL dataset. The results suggested that the word images in the IAM and CVL datasets can share a more similar sample space. Tang and Wu [43] trained his model on the ICDAR2013 dataset, tested on CVL the dataset and provided an impressive identification effect, which revealed that the patches of the CVL and ICDAR2013 datasets have a highly similar sample space. Therefore, we take IAM word images and CVL patches as the extra unlabeled data to evaluate CVL word images and ICDARA2013 patches, respectively.

ICDAR2013 [29]: The ICDAR2013 benchmark dataset is divided into a training set with documents written by 100 writers and a test set with documents written by 250 writers. Every writer contributed four documents, including two Greek documents and two English documents.

CVL [24]: There are 310 writers who contributed documents for the CVL dataset. The 27 writers of the training set contributed seven documents each, and the 283 writers of the test set contributed five documents each. All writers contributed one German document, and the others are English documents.

IAM [31]: The IAM dataset was contributed to by approximately 400 writers with 1066 forms. In the collection, 82,227-word examples are built from a vocabulary of 10,841 words. All of the documents were written in English.

4.2. Evaluation metrics

The mean average precision (mAP) and hard TOP-k, which are common evaluation metrics in image and information retrieval tasks, are used for our experimental evaluation.

A ranked list of all documents in the query library is generated according to the similarity of each query document. Suppose that there are N handwritten documents from the query; thus, the average precision $AP(i)$ of the i_{th} $(1 \le i \le N)$ query document is Eq. (11).

$$AP(i) = \frac{\sum_{k=1}^{M} P(k) \cdot rel(k)}{R} \tag{11}$$

where M is the number of documents in the query library and R is the number of relevant documents of the i_{th} query document in the query library. $P(k)$ is the precision at rank k, which is given by the number of documents from the same writer in the query up to rank k divided by k. $rel(k)$ is an indicator function, where $rel(k) = 1$ when the document retrieved at rank k is from the same writers, and $rel(k) = 0$ otherwise.

The mAP is the mean value of the average precision of all query documents. It can be written as:

$$mAP = \frac{\sum\limits_{i=1}^{N} AP(i)}{N}. \tag{12}$$

The hard TOP-k depends on the calculation of the percentage of the query result, where the k highest ranked documents are from the same writer.

4.3. Experiments

The proposed method was evaluated on the ICDAR2013, CVL and IAM benchmark datasets. We present the implementation details and analysis of the experimental results in the following.

4.3.1. Implementation details

In this work, we adopt the ResNet-50 model as a baseline. To gather more abstract features, we take the global average pooling layer to replace the original pooling layer and add a ReLU activation feature layer. Furthermore, the last fully connected layer was modified to have 100 and 27 neurons for ICADAR2013 and CVL, respectively. We add a dropout layer before the last convolutional layer and set the dropout rate to 0.5 for training. The momentum of stochastic gradient descent is set to 0.9. We set the learning rate of the convolutional layers to 0.1 and have it decay to 0.01 after 45 epochs. To evaluate ICDAR2013, we take the ICDAR2013 training image patches as the original labeled data and the CVL training image patches as the extra unlabeled data. The CVL and IAM datasets already provide a segmentation of words. Thus, we directly take the CVL training words as the original labeled data and the IAM words as the extra unlabeled data to evaluate the CVL dataset. The size of the segmented image patches is set to 256×256, while the width or height of word images was set to 256 pixels and the original aspect ratio was maintained. We extracted the local features of the test images in the first fully connected layer. The similarity between two handwritten documents was calculated by the Euclidean distance for ranking.

4.3.2. Experimental results

First, we evaluate how the number of neurons of the fully connected layer affects writer identification. The number of neurons is set to 512, 1024, 2048, and 4096, which are assessed on the CVL dataset, as shown in Table 1. The semi-supervised feature learning pipeline achieves the best performance on the hard TOP-k and mAP metrics when the number of neurons of the first fully connected layer is set to 2048. Thus, all the following experiments use this configuration.

Table 1
The influence of the number of neurons of the fully connected layer on the CVL test set evaluated with the hard TOP-k and mAP metrics (%).

	TOP-1	TOP-2	TOP-3	TOP-4	mAP
Fc-512	97.9	97.0	93.6	85.0	96.4
Fc-1024	98.4	97.4	94.9	87.9	97.0
Fc-2048	**99.2**	**98.2**	**96.0**	**90.2**	**98.0**
Fc-4096	98.5	97.6	94.7	88.0	97.3

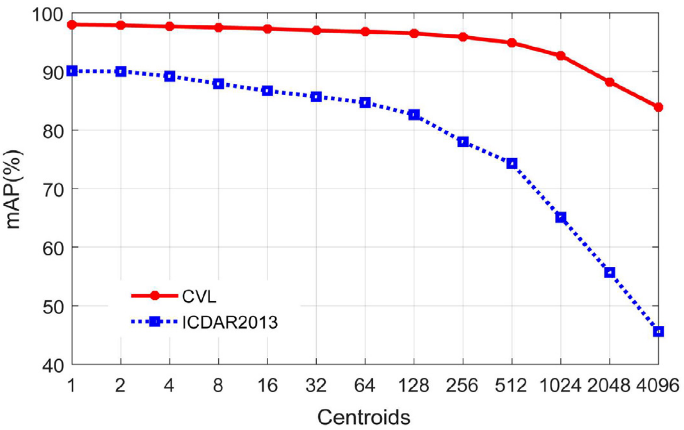


Fig. 6. The influence of the number of centroids during encoding with VLAD. The mAP of the CVL dataset (red solid line) and the ICDAR2013 dataset (blue dotted line) for various numbers of k-means centroids. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 2
Comparison: The proposed semi-supervised feature learning vs. baseline on the CVL and ICDAR2013 test sets.

		TOP-1	TOP-2	TOP-3	TOP-4	mAP
	0 (baseline)	98.3	97.0	92.5	87.0	95.7
CVL	12000 (baseline)	98.4	97.0	94.0	87.2	96.8
	12000(baseline+WLSR)	**99.2**	**97.9**	**96.0**	**90.2**	**97.8**
	0 (baseline)	94.9	74.6	55.1	N/A	88.0
ICDAR2013	1000 (baseline)	95.1	74.3	57.3	N/A	88.1
	1000 (baseline+WLSR)	**96.6**	**79.0**	**61.1**	N/A	**90.1**

Second, we analyze the influence of the number of centroids k during encoding with VLAD. In general, when k is larger, the retrieval performance is better for a large dataset. The experimental results on the ICDAR2013 and CVL datasets are shown in Fig. 6. As shown, when the number of centroids is set to 1, we achieve the largest mAP (98.0% and 90.1% on ICDAR2013 and CVL, respectively). Moreover, the mAP of the two benchmarks consistently decreases as the number of centroids increases. Three reasons may explain the experimental results: **A**. The ICDAR2013 and CVL datasets are too small; therefore, they do not need more image vocabulary to represent themselves. **B**. Every writer wrote the documents with the same content in one dataset, which means that the diversity of the dataset is limited. **C**. The dimensions of the local feature are so large (2048 in this work compared to 64 in [22]) that the local features are discriminative.

Third, we verify the regularization ability of the WLSR method in the semi-supervised feature learning pipeline. The same extra labeled and unlabeled data were added to the supervised baseline and the proposed semi-supervised feature pipeline for training, respectively. As shown in Table 2, the extra labeled data have almost no effect on writer identification for the baseline, while the same unlabeled data improves the identification rate of the semi-supervised feature learning pipeline (on the CVL and ICDAR2013 datasets), which shows that the regularization of WLSR improves the performance of the baseline.

Moreover, we compare the proposed semi-supervised feature learning pipeline with the baseline. As shown in Table 2, when we add 12,000 extra unlabeled IAM words into the CNN for training, our method significantly improves the writer identification performance on the CVL test set, which reveals that the WLSR method achieves improvements of 0.9% (from 98.3% to 99.2%), 0.9% (from 97.0% to 97.9%), 3.5% (from 92.5% to 96.0%), 3.2% (from 87.0% to 90.2%) and 2.1% (from 95.7% to 97.8%) in hard TOP-1, hard TOP-2, hard TOP-3, hard TOP-4, and mAP, respectively. On ICADAR2013, we observe improvements of 1.7%, 4.4%, 6.0% and 2.1% in hard TOP-1, hard TOP-2, hard TOP-3, and mAP, respectively, when 1000 extra unlabeled

Table 3
Comparison of the effect of various numbers of extra unlabeled images on the CVL test set evaluated with the hard TOP-k and mAP metrics (%).

	TOP-1	TOP-2	TOP-3	TOP-4	mAP
0 (baseline+WLSR)	98.3	97.0	92.5	87.0	95.7
1000 (baseline+WLSR)	98.8	97.9	95.0	88.5	97.3
5000 (baseline+WLSR)	98.9	97.9	95.4	88.9	97.5
12000 (baseline+WLSR)	**99.2**	**97.9**	**96.0**	**90.2**	**97.8**
24000 (baseline+WLSR)	99.0	97.9	95.2	89.9	97.6

Table 4
Comparison of the effects of the numbers of extra unlabeled images on the ICDAR2013 test set evaluated with the hard TOP-k and mAP metrics (%).

	TOP-1	TOP-2	TOP-3	mAP
0 (baseline+WLSR)	94.9	74.6	55.1	88.0
500 (baseline+WLSR)	94.8	75.5	56.3	88.1
1000 (baseline+WLSR)	**96.6**	**79.0**	**61.1**	**90.1**
2000 (baseline+WLSR)	96.5	78.6	59.6	90.0
5000 (baseline+WLSR)	94.9	74.3	56.5	88.0

Table 5
Comparison of the performance with other methods on the CVL test set. The hard TOP-k and mAP metrics are listed (%).

	TOP-1	TOP-2	TOP-3	TOP-4	mAP
CS-UMD [24]	97.9	90.0	71.2	48.3	N/A
QUQA A [24]	30.5	5.7	0.5	0.1	N/A
QUQA B [24]	92.9	84.9	71.5	50.6	N/A
TEBESSA-c [24]	97.6	94.3	88.2	73.9	N/A
TSINGHUA [24]	97.7	95.3	94.5	7.30	N/A
Fiel and Sablatnig [11]	98.9	97.6	93.3	79.9	N/A
Christlein et al. [6]	99.2	98.1	95.8	88.7	97.1
Nicolaou et al. [32]	99.0	97.7	95.2	86.0	N/A
Christlein et al.[7]	98.8	97.8	95.3	88.8	96.4
Ours (single)	99.2	97.9	96.0	90.2	97.8
Ours (2-streams)	**99.2**	**98.4**	**96.1**	**91.5**	**98.0**

CVL patches are added in ICDAR2013, as shown in Table 2. Thus, the proposed semi-supervised feature learning pipeline effectively improves upon the performance of the baseline.

In addition, we noted that the amount of additional unlabeled data substantially affects the regularization ability of the WLSR, as shown in Tables 3 and 4. In terms of Table 3 (CVL dataset), the regularization of the WLSR is insufficient if a small amount of extra unlabeled data (e.g., 0, 1000 or 5000 cases) is incorporated into the pipeline. However, if a large amount of extra unlabeled data (e.g., 24,000 cases) is added, the pipeline tends to assign weighted uniform prediction probabilities to all training data. This result indicates that our model is over-regularized by the WLSR; thus, the model may incorrectly predict the original labeled data. The similar results in Table 4 (ICDAR2013 dataset) indicate that the best performance is achieved with 1000 extra unlabeled data. Under-regularization and over-regularization occur with 0/500 extra unlabeled data and 2000/5000 extra unlabeled data, respectively. In summary, an appropriate amount of extra unlabeled data can ensure that the proposed semi-supervised feature learning avoids under-regularization and over-regularization.

Finally, we combined the two models generated by our method to form an ensemble (2-stream) to further enhance the identification performance and compared our proposed method with the other published methods on the ICDAR2013 and CVL datasets, as listed in Tables 5 and 6, respectively. The semi-supervised feature learning pipeline achieves better results than most other supervised approaches. On the CVL dataset, we achieve hard TOP-1 = 99.2%, hard TOP-2 = 98.4%, hard TOP-3 = 96.1%, hard TOP-4 = 91.5, and mAP = 98.0%, which are better results that those achieved by the other supervised methods. On ICDAR2013, we achieved hard TOP-1 = 97.7%, hard TOP-2 = 83.3%, hard TOP-3 = 63.0, and mAP = 91.1%, which are also very competitive results compared to the results of the other methods. Specifically, the proposed semi-supervised feature learning method produces the desired performance on the ICDAR2013 test set with a few extra unlabeled patches of the CVL training set, while Christlein et al. [6] added the entire CVL training set to ICDAR2013 for training and achieved ordinary results. The results in Tables 5 and 6 show that the semi-supervised feature learning method takes full advantage of the extra data and is more convenient to use in practice than other supervised methods [6–8,11,24,29,32]. Fig. 7 presents some identification results achieved by the proposed semi-supervised feature learning method (single) on the ICDAR2013 dataset (sample 1–2, sample 22-4, sample 24-3, and sample 248-1). The images (gray border) are the query images. The identification images (red border and green border) are sorted according to the similarity scores from top to bottom (from

Table 6
Comparison of the performance with the other methods on the ICDAR2013 test set. The hard TOP-k and mAP metrics are shown (%).

	TOP-1	TOP-2	TOP-3	mAP
CS-UMD-b [29]	95.0	20.2	8.4	N/A
HIT-ICG [29]	94.8	63.2	36.5	N/A
TEBESSA-c [29]	93.4	62.6	36.5	N/A
CVL-IPK [29]	90.9	44.8	24.5	N/A
Fiel and Sablatnig [11]	88.5	40.5	15.8	N/A
Christlein et al. [6]	97.1	42.8	23.8	67.1
Nicolaou et al. [32]	97.2	52.9	29.2	N/A
Christlein et al. [7]	**98.2**	71.2	47.7	81.4
Ours (single)	96.6	79.0	61.1	90.1
Ours (2-streams)	97.7	**83.3**	**63.7**	**91.8**

Rank-1 to Rank-5). Images with a green border are correct candidates, and images with a red border images are incorrect candidates. Most ground-truth candidate images are correctly identified.

5. Discussion

In this study, we visualized the intermediate feature maps of the baseline and semi-supervised feature learning pipeline (Sec. 3.2.2). The results showed that the activation maps of the semi-supervised feature learning more accurately represent the contents of the test patches than do the activation maps extracted from the baseline. Then, we analyzed the impact of the dimensions of the local features, the centroids of VLAD encoding and the amount of extra unlabeled data Section 4.3.2. Moreover, we experimentally showed that the proposed method could significantly improve the baseline and perform competitively with existing writer identification approaches, which benefit from the potential of regularization of WLSR. WLSR takes full advantage of extra unlabeled data for regularizing the baseline, and thus, the CNN learns effective and discriminative features.

Due to some common representations in the extracted features, some researchers combined multiple handcrafted elements to derive a more reliable discriminative feature, yet restraining the impact of common features. For example, Helli extracted features using Gabor and XGabor filters and then developed a feature relation graph [20]. In terms of the width of ink traces, a powerful source of information for offline writer identification consisted of a powerful feature (Quill) in combination with the direction [3]. In [18], they proposed a novel junction detection method for writer identification using stroke-length distribution and direction of ink of texts. Motivated by the above methods, we proposed a WLSR method to regularize and penalize the common features that are automatically learned features by the CNN and reducing the negative influence of common features.

Our proposed semi-supervised feature learning approach suffers from the limitation that WLSR depends on the similarity of the sample spaces of the original labeled data and the extra unlabeled data. In the future, the generative adversarial networks (GANs), a system of two neural networks competing with each other in a zero-sum game framework, may be a potential choice to overcome this limitation. Because we can generate data by GANs and original data share the same sample space, we do not require any extra data from other datasets.

6. Conclusion

In this paper, we proposed a semi-supervised feature learning pipeline for offline writer identification. To the best of our knowledge, this is a pioneering work that uses semi-supervised feature learning to automatically learn discriminative features to improve writer identification. Notably, the WLSR method is introduced to train on the extra unlabeled data and the original labeled data simultaneously to provide the semi-supervised feature learning pipeline with regularization ability, which improved the identification results of the baseline model and achieved better performance than other popular methods on the CVL and ICDAR2013 datasets.

In the future, we will choose a better encoding method that is suitable for small datasets of writer identification tasks to replace VLAD. Furthermore, we will adopt the unlabeled data generated by GANs to train the semi-supervised feature learning network because the generated data share a similar sample space with the original labeled data.

Acknowledgments

We thank reviewers for their helpful comments. Shiming and Yisong's work is partially supported by the National Natural Science Foundation of PR China (under Grants 61370161, 61562009 and U1836105) and Outstanding Young Talent Training Fund of Guizhou Province of PR China (under grant (2015)01).

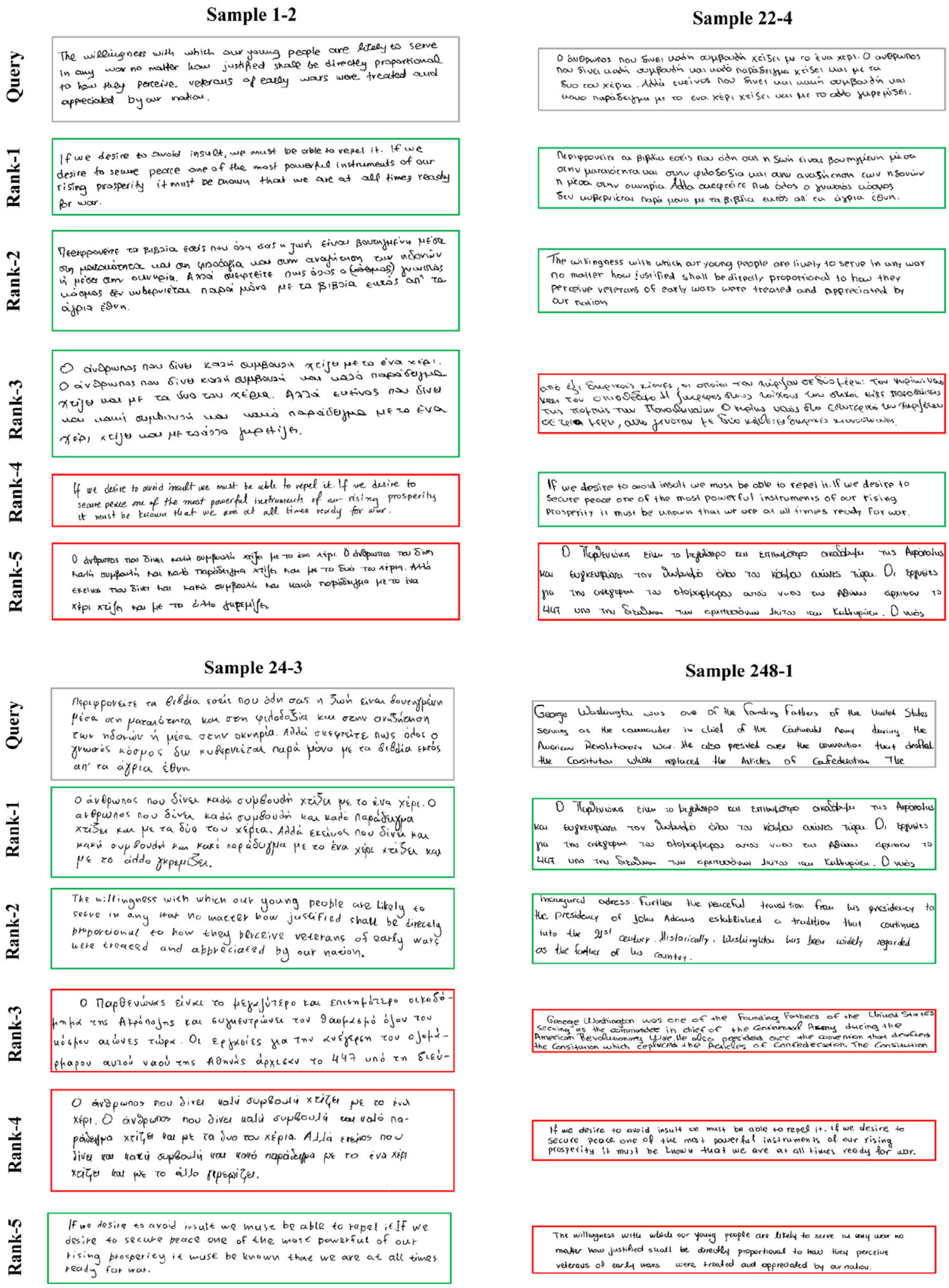


Fig. 7. Writer identification results of the proposed semi-supervised feature learning method (single) on the ICDAR2013 dataset (sample 1–2, sample 22-4, sample 24-3, and sample 248-1). The images (gray border) are the query images. The identification images are sorted according to the similarity scores from top to bottom (from Rank-1 to Rank-5). We maintain the original aspect ratio of the images. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

References

[1] M.N. Abdi, M. Khemakhem, A model-based approach to offline text-independent arabic writer identification and verification, Pattern Recognit. 48 (5) (2015) 1890–1903.
[2] G.R. Ball, S.N. Srihari, Semi-supervised learning for handwriting recognition, in: Proceedings of the International Conference on Document Analysis and Recognition, 2009, pp. 26–30.
[3] A.A. Brink, J. Smit, M.L. Bulacu, L.R.B. Schomaker, Writer identification using directional ink-trace width measurements, Pattern Recognit. 45 (1) (2012) 162–171.
[4] M. Bulacu, L. Schomaker, Text-independent writer identification and verification using textual and allographic features, IEEE Trans. Pattern Anal. Mach. Intell. 29 (4) (2007) 701–717.

[5] C. Chattopadhyay, S. Das, Supervised framework for automatic recognition and retrieval of interaction: a framework for classifcation and retrieving videos with similar human interactions, Iet Comput. Vision 10 (3) (2016) 220–227.
[6] V. Christlein, D. Bernecker, F. Honig, E. Angelopoulou, Writer identification and verification using GMM supervectors, in: Proceedings of the IEEE Winter Conference on Applications of Computer Vision, 2014, pp. 998–1005.
[7] V. Christlein, D. Bernecker, F. Hnig, A. Maier, E. Angelopoulou, Writer identifcation using GMM supervectors and exemplar-SVMS, Pattern Recognit. 63 (2017) 258–267.
[8] V. Christlein, D. Bernecker, A. Maier, E. Angelopoulou, Offline writer identification using convolutional neural network activation features, Proc. German Conf. Pattern Recognit. (2015) 540–552.
[9] V. Christlein, M. Gropp, S. Fiel, A. Maier, Unsupervised feature learning for writer identification and writer retrieval, in: Proceedings of the International Conference on Document Analysis and Recognition, 2018, pp. 991–997.
[10] R. Eldan, O. Shamir, The power of depth for feedforward neural networks, 2015, arXiv:1512.03965.
[11] S. Fiel, R. Sablatnig, Writer identification and retrieval using a convolutional neural network, in: Proceedings of the International Conference in Computer Analysis of Images and Patterns, 2015, pp. 26–37.
[12] B.G.G. Louloudis, N. Stamatopoulos, ICDAR 2011 writer identification contest, in: Proceedings of the International Conference on Document Analysis and Recognition, 2011, pp. 1475–1479.
[13] X. Glorot, A. Bordes, Y. Bengio, Deep Sparse Rectifier Neural Networks, in: Proceedings of the International Conference on Artificial Intelligence and Statistics, 2011, pp. 315–323.
[14] Y. Hannad, I. Siddiqi, M.E.Y.E. Kettani, Writer identification using texture descriptors of handwritten fragments, Expert Syst. Appl. 47 (2016) 14–22.
[15] K. He, J. Sun, Cvolutional neural networks at constrained time cost, in: Proceedings of the IEEE Computer Vision and Pattern Recognition, 2014, pp. 5353–5360.
[16] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Computer Vision and Pattern Recognition, 2016, pp. 770–778.
[17] S. He, L. Schomaker, Writer identification using curvature-free features, Pattern Recognit. 63 (2017) 451–464.
[18] S. He, M. Wiering, L. Schomaker, Junction detection in handwritten documents and its application to writer identification, Pattern Recognit. 48 (12) (2015) 4036–4048.
[19] Z. He, X. You, Y.Y. Tang, Writer identification of chinese handwriting documents using hidden Markov tree model, Pattern Recognit. 41 (4) (2008) 1295–1307.
[20] B. M. M. Helli, A text-independent persian writer identification based on feature relation graph (FRG), Pattern Recognit. 43 (6) (2010) 2199–2209.
[21] G. Huang, S. Song, J.N. Gupta, C. Wu, Semi-supervised and unsupervised extreme learning machines, IEEE Trans. Cybern. 44 (12) (2017) 2405–2417.
[22] H. Jegou, F. Perronnin, M. Douze, J. Sanchez, P. Perez, C. Schmid, Aggregating local image descriptors into compact codes, IEEE Trans. Pattern Anal. Mach. Intell. 34 (9) (2012) 1704–1716.
[23] H. Jgou, O. Chum, Negative evidences and co-occurences in image retrieval: the benefit of PCA and whitening, in: Proceedings of the European Conference on Computer Vision, 2012, pp. 774–787.
[24] F. Kleber, S. Fiel, M. Diem, R. Sablatnig, CVL-database: an off-line database for writer retrieval, writer identification and word spotting, in: Proceedings of the International Conference on Document Analysis and Recognition, 2013, pp. 560–564.
[25] D.H. Lee, Pseudo-label : the simple and efficient semi-supervised learning method for deep neural networks, in: Proceedings of the ICML 2013 Workshop: Challenges in Representation Learning, 2013, pp. 1–6.
[26] M. Lin, Q. Chen, S. Yan, Network in network, 2013, arXiv:1312.4400.
[27] C.L. Liu, F. Yin, D.H. Wang, Q.F. Wang, Online and offline handwritten chinese character recognition: benchmarking on new databases, Pattern Recognit. 46 (1) (2013) 155–162.
[28] H. Liu, Z. Fu, J. Han, L. Shao, S. Hou, Y. Chu, Single image superresolution using multi-scale deep encoder-decoder with phase congruency edge map guidance, Inf. Sci. 473 (2019) 44–58.
[29] G. Louloudis, B. Gatos, N. Stamatopoulos, A. Papandreou, Icdar 2013 competition on writer identification, in: Proceedings of the International Conference on Document Analysis and Recognition, 2013, pp. 1397–1401.
[30] B.P. Mallikarjunaswamy, Writer identification based on offline handwritten document images in kannada language using empirical mode decomposition method, Int. J. Comput. Appl. 30 (6) (2013) 31–36.
[31] U.V. Marti, H. Bunke, The IAM-database: an english sentence database for offline handwriting recognition, Int. J. Doc. Anal. Recogn. 5 (1) (2002) 39–46.
[32] A. Nicolaou, A.D. Bagdanov, M. Liwicki, D. Karatzas, Sparse radial sampling LBP for writer identification, in: Proceedings of the International Conference on Document Analysis and Recognition, 2015, pp. 716–720.
[33] N. Otsu, A threshold selection method from gray-level histograms, IEEE Trans. Syst. Man Cybern.: Syst. 9 (1) (2007) 62–66.
[34] G. Papandreou, L.C. Chen, K.P. Murphy, A.L. Yuille, Weakly- and semi-supervised learning of a deep convolutional network for semantic image segmentation, in: Proceedings of the IEEE International Conference on Computer Vision, 2016, pp. 1742–1750.
[35] M. Paulin, J. Mairal, M. Douze, Z. Harchaoui, F. Perronnin, C. Schmid, Convolutional patch representations for image retrieval: an unsupervised approach, Int. J. Comput. Vis. 121 (1) (2016) 1–20.
[36] U. Porwal, V. Govindaraju, Semi-supervised framework for writer identification using structural learning, IET Biom. 2 (4) (2013) 208–215.
[37] U. Porwal, C. Ramaiah, A. Shivram, V. Govindaraju, Structural learning for writer identification in offline handwriting, in: Proceedings of the International Conference on Frontiers in Handwriting Recognition, 2012, pp. 417–422.
[38] A. Rasmus, H. Valpola, M. Honkala, M. Berglund, T. Raiko, Semi-supervised learning with ladder networks, in: Proceedings of the International Conference on Neural Information Processing Systems, 2015.
[39] L. Schomaker, M. Bulacu, Automatic writer identification using connected-component contours and edge-based features of uppercase western script, IEEE Trans. Pattern Anal. Mach. Intell. 26 (6) (2004) 787–798.
[40] H. Srinivasan, S. Srihari, A statistical approach to line segmentation in handwritten documents, in: Proceedings of the SPIE Document Recognition and Retrieval XIV, 2007, pp. 1–11.
[41] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going deeper with convolutions, in: Proceedings of the IEEE Computer Vision and Pattern Recognition, 2015, pp. 1–9.
[42] C. Szegedy, V. Vanhoucke, S. Ioe, J. Shlens, Z. Wojna, Rethinking the inception architecture for computer vision, in: Proceedings of the IEEE Computer Vision and Pattern Recognition, 2016, pp. 2818–2826.
[43] Y. Tang, X. Wu, Text-independent writer identification via CNN features and joint bayesian, in: Proceedings of the International Conference on Frontiers in Handwriting Recognition, 2017, pp. 566–571.
[44] R.R. Varior, B. Shuai, J. Lu, D. Xu, G. Wang, A siamese long short-term memory architecture for human re-identification, in: Proceedings of the European Conference on Computer Vision, 2016, pp. 135–153.
[45] J. Weston, F. Ratle, R. Collobert, Deep learning via semi-supervised embedding, in: Proceedings of the International Conference on Machine Learning, 2012, pp. 1168–1175.
[46] X. Wu, Y. Tang, W. Bu, Offline text-independent writer identification based on scale invariant feature transform, IEEE Trans. Inf. Forensics Secur. 9 (3) (2014) 526–536.
[47] L. Xing, Y. Qiao, Deepwriter: a multi-stream deep CNN for text independent writer identification, in: Proceedings of the International Conference on Frontiers in Handwriting Recognition, 2016, pp. 584–589.
[48] Z. Zheng, L. Zheng, Y. Yang, Unlabeled samples generated by GAN improve the person re-identification baseline in vitro, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 3774–3782.

Shiming Chen has been a full-time M.D. student in the School of Computer Science and Technology, Guizhou University, China since 2016. He will pursue a Ph.D. in 2019. His research interests include pattern recognition, computer vision and machine learning. He concentrates most of his attention on content-based image retrieval (CBIR).

Yisong Wang is currently a Professor at Guizhou University. He received B.S., M.S. and Ph.D. degrees from Guizhou University in 1998, 2004 and 2007, respectively. Dr. Wang currently serves as an Associate Editor of the Annals of Mathematics and Artificial Intelligence (2018-). His main research interests include knowledge representation and reasoning, nonmonotonic reasoning and answer set programming in particular, inductive logic programming and their applications.

Chin-Teng Lin is currently a Distinguished Professor with the University of Technology Sydney, Sydney, NSW, Australia. He has authored over 200 journal articles (Total Citation: 18 960, H-index: 63, and i10-index: 242), including 97 IEEE journal articles, in the areas of neural networks, fuzzy systems and cognitive neuroengineering. Dr. Lin was a Fellow of IEEE in 2005 and a Fellow of the International Fuzzy Systems Association in 2012. He served as the Editor-in-Chief of IEEE Transactions on Fuzzy Systems from 2011 to 2016. He is a Distinguished Lecturer in IEEE Computational Intelligence Society in 2018.

Weiping Ding is currently an Associate Professor at Nantong University and was awarded the Excellent Paper (First Prize) in Computer Education by the Chinese National Committee of Computer Education and a Chinese Government Scholarship for Overseas Studies in 2011 and 2016. He was an Excellent Young Teacher in Jiangsu Province in 2014 and a High-level Talent in Jiangsu Province in 2016. Dr. Ding currently serves as an Associate Editor of IEEE Transactions on Fuzzy Systems (2015-) and Information Sciences (2016-). His main research interests include data mining, quantum co-evolutionary computing, and machine learning, and their applications in big data.

Zehong Cao completed the dual PhD program with the Faculty of Engineering and Information Technology, University of Technology Sydney, Australia, and the Institute of Electrical and Control Engineering, National Chiao Tung University, Taiwan in 2017. He is currently a Lecturer (Assistant Professor) in Discipline of ICT, University of Tasmania, Australia and honorary research fellow in the University of Technology Sydney, Australia. He was honored as an Associate Editor of IEEE Access and as a Guest Editor of Swarm and Evolutionary Computation, and Neurocomputing. His research interests include signal processing, data mining, human-machine interfaces, fuzzy systems, neural networks, optimization, pattern recognition, and machine learning.