PS-LSTM: Capturing Essential Sequential Online Information with Path Signature and LSTM for Writer Identification

Manfei Liu, Lianwen Jin*, Zecheng Xie
College of Electronic and Information Engineering
South China University of Technology
Guangzhou, China
manfei.l.liu@gmail.com, *lianwen.jin@gmail.com, zchengxie@gmail.com

***Abstract*—Writer identification plays a significant role in several applications including forensic trace evidence identification, mobile bank transaction authentication, and handwritten character/text recognition. However, performing writer identification requires a considerable amount of experimental work and labor, as well as professional skills. In this paper, we propose a novel path-signature long short-term memory (PS-LSTM) recurrent neural network for writer identification that contributes as follows: 1) A mathematical feature set, path signature, is successfully applied to writer identification to characterize the essential geometric and analytic properties of pen-tip trajectory, which help distinguish between the diverse writing styles of different people, especially in confusing situations. 2) Different iterated integrals of path signature are investigated to model the local curvature nature of the pen-tip trajectory for writer identification. 3) Our study is the first to demonstrate the significance of integrating path signature with long short-term memory for writer identification. On Database CASIA-OLHWDN1.0, the proposed PS-LSTM writer identification achieves significantly superior results and outperforms previous works.**

***Keywords*-Writer identification; path-signature; lstm;**

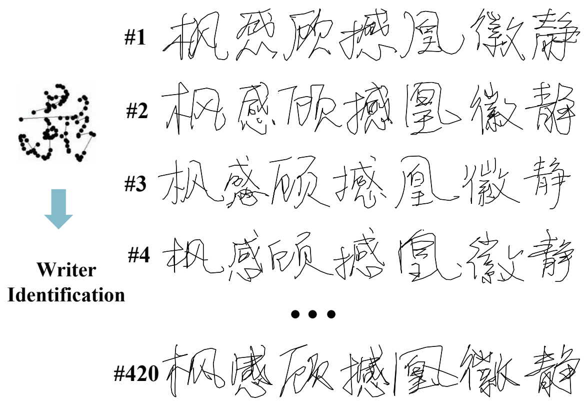

Fig. 1. Typical samples of the writings of different writers in CASIA-OLHWDN1.0. Different writers usually have different writing styles. PS-LSTM learns from part of classes of characters, i.e., 200 classes, and tries to predict the writers of all the remaining classes of characters, i.e., 3666 classes.

I. Introduction

Given handwritten pen-tip trajectories, writer identification involves predicting a probability distribution over all the writers according to the degree of similarity between them and the given trajectories. Writer identification has attracted significant attention from researchers owing to its immense potentials in practical applications, like forensic trace evidence identification, mobile bank transaction authentication, and handwritten character/text recognition augmentation. Given the complexity and diversity of handwriting styles of different writers, traditional approaches utilized sophisticated hand-crafted features for writer identification, such as histogram-based features [1], allographic and textural features [2], interval-valued symbolic features [3], and a designed feature [4] of 45 types including curvature, (x, y)-coordinates, stroke length, etc.

With the development and popularity of the deep learning technique, numerous approaches leverage the outstanding networks, e.g., deep convolutional neural network (CNN) [5] and long short-term memory (LSTM) [6], to improve writer identification. Yang et al [7], [8] integrated path signature with deep CNN as well as data augmentation techniques, namely, DropStroke [8] and DropSegment [7], for writer identification and substantially improved the accuracy. Recently, the research community has witnessed several impressive capabilities of LSTM in different research areas, such as handwritten character/text recognition [9]–[11], visual captioning [12], [13] and language translation [14]. As for writer identification, Zhang et al [15] presented an end-to-end online writer identification approach with Recurrent Neural Networks (RNNs) with random hybrid strokes (RHSs), which achieved a state-of-the-art result among the existing methods.

Among above-mentioned methods, the recently developed deep learning architecture LSTM is significantly advantageous over traditional approaches for writer identification. However, the system proposed by Zhang et al [15] only considered the (x, y)-coordinate system and binary pressure information, i.e, pen-up and pen-down, without properly utilizing the domain knowledge required for feature extraction. As pointed out by Yang et al [16], domain-specific information can help improve the recognition performance with a deep learning network. It is noteworthy that path-signature, a novel mathematical feature set from rough theory [17]–[19] has been successfully applied to extracting complex online information from pen-tip trajectories [16], [20]–[22]. Therefore, in this paper, we propose a

2379-2140/17 $31.00 © 2017 IEEE
DOI 10.1109/ICDAR.2017.114

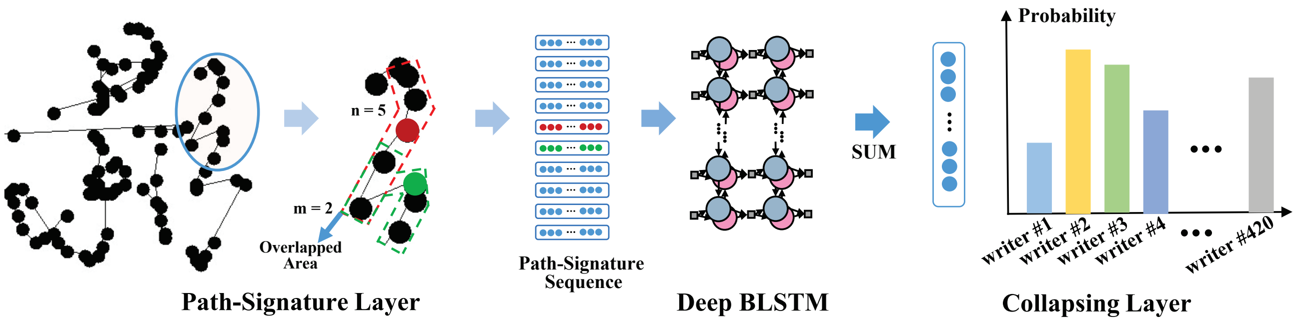

Fig. 2. Illustration of network architecture and flow schematic of PS-LSTM. PS-LSTM is constructed by the path-signature layer, Deep BLSTM and the Collapsing layer, from bottom to top. Given the pen-tip trajectory collected from a certain writer, path-signature layer extracts features for the overlapped sequential local area and generates the path-signature sequence. Then, Deep BLSTM takes the path-signature sequence as well as the past hidden states as input and outputs the predicting vector for each time step. Finally, the collapsing layer sums over all the predicting vectors along each dimension independently and provides a probability distribution over all the writers.

path-signature long short-term memory (PS-LSTM) for writer identification; it contributes as follows. (1) Our work is the first work to investigate and demonstrate the performance of integrating path signature with long short-term memory for writer identification. (2) The path-signature is proved able to characterize the essential sequential online information of pen-tip trajectory, which helps distinguish between the diverse complex writing styles of different people, even in confusing situations. (3) We investigate different iterated integrals of path signature for modeling the local curvature nature of the pen-tip trajectory in the process of writer identification.

The remainder of this paper is organized as follows. Section II presents the network architecture of PS-LSTM. Section III provides details on the data augmentation technique used in this paper. Section IV presents the experimental results. Lastly, Section V concludes the paper.

II. Path-Signature LSTM Recurrent Neural Network

Given training set χ with a training sample (x, z), where x represents the pen-tip trajectories of a writer and z denotes the corresponding label, the proposed PS-LSTM, denoted by g, is intended at minimizing loss function $L(\chi)$ as the negative log probability of correctly predicting the writer for all the training samples in χ:

$$L(Q) = -\ln \prod_{(x,z)\subset Q} p_g(z|x) = -\sum_{(x,z)\subset Q} \ln p_g(z|x). \quad (1)$$

As described in Fig. I, the proposed PS-LSTM consists of three main components, namely the path-signature layer, deep bidirectional LSTM (BLSTM) recurrent neural network, and collapsing layer. Given a sequential pen-tip trajectory, the path-signature layer will generate a path-signature sequence for each subsequence to represent the whole path. Then, the BLSTM recurrent neural network takes the feature vector and past hidden state as input for each time step and outputs the corresponding predicting feature sequence. Finally, the collapsing layer will considers the sum over all the predicting feature for each dimension along the time dimension individually, which produces a feature vector for predicting a probability distribution over all the writer.

A. Path-Signature Layer

Given a sequential pen-tip trajectory $x = x_1, x_2, \cdots, x_l$ from a certain writer, we apply the path-signature layer to extract the path-signature feature from the overlapped subsequences of the original pen-tip trajectories. Specifically, the length of these subsequences is n and that of the overlapped area is m. After performing the path-signature feature extraction, the pen-tip trajectory with length l reduces to a feature sequence with length $\frac{l-n}{m}+1$, namely the path-signature sequence. Specifically, each feature vector in the path-signature sequence characterizes the analytic and geometric property of a local area. By concatenating all these feature vectors, we create the path-signature sequence that represents the essential sequential online information of the original pen-tip trajectory.

Path signature is pioneered by Chen [23] and advocated by Lyons et al [17]–[19] as playing a fundamental role in rough theory. Essentially, path signature is intended at extracting sufficient information to represent the paths, such as pen-tip trajectories and acoustic waves, with finite length. In the following, we briefly introduce the principle and calculation of path-signature.

Typically, a pen-tip trajectory P can be represented by: $P : [T_a, T_b] \rightarrow P$ with $W \subset R^2$ denoting a writing plane and $[T_a, T_b]$ denoting a time interval. Given a specific interval $[t_a, t_b] \subset [T_a, T_b]$, we can calculate the k-th iterated integral of the pen-tip trajectory P as follows:

$$P^k_{t_a,t_b} = \int_{t_a<q_1<\cdots<q_k<t_b} 1 dP_{q_1} \otimes, \cdots, \otimes dP_{q_k}. \quad (2)$$

where $P^k_{t_a,t_b}$ is a 2^k dimensional vector. It is noteworthy that the dimension of the k-th iterated integral of pen-tip trajectory P grows exponentially while carrying merely little more knowledge about the path; thus, using truncated signature in practical application is preferable. Specifically, the truncated path signature with level n can be represented as follows:

$$S(P)^n_{t_1,t_2} = (1, P^1_{t_1,t_2}, \cdots, P^n_{t_1,t_2}). \quad (3)$$

The dimension of the truncated path signature is $2^{(n+1)} - 1$. However, since $k = 0$ iterated integral always remain one, it carries no valuable information for writer identification problem, we eliminate this dimension and use the novel truncated path signature $X(P)^n_{t_1,t_2}$ as follows:

$$X(P)^n_{t_1,t_2} = (P^1_{t_1,t_2}, \cdots, P^n_{t_1,t_2}). \quad (4)$$

Suppose P is a straight line, which is true for the writer identification problem, we can calculate iterated integrals $P^k_{t_1,t_2}$ using:

$$\begin{aligned} P^1_{t_1,t_2} &= \triangle_{t_1,t_2}, \\ P^2_{t_1,t_2} &= (\triangle_{t_1,t_2} \otimes \triangle_{t_1,t_2})/2!, \\ P^3_{t_1,t_2} &= (\triangle_{t_1,t_2} \otimes \triangle_{t_1,t_2} \otimes \triangle_{P_1,t_2})/3!, \cdots, \end{aligned} \quad (5)$$

where $\triangle_{t_1,t_2} := S_{t_2} - S_{t_1}$ denotes path displacement.

B. Deep Bidirectional LSTM

Recurrent neural network (RNN), an extension of the feedforward neural network, has been widely studied in the research community owing to its inherent nature of being useful in solving sequence problem. Essentially, RNN can be applied in problems such as video classification [24], i.e., encoding variable-length input to a fixed-length feature vector, image description [25], i.e., translating fixed-length input to variable-length output, video description and language translation [12], [14], i.e., translating variable-length input to variable-length output. In this paper, we regarded the writer identification problem as translating the input path-signature sequence into a sequence of predicting vectors, which will be summed up in the collapsing layer. Specifically, given input sequence $\boldsymbol{x} = x_1, x_2, \cdots, x_T$ with length T, RNN takes x_t as well as a past hidden state as input, and it outputs the corresponding predicting feature vector for each time step t, and generates hidden states $\boldsymbol{h} = h_1, h_2, \cdots, h_T$ with length T:

$$h(t) = f(x_t, h_{t-1}). \quad (6)$$

However, traditional RNN suffers from gradient vanishing and exploding problem, which in turn prevents the RNN from learning long-term context information in the sequential data. Recently, long short-term memory (LSTM), essentially an extension of RNN, has attracted the attention of the research community. As described in Fig. 3, the core of LSTM is a memory cell c, with three gates, namely input gate (i), forget gate (f) and output gate (o). Specifically, the input gate controls the input of the data, forget gate determines whether to remember the past memory dynamics, and output gate allows the the information to flow out the cell. Formally, the underlying mechanism of LSTM can be represented as follows (for each time step t):

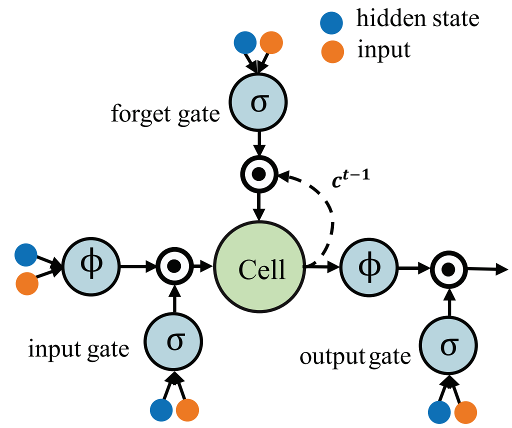

Fig. 3. Long short-term memory (LSTM) cell.

$$\begin{aligned} i_t &= \sigma(W_{xi}x_t + W_{hi}h_{t-1}), \\ f_t &= \sigma(W_{xf}x_t + W_{hf}h_{t-1}), \\ o_t &= \sigma(W_{xo}x_t + W_{ho}h_{t-1}), \\ c_t &= f_t \odot c_{t-1} + i_t \odot \phi(W_{xc}x_t + W_{hc}h_{t-1}), \\ h_t &= o_t \odot \phi(c_t). \end{aligned} \quad (7)$$

where the weight matrices W_ij represents the trained parameters of the cell. ϕ is the hyperbolic tangent non-linearity, σ is the sigmoidal non-linearity, and $\odot$ denotes the simple scalar multiplication.

Although LSTM is capable of capturing the past long-term dependencies, there remains context information from the other direction; thus, it is natural to apply bidirectional LSTM (BLSTM) to learning complex contextual knowledge underlying the sequential data, e.g., the pen-tip trajectories of the writers. Furthermore, we followed the suggestion of Pascanu et al [26] to stack multiple BLSTM layers to construct our PS-LSTM, so as to exploit the essential contextual information of the pen-tip trajectory.

C. Collapsing Layer

Given the predicting sequence of the deep BLSTM, the collapsing first sums over all the predicting vector of the sequence along the time axis for each dimension, individually, which results in a summed feature vector with the same dimension of the these predicting feature vector. Then, the collapsing layer can generate the probability distribution over all the writers based on this summed feature vector.

III. Data Augmentation Technique

In this paper, we followed the augmentation technique, *DropSegment*, proposed by Yang et al [7] to overcome the following problems commonly faced in writer identification. First, given the fact that identification materials as well as characters vary in practical application, researchers often waste time on a trivial segmentation problem. Such a phenomenon increases in severity when dealing with a different language, such as Chinese and English. Secondly, suppose there is a character/word written by two different writers, which results into two different samples. Although these two samples possess different writing styles which is crucial for writer identification, they are essentially very similar in structure, which would easily cause confusion and affect the performance of the identification system. Furthermore, in actual application, for example, Bank Transaction, there is barely sufficient training data from the customers; thus, developing a data augmentation technique to enrich the training samples would play an important role in writer identification. Lastly, data augmentation technique *DropSegment* can also promote the recognition performance of the system by averaging the results under different settings of *DropSegment*. This is completely different from the traditional ensemble techniques that rely on multiple different models.

In the following, we briefly introduce the mechanism of *DropSegment*. Given a character $\boldsymbol{w} = w_1, w_2, \cdots, w_N$ with N strokes, wherein the n-th stroke is denoted by w_n, we employ the corner detection algorithm [27] to perform character segmentation. As described in the paper [27], the bending value of point (x_i, y_i) is calculated by considering its k forward/backward points:

$$\beta = \frac{max(|x_{i+k} + x_{i-k} - 2x_i|, |y_{i+k} + y_{i-k} - 2y_i|)}{2k} \quad (8)$$

Essentially, the bending value of a certain point indicates the curvature of a local area of the character, and the maximum values will be treated as the corners. Now since we have found the corners, each stroke can then be split into segmentations; thus, we have $w_n = w_{n1}, w_{n2}, \cdots, w_{ns_n}$ with s_n segmentations. Typically, *DropSegment* is simply conducted by dropping a certain number of segments from the character. Suppose we drop d_n number of segments from the n-th stroke, the possible number of the remaining combinations of the segments should be:

$$C_{s_n}^{d_n} = \frac{s_n!}{(s_n - d_n)!d_n!} \quad (9)$$

Then, following the addition principle of combinatorics, we have the total possible remaining combinations of the segments $\sum_{d_n}^{s_n} C_{s_n}^{d_n}$ by considering all the possible d_n with $0 \leq d_n \leq s_n$. Now, we can calculate the number of new characters $N(\boldsymbol{w})$ by using *DropSegment* to the prototype character, which is presented as follows:

$$N(\boldsymbol{w}) = \prod_{n=1}^{N} \sum_{d_n=0}^{s_n} C_{s_n}^{d_n} \quad (10)$$

Note, however, that we cannot drop all the segments from the character, because then there would be insufficient information for identifying a certain writer. Given the above-mentioned analysis, we now use a simple example to show the advantage of data augmentation technique *DropSegment*. Suppose we have a character with four strokes, each stroke can be split into 2, 3, 4, and 5 segments, respectively. Then, the *DropSegment* data augmentation technique can generate a total number of $2^2 \times 2^3 \times 2^4 \times 2^5 - 1 = 16383$ characters; thus, *DropSegment* data augmentation technique can substantially increase the training sample numbers, and more importantly for improving the generalization of our system. In Fig. 4, we show some typical samples to illustrate the concept of *DropSegment*.

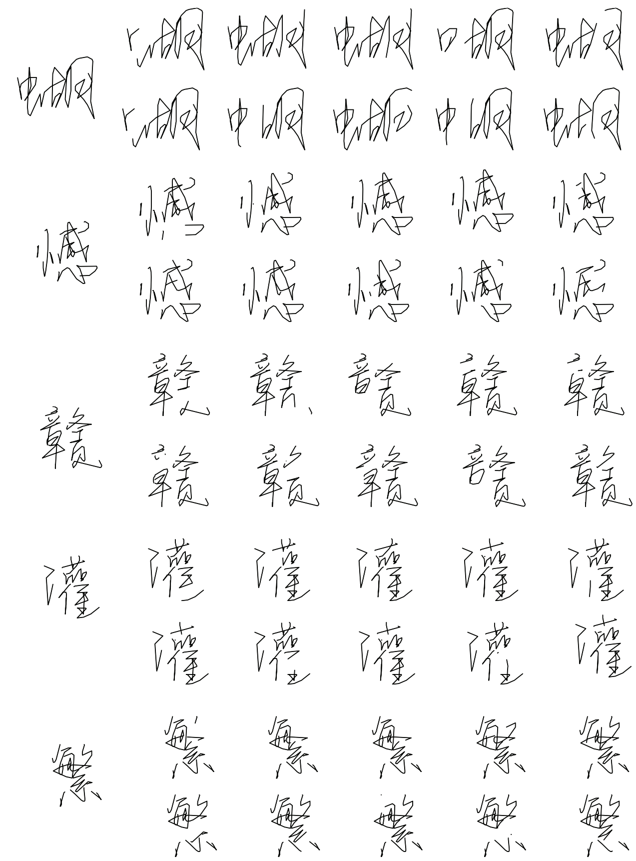

Fig. 4. Illustration of *DropSegment*. For each character, we show ten examples processed after *DropSegment*.

IV. Experimental Results

A. Database

The National Laboratory of Pattern Recognition (NLPR) handwriting database [28], which contains pages and very detailed information such as position, pen-down and pen-up state, azimuth, altitude and pressure, is not evaluated in this paper. This paper is intended at providing a general-purpose writer identification system that can be applied not only for page level writer identification but also at the level of individual characters. Therefore, we turn to the online handwritten Chinese character dataset CASIA-OLHWDB1.0 [29], which consists of 3866 classes from 420 writers. In our experiments, we follow the method of Yang et al [8] to split the dataset, resulting in a very small split of 200 classes for training, and the remaining 3666 classes are for testing, which require high expansibility of the writer identification system.

TABLE I
WRITER IDENTIFICATION RATES WITH THE PATH SIGNATURES IN DIFFERENT TRUNCATED VERSIONS.

Path signatures	feature dimension	Accuracy Rate
Δx,Δy	2	63.28
Sig1	4	64.54
Sig2	8	64.97
Sig2 (with *DropSegment*)	8	69.96

TABLE II
COMPARISON WITH STATE-OF-THE-ART METHODS BASED ON WRITER IDENTIFICATION RATES

Methods	Accuracy Rate
Zhang et al [15]	63.28
Our System	**64.97**
yang et al [8]	54.06
Our System	**68.84** (DropStroke)
Our System	**69.96** (DropSegment)

Upper: No data augmentation technique applied.
Lower: Data augmentation technique applied.

B. Experimental setting

As described in Section II, our network architecture is constructed by a path-signature layer, Deep BLSTM layer and collapsing layer. In this study, we fix the length of input string by 100, i.e., each character has a total point number of 100 with the spare space filled with zero and the extra points neglected. Note that, after applying data augmentation technique *DropSegment*, there are essentially very few samples with more than 100 points. In our baseline experiments, we simply used the raw data, i.e., Δx and Δy, just like Zhang et al [15]. Then, we applied path-signature layer to extract different iterated integrals of path-signature feature as well as Δx and Δy, which results in a feature dimension of 4, 8, and 16 for each local area. Our Deep BLSTM layer is constructed by two BLSTM layers with the output number set as 200, followed by a fully-connected layer with output number of 200. Then, the collapsing layer would sum over each dimension of the output of the deep BLSTM layer along the time axis, and make predictions for the 420 writers.

Writer identification is constructed within CAFFE [30] deep learning framework with LSTM implemented by Venugopalan et al [12] and the remaining contributed by us. Our optimization algorithm followed the typical stochastic gradient decent with a decay of 0.00005 and a clip-gradients of 10. The system is trained with GeForce Tian-X GPUs and it took approximately six hours to reach convergence.

C. Experimental Results

As listed in Table I, we compared the system with different iterated integrals of path signature. When the system only used the raw sequential data, Δx and Δy as input, the writer identification rate was 63.28%. After we applied the path-signature layer to extract signature feature for exploiting the essential online information for the pen-tip trajectories, the system performance consistently improved. However, we notice that the improvement of system performance slowed down when we used *sig2*. In conclusion, our PS-LSTM can integrate the advantage of both path signature and LSTM for sequential data learning, and can improve the writer identification rate. In the above experiments, we did not apply the data augmentation technique. After we applied *DropSegment* to enrich the training samples, we observed that the result improved with the absolute error rate reduced by approximately 5%. This phenomenon verifies the outstanding potential of data augmentation technique *DropSegment*, especially when training data is scarce or difficult to collect, e.g., Bank Transaction.

In Table II, we compare our method with state-of-the-art methods based on writer identification rates. Note that Zhang et al did not evaluate their methods on Database CASIA-OLHWDN1.0 and their data augmentation technique *RHS* is different from the data augmentation technique *DropSegment* we used in this paper. For a fair comparison, we implemented their LSTM-based method without *RHS* and compared the result with our system without *DropSegment*. As listed in Table II, the comparison demonstrates the outstanding capability of our PS-LSTM and the significance of the integration of path-signature and LSTM. Since Yang et al [8] applied data augmentation technique *DropStroke* to enrich the training samples for higher performance. For a fair comparison, we also used the same technique. By utilizing the advantages of the mechanism of *DropStroke*, Yang et al [8] can make not only one evaluation for the original test set with an accuracy rate of 54.06%, but 20 evaluations for the test set using *DropStroke* and average the probability distribution to get the final result with an accuracy rate of 55.45%. However, as listed in Table II, our result is superior to that of Yang et al [8] with an absolute error reduction of 14.78% in one test situation. Since both methods apply path-signature for feature extraction, the superiority of our writer identification system can be attributed to the application of LSTM. Note that when we apply the data augmentation technique *DropSegment*, the proposed writer identification system achieves better result of 69.96%.

V. CONCLUSION

In this paper, we propose a novel path-signature long short-term memory (PS-LSTM) recurrent neural network for writer identification. This is the first work to integrate path signature with Deep LSTM network and demonstrate its significance in the writer identification problem. In our experiments, we evaluate the effect of different iterated integrals of path signature and show that *sig2* can improve the accuracy rate of the deep LSTM system with an absolute error reduction of 1.69%, which indicates that by combining the outstanding capability of path signature and LSTM for path feature modeling, our PS-LSTM manages to capture more essential underlying dynamic sequential information of a path.

It is worthy noting that the proposed method now can not allow registration for new writers that are not included in the training set, which is worth studying for flexible practical application. Moreover, our PS-LSTM now only focuses on single-character recognition. How to combine context information between characters for writer identification deserve further study in the future work.

ACKNOWLEDGMENT

This research is supported in part by NSFC (Grant No.: 61472144, 61673182), the National Key Research & Development Plan of China (No. 2016YFB1001405), GDSTP (Grant No.: 2014A010103012, 2015B010101004, 2015B010131004) , GZSTP(No. 201607010227).

REFERENCES

[1] N. Sae-Bae and N. Memon, "Online signature verification on mobile devices," *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 6, pp. 933–947, 2014.

[2] M. Bulacu and L. Schomaker, "Text-independent writer identification and verification using textural and allographic features," *IEEE transactions on pattern analysis and machine intelligence*, vol. 29, no. 4, 2007.

[3] D. Guru and H. Prakash, "Online signature verification and recognition: An approach based on symbolic representation," *IEEE transactions on pattern analysis and machine intelligence*, vol. 31, no. 6, pp. 1059–1073, 2009.

[4] A. Schlapbach, M. Liwicki, and H. Bunke, "A writer identification system for on-line whiteboard data," *Pattern recognition*, vol. 41, no. 7, pp. 2381–2397, 2008.

[5] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[6] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[7] W. Yang, L. Jin, and M. Liu, "Deepwriterid: An end-to-end online text-independent writer identification system," *IEEE Intelligent Systems*, vol. 31, no. 2, pp. 45–53, 2016.

[8] ——, "Chinese character-level writer identification using path signature feature, dropstroke and deep cnn," in *Document Analysis and Recognition (ICDAR), 2015 13th International Conference on*. IEEE, 2015, pp. 546–550.

[9] A. Graves, *Supervised sequence labelling*. Springer, 2012.

[10] A. Graves, M. Liwicki, S. Fernández, R. Bertolami, H. Bunke, and J. Schmidhuber, "A novel connectionist system for unconstrained handwriting recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 5, pp. 855–868, 2009.

[11] M. Liwicki, A. Graves, H. Bunke, and J. Schmidhuber, "A novel approach to on-line handwriting recognition based on bidirectional long short-term memory networks," *Proc. 9th Int. Conf. on Document Analysis and Recognition*, vol. 1, pp. 367–371, 2007.

[12] S. Venugopalan, M. Rohrbach, J. Donahue, R. Mooney, T. Darrell, and K. Saenko, "Sequence to sequence-video to text," pp. 4534–4542, 2015.

[13] Y. Pan, T. Mei, T. Yao, H. Li, and Y. Rui, "Jointly modeling embedding and translation to bridge video and language," *CoRR*, vol. abs/1505.01861, 2015.

[14] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," *Advances in neural information processing systems*, pp. 3104–3112, 2014.

[15] X.-Y. Zhang, G.-S. Xie, C.-L. Liu, and Y. Bengio, "End-to-end online writer identification with recurrent neural network," *IEEE Transactions on Human-Machine Systems*, 2016.

[16] W. Yang, L. Jin, D. Tao, Z. Xie, and Z. Feng, "Dropsample: a new training method to enhance deep convolutional neural networks for large-scale unconstrained handwritten chinese character recognition," *Pattern Recognition*, vol. 58, pp. 190–203, 2016.

[17] T. Lyons and Z. Qian, "System control and rough paths,(2002)."

[18] T. Lyons, "Rough paths, signatures and the modelling of functions on streams," *CoRR*, vol. abs/1405.4537, 2014.

[19] B. Hambly and T. Lyons, "Uniqueness for the signature of a path of bounded variation and the reduced path group," *Annals of Mathematics*, pp. 109–167, 2010.

[20] B. Graham, "Sparse arrays of signatures for online character recognition," *CoRR*, vol. abs/1308.0371, 2013.

[21] W. Yang, L. Jin, Z. Xie, and Z. Feng, "Improved deep convolutional neural network for online handwritten chinese character recognition using domain-specific knowledge," *13th International Conference on Document Analysis and Recognition (ICDAR)*, pp. 551–555, 2015.

[22] Z. Xie, Z. Sun, L. Jin, Z. Feng, and S. Zhang, "Fully convolutional recurrent network for handwritten chinese text recognition," *CoRR*, vol. abs/1604.04953, 2016.

[23] K.-T. Chen, "Integration of paths–a faithful representation of paths by noncommutative formal power series," *Transactions of the American Mathematical Society*, vol. 89, no. 2, pp. 395–407, 1958.

[24] J. Donahue, L. Anne Hendricks, S. Guadarrama, M. Rohrbach, S. Venugopalan, K. Saenko, and T. Darrell, "Long-term recurrent convolutional networks for visual recognition and description," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2625–2634, 2015.

[25] K. Xu, J. Ba, R. Kiros, A. Courville, R. Salakhutdinov, R. Zemel, and Y. Bengio, "Show, attend and tell: Neural image caption generation with visual attention," *arXiv preprint arXiv:1502.03044*, 2015.

[26] R. Pascanu, C. Gulcehre, K. Cho, and Y. Bengio, "How to construct deep recurrent neural networks," *CoRR*, vol. abs/1312.6026, 2013.

[27] M.-J. J. Wang, W.-Y. Wu, L.-K. Huang, and D.-M. Wang, "Corner detection using bending value," *Pattern Recognition Letters*, vol. 16, no. 6, pp. 575–583, 1995.

[28] "Nlpr." NLPRdatabase.http://www.cbsr.ia.ac.cn/english/Databases.asp., the National Laboratory of Pattern Recognition (NLPR) handwriting database.

[29] C.-L. Liu, F. Yin, D.-H. Wang, and Q.-F. Wang, "Casia online and offline chinese handwriting databases," *2011 International Conference on Document Analysis and Recognition (ICDAR)*, pp. 37–41, 2011.

[30] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, "Caffe: Convolutional architecture for fast feature embedding," *Proceedings of the ACM International Conference on Multimedia*, pp. 675–678, 2014.