DRAWING ORDER RECOVERY FROM TRAJECTORY COMPONENTS

Minghao Yang[1*], *Xukang Zhou*[1,2*], *Yangchang Sun*[1], *Jinglong Chen*[2], *Baohua Qiang*[2]

[1]Research Center for Brain-inspired Intelligence, Institute of Automation, Chinese Academy of Sciences
[2] Guilin University Of Electronic Technology, China

ABSTRACT

In spite of widely discussed, drawing order recovery (DOR) from static images is still a great challenge task. Based on the idea that drawing trajectories are able to be recovered by connecting their trajectory components in correct orders, this work proposes a novel DOR method from static images. The method contains two steps: firstly, we adopt a convolution neural network (CNN) to predict the next possible drawing components, which is able to covert the components in images to their reasonable sequences. We denote this architecture as Im2Seq-CNN; secondly, considering possible errors exist in the reasonable sequences generated by the first step, we construct a sequence to order structure (Seq2Order) to adjust the sequences to the correct orders. The main contributions include: (1) the Img2Seq-CNN step considers DOR from components instead of traditional pixels one by one along trajectories, which contributes to static images to component sequences; (2) the Seq2Order step adopts image position codes instead of traditional points' coordinates in its encoder-decoder gated recurrent neural network (GRU-RNN). The proposed method is experienced on two well-known open handwriting databases, and yields robust and competitive results on handwriting DOR tasks compared to the state-of-arts.

Index Terms— Handwriting trajectory image, Drawing order recovery (DOR), Convolution neural network (CNN), encoder-decoder GRU

1. INTRODUCTION

Drawing order recovery (DOR) is widely discussed in various fields, including off-line to on-line handwriting recognition [1–4], robot based in-home learning companion for writing [5], video-based storytelling animation [6], etc. In spite of various methods proposed before [1, 3, 7, 8], etc., DOR from static images is still a great challenge.

The target and challenge of DOR are to predict the next location correctly at the travel beginning or cross points are ahead along the drawing trajectory [4, 9]. Various methods have been proposed to solve the challenge. In general, DOR task contains two categories: single-stroke DOR (ss-DOR) [2, 4] and multi-stroke DOR (ms-DOR) task [1]. The trajectories in ss-DOR mean that they are strictly continuous, which are generated by one stroke, while the trajectories in ms-DOR task contain multiple discontinuous components. In early years, knowledge based rules were usually used to locate the start points or determine the next movement direction when cross points are ahead, such as instance-based matching [7], double-traced lines check [10] methods. Until now, rules are still efficient in trajectory direction determination. However, rules based methods have their limits because it is impossible to list all traveling rules.

Graph model based global strategies and its varieties were widely discussed in order discovery, including graph based ambiguous zone analysis [1], optimal Euler path method [2], global energy minimization [4, 6], semi-Eulerian graph for double-traced lines detection [10], etc. These methods determine the start vertex and edges at first, then trace the graph from the start vertex combined with specified restrictions. They were suitable for ss-DOR tasks. While for the ms-DOR tasks, graph model is challenged when multiple possible next locations are not close to ones previously recovered.

For the task of multiple-stroke order decision, nearest neighbor principle [7], circle segmentation method [3], etc., were proposed to find the best next strokes from previous locations. However, most of these methods are still heuristic and they are limited with some constraints: for example, the drawing trajectories must have clear start and end points, or similar trajectories are needed in database, etc. These constraints limit the popularization and application of them.

Deep architecture has presented its high performance on order recovery, among which image based location prediction [11], sequence to sequence transfer [12, 13], attention model (AM) [14] are much related to the task of DOR. Convolution neural network (CNN) has been used to estimate human's writing orders pixel by pixel [9] and some researchers combined CNN and Long Short-Term Memory (LSTM) to a sequence generation method for the points sampled uniformly with a fixed number over the image trajectory [15]. These CNN and RNN (LSTM) methods modeled the image to sequence by learning all pixels orders [9] or full trajectory features [15]. However, how to search and gather the most relevant information from source sequences for target order from stroke components in static images is still not discussed.

*These authors made equal contributions to this work.

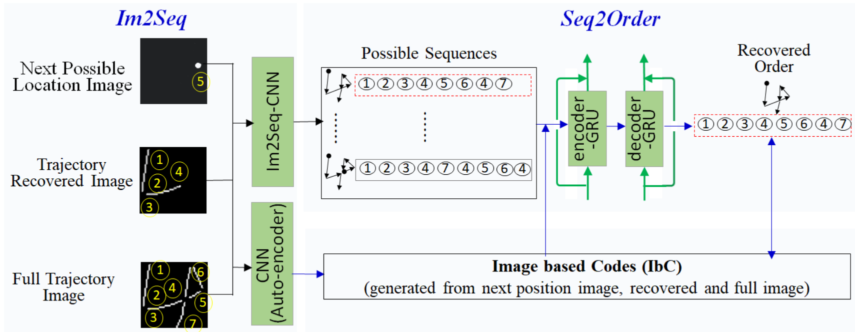

Fig. 1. The outline of the proposed DOR model

2. THE PROPOSED MODEL

In this work, we consider the DOR task as a reconstruction problem by connecting the trajectory components segmented at the cross points. The drawing trajectory images are easily converted into single-pixel wide image using the thinning method [16] and are conveniently converted into disconnected components at the cross points [17]. Following the ideas presented in [9, 18], we call the End Points (EPs) and Cross Points (CPs) on drawing trajectories as key points in this work. Once the key points' orders are determined, the drawing order is recoverd.

Definition: let Γ (ξ_1, ξ_2,...,ξ_i,..., ξ_M, $1 \leq i \leq M$) be a trajectory which length is M pixels and ξ_i is the i^{th} pixel in order on Γ. $Loc(\xi_i)$ = (x_i,y_i) is the coordinate of ξ_i in image and $Ord_{Loc(\xi_i)}$ is the correct order of $Loc(\xi_i)$, which is known in training phase and unknown in recovery phase. Supposing that Ψ (ψ_1, ψ_2,...,ψ_j,..., ψ_N, $1 \leq j \leq N$, $N << M$) is the key point set on Γ, and $Pos(\psi_j)$ = (u_j,v_j) is the coordinate of ψ_j in image. The function $Ord_{Pos(\psi_j)}$ = $h(j$,$Pos(\psi_j)$,$Loc(\xi_i)$,$Ord_{Loc(\xi_i)})$ return the correct order on Γ for each key point in Ψ by finding the index i according to $Pos(\psi_j)$=$Loc(\xi_i)$. Similar to $Ord_{Loc(\xi_i)}$, $Ord_{Pos(\psi_j)}$ is known in training phase and unknown in recovery phase. The task of DOR is defined as: for a static trajectory $\widehat{\Gamma}$ and its key points set $\widehat{\Psi}$ ($\widehat{\psi}_1$, $\widehat{\psi}_2$,...,$\widehat{\psi}_j$,..., $\widehat{\psi}_N$), to find the $Ord_{Pos(\widehat{\psi}_j)}$ ($1 \leq j \leq N$).

Fig. 1 outlines the framework of the proposed DOR model, which contains two steps: Im2Seq and Seq2Order. Im2Seq is used to predict the next possible drawing components and the Seq2Order is used to convert the reasonable sequences to the correct orders. In Fig. 1, we take a handwritten digital number "4" as an example. Supposing at current frame j, the previous stroke components ((1) $\rightarrow$ (2),(3) $\rightarrow$ (4)) have been correctly traveled, where the traveled image is presented as the "Trajectory Recovered Image" items in the left column of the Img2Seq block. In addition, the items in the left column of the Seq2Order area list all the possible sequences for the left four points ψ_j ($5 \leq j \leq 7$), where the ground truth sequence $Ord_{Pos(\psi_j)}$ ($1 \leq j \leq 7$) is given in red dot line box both on the top of the column of "Possible Sequences" and the output of "decoder-GRU" module.

2.1. Im2Seq: component sequences generation from static Images

The tasks of transfer the pixels on images to their orders have been well discussed [4]. Among these work, deep architectures have presented its high performances [9]. Inspired by the DEN-CNN idea presented in [9], we use it as the Img2Seq-CNN in this work.

2.2. Seq2Order: adjust component sequences to order

Im2Seq step learns the next possible locations from images and converts key points' locations to reasonable sequences. Considering possible errors that exist in the reasonable sequences generated by Im2Seq step, we continually propose a Seq2Order attention model to adjust these possible error sequences to correct orders. Encoder-decoder LSTM (Long Short Term Memory) has been used to learn the probability of distribution of successive coordinate points for every stepwise step from the static image encoded by CNN [15]. In this work, we adopt Encoder-decoder gated recurrent units (GRU) RNN in the Seq2Order step. The proposed Seq2Order model is different from the ones proposed in [15, 19]: (1) the encoder-GRU uses the three channels' images (full trajectory, trajectory traveled, position) information instead of single image codes [15]; (2) the item length of sequences are dynamically decided by the number of the key points instead of fixed input number adopt in [18, 19].

Fig. 2 outlines the proposed encoder-decoder GRU Seq2Order in this work. The hidden layer features of GRU cells obtained from three images are used to calculated GRU weights, where the three images are next position image, recovered and full image, which we called as image based codes (IbC) attention model in this work. The workflow of decoder-GRU is presented as blue arrows in the right and bottom parts of Fig. 1 and the red lines between the hidden layer of encoder-GRU and decoder-GRU in Fig. 2. The target of encoder-decoder

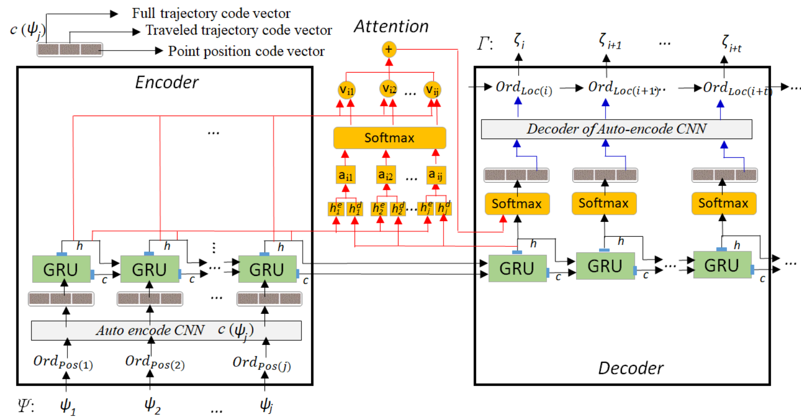

Fig. 2. The outline of the proposed encoder-decoder GRU in Seq2Order and IbCAM.

GRU is to connect the sequence ψ_j $(1 \leq j \leq N)$ generated by Im2Seq and the item ξ_i or $Ord_{Loc(\xi_i)}$ coded by IbC model. In Fig. 2, h_j^e and h_i^d are the hidden features for the j^{th} encoder-GRU and the i^{th} decoder-GRU respectively. a_{ij} and v_{ij} are the parameters and the feature of encoder-decoder GRU respectively.

3. EXPERIMENTS

3.1. Handwriting Drawing order recovery

3.1.1. *Experiment setting*

We use UNIPEN [20] for English and digit ss-DOR task and OLHWDB 1.1 [21] for Chinese ms-DOR task. These two datasets are widely recognized online handwriting database. Being similar to [9, 22], all images are normalized into 48×48 resolution, and the strcutrue of Img2Seq-CNN is similar to the DEN proposed in [9] and we adopt the structure proposed in [23] as the GRU-RNN in Seq2Order step using MindSpore framework [24].

3.1.2. *DOR for English and digit handwriting*

UNIPEN includes 708,881 static English and digit images. In the experiments, 500,000 random images were used in training, and other 100,000 images unseen before were used in evaluation. As far as we know, there was no DOR work especially proposed for English letters and Arab digital symbols since [2, 4, 25]. Therefore, the excellent results obtained by them could be still viewed as the state-of-the-art.

Table 1 presents the comparison results for the randomly picked 100,000 English letters and numbers in UNIPEN. Being similar to previous work [9], we compute the Dynamic Time Warping (DTW) between our results and standard writing orders to evaluate the performance of English and digit handwriting DOR. We can see from Table 1 that the proposed DOR models (Im2Seq + Seq2Order + PbC/IbC) outperforms the results given by previous researchers. "PbC" means the Seq2Order step adopts the key points' coordinates as codes in encoder-decoder GRU, like the strategy adopted in [15].

Table 1. ss-DOR Results for English handwriting

Method	**Single-stroke restriction?**	
	NO	**YES**
Kato[2]	/	91.6
Qiao[25]	/	93.7
Nishiara[4]	/	96.0
Iwakiri[7]	51.7	/
Zhao[9]	95.5	98.1
Im2Seq+Seq2Order	95.3	97.9
Im2Seq+Seq2Order+PbC	95.7	98.2
Im2Seq+Seq2Order+IbC	**95.8**	**98.4**

*Numbers in the table are prediction accuracy and measurement is percentage (%)

3.1.3. *Multi-stroke DOR for Chinese characters*

We continually use OLHWDB 1.1 [21] in Chinese ms-DOR task. OLHWDB 1.1 includes 3755 Chinese characters written by 240 users. In the experiments, we choose the most η commonly used Chinese characters categories [26] written by 120 different writers as the training set. Then the size of Im2Seq training data is $\sum_{\Gamma=1}^{\eta} 120 \times m_\Gamma \times m_\Gamma$, where m_Γ is the number of key points in a character. Then it is about 240,000 and 614,400 images for η=80 and m_Γ is 5 and 8 respectively.

We randomly choose other 4000 Chinese characters from about 100 categories written by 100 users from OLHWDB 1.1 as ms-DOR test set. For simplification, we denote these five kinds of characters as five levels: Cl.i (1≤i≤5). Chinese character maintains less than 5 strokes are configured Cl.1 and upgrades every 5 more strokes. We compute the Dynamic Time Warping (DTW) [9] between our results and standard writing orders to evaluate the performance of ms-DOR.

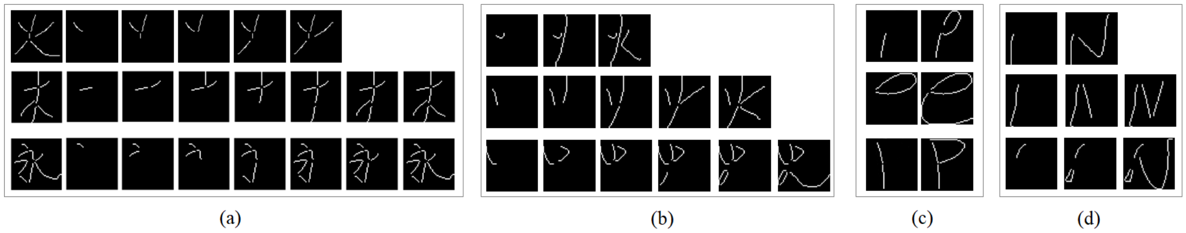

Fig. 3. Handwriting DOR for some English and Chinese characters written by unknown users which have similar structures (all junction points are removed in the images, which contributes to the key points along the stroke components).

Table 2. ms-DOR Results for Chinese handwriting

$Cl.i$	$Cl.1$	$Cl.2$	$Cl.3$	$Cl.4$	$Cl.5$
$\phi(Cl.i)$	1156	2428	313	91	12
Yang[22]	65.5	49.1	33.1	12.6	5.8
Zhao[9]	83.1	68.2	50.5	35.6	23.3
Im2Seq+Seq2Order	83.9	70.8	51.7	36.5	23.9
Im2Seq+Seq2Order+PbC	84.8	71.2	52.6	37.2	23.4
Im2Seq+Seq2Order+IbC	**85.3**	**72.1**	**53.9**	**38.6**	**24.8**

*Numbers in the table are prediction accuracy and measurement is percentage (%).

Table 2 lists the ms-DOR DTW results of the proposed DOR for Cl.1 to Cl.5, where Φ(Cl.i) is the characters' number contained in set Cl.i. We can see from Table 2 the proposed DOR model outperforms [9] on ms-DOR accuracy at least 2.2% and outperforms [22] at least 19.8% for Cl.1. For other levels of Chinese characters, the proposed DOR model also yields obviously better accuracies compared with [9, 22] on the same Chinese handwriting dataset OLHWDB 1.1.

Using the results given by [9] as a baseline, we continue to discuss the performance of each step in DOR. The models, incuding Im2Seq + Seq2Order + PbC/IbC obtain obvious higher performances than those of [22] and baseline. The ms-DOR accuracies of Im2Seq + Seq2Order and Im2Seq+Seq2Order+IbC gradually increase 0.8%, 2.2% on Cl.1 and 2.6%, 3.9% on Cl.2 respectively from SEN+DEN [9]. As for other levels of Chinese characters, the proposed imTOR model also yields obviously better DOR accuracies compared with [9]. And the proposed IbC model outperforms the key points' coordinates model (PbC) high about 0.9% in Cl.2 and 1.4% in Cl.4. It demonstrates that with the connection of Seq2Order and the proposed IC module in the DOR, the Seq2Order and IC module contribute to performance improvement. In addition, it is noticeable that with the increasing of characters' complex degree from Cl.1 to Cl.5, DOR accuracies present exaggeratedly decay reduction from Cl.1 to Cl.5. We think a hidden reason behind this phenomenon is: the number of possible writing orders increases in the power exponent level with the increasing of strokes number, which increases the challenge for DOR to choose the only one correct order from exaggeratedly increased candidates.

Fig. 3(a)(b)(c) presents several DOR results for some multi-stroke Chinese characters, Greek letter "ρ", English letters "e" and "p" written by unknown users. We can see from Fig. 3(a)(b)(c) that although they share similar structures, the proposed model could recover the drawing progresses for these letters correctly. Fig. 3(d) presents several DOR results for unlearned English letter "N" written by different unknown users. It demonstrates that the proposed algorithm owns the ability to infer the drawing orders for the characters which are unknown before.

4. CONCLUSIONS

This work proposed a handwriting drawing order recovery method from their static image presentation. As far as we know, the proposed DOR method is the first one to regard DOR problem as a trajectory components sequences estimation task. And the proposed DOR model is able to consider single-stroke and multiple-stroke DOR tasks in a uniform framework and achieves satisfactory performance compared with the state-of-the-art methods. Despite competitive results obtained, the proposed method still needs to improve its performance, especially for those complex Chinese characters with 10 or more stroke numbers in writing.

5. ACKNOWLEDGEMENTS

This work is supported by the National Key Research & Development Program of China (No. 2018AAA0102902), the National Natural Science Foundation of China (NSFC) (No.61873269), the Beijing Natural Science Foundation (No: L192005), the CAAI-Huawei MindSpore Open Fund (CAAIXSJLJJ-20202-027A), the Guangxi Key Research and Development Program (AD19110137, AB18126063), the Natural Science Foundation of Guangxi of China (2019-GXNSFDA185006).

References

[1] Guee Sang Lee Minh Dinh, Hyung Jeong Yang, "Recovery of drawing order from multi-stroke english handwritten images based on graph models and ambiguous zone analysis," *Expert Systems with Applications*, vol. 64, pp. 352–364, 2016.

[2] Yoshiharu Kato and Makoto Yasuhara, "Recovery of drawing order from single-stroke handwriting images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 9, pp. 938–949, 2000.

[3] Eric Anquetil Rousseau Laetitia and Jean Camillerapp, "Recovery of a drawing order from off-line isolated letters dedicated to on-line recognition," *International Conference on Document Analysis and Recognition*, vol. 2, pp. 1121–1125, 2005.

[4] Mikihiko Nishiara Qiao Yu and Makoto Yasuhara, "A framework toward restoration of writing order from single-stroked handwriting image," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 11, pp. 1724–1737, 2006.

[5] Wu Na Tan Huan, Du Qian, "Robots learn writing," *Journal: Journal of Robotics*.

[6] Ligang Liu Hongbo Fu, Shizhe Zhou and Niloy J. Mitra, "Animated construction of line drawings," *ACM Transactions on Graphics*, vol. 30, no. 6, pp. 1–10, 2011.

[7] Yaokai Feng Yutaro Iwakiri, Soma Shiraishi, "On the possibility of instance-based stroke recovery," *IEEE International Conference on Frontiers in Handwriting Recognition*.

[8] Makoto Yasuhara Yu Qiao, "Recovering drawing order from offline handwritten image using direction context and optimal euler path," *IEEE International Conference on Acoustics, Speech and Signal Processing*.

[9] Minghao Yang Bocheng Zhao and Jianhua Tao, "Drawing order recovery for handwriting chinese characters," *International Conference on Acoustic, Speech, and Signal Processing (ICASSP)*, 2019.

[10] Takayuki Nagoya and Hiroyuki Fujioka, "Recovering human-like drawing order from static handwritten images with double-traced lines," *Lecture Notes in Electrical Engineering*, vol. 253, pp. 941–948, 2013.

[11] Douglas Eck David Ha, "A neural representation of sketch drawings," *ArXiv e-prints*, 2017.

[12] Trisha Mittal R. Venkatesh Babu Santosh Ravi Kiran Sarvadevabhatla, Shiv Surya, "Game of sketches: Deep recurrent models of pictionary-style word guessing," *Thirty-Second AAAI Conference on Artificial Intelligence(AAAI)*, February 2018.

[13] Le Quoc V Sutskever Ilya, Vinyals Oriol, "Sequence to sequence learning with neural networks," *Neural information processing systems*, pp. 3104–3112, 2014.

[14] KyungHyun Cho Bahdanau Dzmitry and Yoshua Bengio, "Neural machine translation by jointly learning to align and translate," *The 3rd International Conference on Learning Representations, ICLR 2015*, 2015.

[15] Ankan Kumar Bhunia Aishik Konwer Prithaj Banerjee Partha Pratim Roy Umapada Pal Ayan Kumar Bhunia, Abir Bhowmick, "Handwriting trajectory recovery using end-to-end deep encoder-decoder network," *24th International Conference on Pattern Recognition (ICPR*, pp. 3639–3644, 2018.

[16] A. Krzy.zak Balazs Kegl, "Piecewise linear skeletonization using principal curves," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 1, pp. 1–74, 2002.

[17] Sabine Barrat Jean-Yves Ramel Anh Pham, Mathieu Delalandre, "Accurate junction detection and characterization in line-drawing images," *Pattern Recognition*, vol. 47, no. 1, pp. 282–295, 2013.

[18] Minghao Yang Zhengkun Tian Cunhang Fan Bocheng Zhao, Jianhua Tao and Ye Bai, "Deep imitator: handwriting calligraphy imitation via deep attention networks," *Pattern Recognition*, vol. 104, 2020.

[19] Chenglin Liu Xuyao Zhang, Yin Fei, "Drawing and recognizing chinese characters with recurrent neural network," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 4, pp. 849–862, 2018.

[20] EH Ratzlaff, "Methods, reports and survey for the comparison of diverse isolated character recognition results on the unipen database," *IEEE Seventh International Conference on Document Analysis and Recognition (ICDAR)*, pp. 623–628, 2003.

[21] Da-Han Wang Qiu-Feng Wang Cheng-Lin Liu, Fei Yin, "Casia online and offline chinese handwriting databases," *International Conference on Document Analysis and Recognition*, pp. 37–41, Sept 2011.

[22] Jianhua Tao Bocheng Zhao, Minghao Yang, "Pen tip motion prediction for handwriting drawing order recovery using deep neural network," *24th International Conference on Pattern Recognition (ICPR)*, 2018.

[23] YanMing Zhang Zhang XuYao, Fei Yin and ChengLin Liu, "Drawing and recognizing chinese characters with recurrent neural network," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 4, pp. 849–862, 2018.

[24] "https://www.mindspore.cn/install/en," *Acces date: Feb 10, 2021*.

[25] Qiao Yu and Makoto Yasuhara, "Recovering drawing order from offline handwritten image using direction context and optimal euler path," *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2006.

[26] SR Kao Henry DingGuo Gao, "Psycho-geometric analysis of commonly used chinese characters," *Cognitive neuroscience studies of the Chinese language*, pp. 195–206, 2002.