Word-level Stroke Trajectory Recovery for Handwriting with Gaussian Dynamic Time Warping

Anonymous ACL submission

Abstract

Handwriting trajectory recovery has recently gained more attention for practical applications such as personalized messages. It is a sequence learning problem from image to handwriting stroke sequence where Dynamic Time Warping (DTW) is a preferred loss function. However, aligning two varying length sequences in DTW loss accumulates the differences of predicted and ground truth strokes for the entire line-level text. As a result, averaging over long sequences in DTW loss, it cannot distinguish between a small number of perceptually significant errors and a large number of visually insignificant errors. To address this issue, we propose two new strategies. First, we propose applying DTW to words instead of line-level text so that the DTW loss for all the words in the line-level text is not averaged out. Moreover, for aligning the predicted and ground-truth sequences for each word, we propose to weight the cost matrix with a Gaussian function so that the far-off predicted strokes from ground truth are penalized heavily. This strategy for word-level stroke trajectory learning improves quantitative and qualitative results.

1 Introduction

Handwriting stroke trajectory recovery from static images is of utmost importance to revolutionize the applications such as personalized message writing on letters or greeting cards, signature verification, and script handwriting learning. The earliest work on handwriting stroke recovery started before the deep learning boom, which utilized handcrafted local and global features with the taxonomy of clues to recover the handwriting trajectory for each alphabet letter (Doermann and Rosenfeld, 1995; Viard-Gaudin et al., 2005a). (Abuhaiba et al., 1998; Viard-Gaudin et al., 2005b) used a semantic rules-based approach for sub-words with a graph traversal to reconstruct stroke trajectory for handwriting recognition. Nevertheless, they considered only alphabets to learn the trajectory of the stroke. (Privitera and Plamondon, 1995) recovered the trajectory information for handwriting by segmenting and dividing the words into a temporal sequence of strokes. Above mentioned researches exhibit limited application for recently introduced handwriting datasets.

Stroke trajectory recovery has made progress towards more realistic and complex handwriting datasets using deep neural networks in recent years. (Bhunia et al., 2018) introduced the first trainable convolution network for stroke trajectory recovery. This LSTM architecture learns strokes with Euclidean distance loss, making it hard to apply on long words with multiple strokes. Moreover, (Moussa et al., 2021) added a CNN before LSTM to recover the stroke trajectory of the handwriting in images. However, this work is limited to stroke learning for mathematical equations, and in the current form, it is not being applied to words in the English language.

The most recent work related to stroke trajectory recovery is presented by (Archibald et al., 2021), where LSTM is trained with a Dynamic Time Warping (DTW) loss function. They also introduced adaptive ground truths to make stroke ordering more flexible during training. (Nguyen et al., 2021) employed an LSTM architecture with an attention layer and Gaussian Mixture Model (GMM) trained with cross-entropy loss, but it learns to encode only a single Japanese alphabet.

All these architectures either use line of text (Archibald et al., 2021; Bhunia et al., 2018) or alphabet letter (Nguyen et al., 2021; Viard-Gaudin et al., 2005b; Privitera and Plamondon, 1995), but to the best of our knowledge, the stroke trajectory recovery network for words has not yet been proposed.

Moreover, we propose to compute the warping path during the alignment of predicted and ground truth sequences in DTW with Gaussian weighting.

In this way, we penalize the warping path heavily if the predicted stroke is far-off the ground truth stroke as it adds a perceptually significant error in stroke trajectory recovery. Whereas, the predicted stroke points in the close vicinity of the ground truth are perceptually indistinguishable from original strokes. The Gaussian function for DTW has been used for time series classification (Jeong et al., 2011) based on the phase difference between two time series, but its potential advantage for stroke trajectory recovery has not been explored before. The main contributions of this work are as follows: 1) A word-level handwriting stroke trajectory recovery method is proposed. It estimates loss for each word rather than averaging DTW loss over the entire line-level text. 2) To better match the human visual perception of handwriting, we employ a Gaussian weighted cost matrix in DTW to generate a loss function for deep learning. It allows our network to tolerate minor deviations in aligning the predicted and ground truth strokes while penalizing large, easily noticeable deviations. 3) Our quantitative and qualitative results demonstrate the superior performance of our approach in comparison to the state-of-the-art (SOTA).

We introduce the method in Sec. 2 and demonstrate the experimental results in Sec. 3.

2 Method

In our work, we introduced two levels of granularity to learn the stroke trajectory for handwriting, the first is dividing a line-level text into words, and the second is to use the Gaussian function to weigh the cost matrix in DTW loss for each word.

2.1 Word-level datasets

IAM-online datasets (Marti and Bunke, 2002) consists of line-level text with stroke ground truth information. To the best of our knowledge, the previous researches (Archibald et al., 2021) for handwriting stroke trajectory recovery considered the text lines as input. The disadvantage of using text lines is the averaging out of loss function for all the words in the line. However, some words have a structure that is harder to learn (such as *stage*) than the less complex words such a *the*. Therefore, in our work, we propose to break the text lines into words to calculate DTW loss for each word. For this purpose, the strokes are divided into words for train and test sets.

For this, we use a simple rule defined below. Let the stroke sequence S be composed of strokes as $[s_1, s_2, s_3..., s_n]$. Stroke s_{i+1} merges with the previous stroke s_i if the following set of conditions are obeyed.

$$\begin{cases} M(s_{i+1}, s_i), & \text{if } \wedge(s_{i+1}) \geq \vee(s_i) \\ M(s_{i+1}, s_i), & \text{elif } (\vee(s_i) - \wedge(s_i + 1)) \geq th \\ Sep(s_{i+1}, s_i), & \text{otherwise} \end{cases} \tag{1}$$

Where the symbol $\wedge(s_{i+1})$ and $\vee(s_i)$ represents the minimum x-coordinate for stroke s_{i+1} and the maximum x-coordinate for stroke s_i respectively. M and Sep stand for merge or separate stroke function. We merge the strokes if the later stroke in S has already started before ending the previous stroke or the distance between the two strokes is less than the threshold (th). The value of th is different for each line. It is calculated based on the average stroke's spacing in each text line. Therefore it is based on handwriting style.

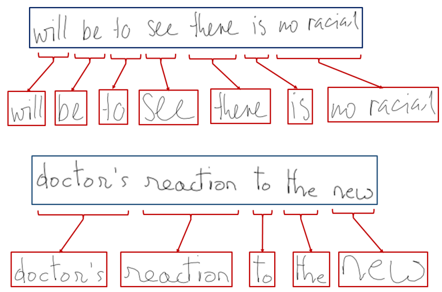

Figure 1: Sample of the word-level IAM-online datasets we created.

Figure 1 shows a reasonably separated words from line-level datasets into word level datasets.

2.2 Network architecture

Our architecture uses a CNN (seven convolutional blocks with ReLU) and LSTM layer. Convolutional filters have a 3x3 kernel size with 2x2 and 2x1 max pooling in each layer. Moreover, the input for the first convolutional block has a fixed height, and variable-width similar to (Bhunia et al., 2018; Archibald et al., 2021) in order to facilitate the processing of variable-length words for different handwriting styles. The block diagram of overall architecture is shown in Figure 2.

2.3 Loss function

In the next section, we introduce a Gaussian weighted cost matrix in DTW loss that emphasizes avoiding the costly alignment of far-off points in loss computation $\mathcal{L}_{DTW_G}$.

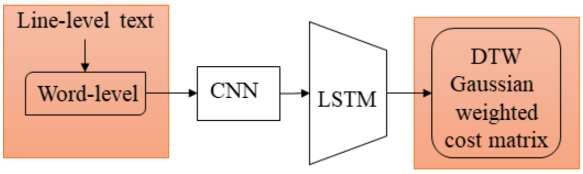

Figure 2: The block diagram of our proposed architecture, the modules in orange highlight our contribution.

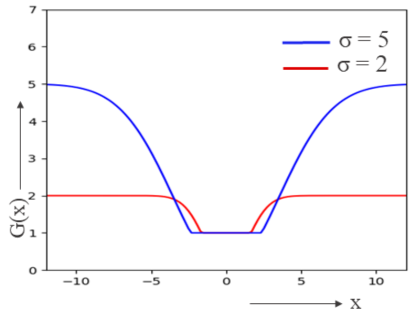

Figure 3: Gaussian function G used to calculate DTW alignment between GT T and predicted P strokes.

2.3.1 Gaussian weighted cost matrix in DTW

In general, DTW (Berndt and Clifford, 1994; Choi et al., 2020) computes the optimal match between GT $T = (t_1, t_2, t_3, ...t_m)$ and predicted sequences $P = (p_1, p_2, p_3, ...p_n)$ of different lengths by finding the warping path between two sequences. In DTW loss, cost matrix A calculates the distance between all the stroke points in P and T to find the optimal warping path that is used to align the two sequences.

In (Archibald et al., 2021), the cumulative cost matrix A at the i^{th} stroke point of P and j^{th} stroke point of T is calculated by their squared Euclidean distance. In our work, we propose to weight the distance ($||p_i - t_j||^2$) by Gaussian function G as:

$$G(||p_i - t_j||) \cdot ||p_i - t_j||^2. \tag{2}$$

To define G, we start with

$$H(||p_i - t_j||) = \sigma \left(1 - e^{-\left(\frac{||p_i - t_j||}{\sigma}\right)^2}\right), \tag{3}$$

where σ is a constant related to Gaussian standard deviation. According to Eq. 3, H ranges from 0 to σ. To define G, we clip the value of H at 1 as follows:

$$G(x) = \begin{cases} H(x) & \text{if } H(x) > 1 \\ 1, & \text{else} \end{cases} \tag{4}$$

Fig. 3 shows the visualization of the Gaussian function G used in our cost matrix. We evaluated our method for $\sigma = 2$ and $\sigma = 5$.

The Gaussian function G directly affects the cumulative cost matrix A, where the $(i, j)^{th}$ entity of A is given as:

$$A(i, j) = G(||p_i - t_j||) \cdot ||p_i - t_j||^2 + \\ min[A(i-1, j), A(i-1, j-1), A(i, j-1)] \tag{5}$$

for $1 \leq i \leq m$ and $1 \leq j \leq n$. Given the cumulative cost matrix A, DTW computes the optimal warping path from $A(m, n)$ to $A(1, 1)$ as the alignment of points in P to points in T is expressed as index mapping $\alpha : \{1, \ldots, m\} \rightarrow \{1, \ldots, n\}$, where α is an onto function. Finally, the Gaussian weighted DTW loss is given by alignment α:

$$\mathcal{L}_{DTW_G}(P, T) = \sum_{i=1}^{m} ||p_i - t_{\alpha(i)}||. \tag{6}$$

The next section details the evaluation of handwriting stroke trajectory and the effect of the Gaussian function.

3 Experimental evaluation

3.1 Data

In the IAM-online dataset (Marti and Bunke, 2002), we have 10,927 annotations for line-level text with strokes information, out of which 7,402 are training, and 3,525 are testing text lines. After splitting text lines into words using the proposed word-level algorithm as described in Section 2.1, the size of training and testing datasets increase to 36,106 and 17,087 words, respectively. Figure 1 shows the sample images of the word-level dataset proposed in our work.

3.2 Evaluation Metrics

We use the same evaluation metric as (Archibald et al., 2021). It considers the percentage of predicted stroke points farther than T_0 pixels from their nearest GT denoted by $\%N_{t,p}$, and similarly, the percentage of GT stroke points father from their nearest predicted stroke denoted by $\%N_{p,t}$. The average distance of points in $\%N_{t,p}$ and $\%N_{p,t}$ is denoted by $dist_{t,p}$ and $dist_{p,t}$. To add the holistic view of GT and predicted sequence matching, we also evaluated our method for DTW distance D_{DTW} between GT and predicted strokes.

Distance metric		% $N_{t,p}$		$dist_{t,p}$		% $N_{p,t}$		$dist_{p,t}$	
		$T_0 = 5$	$T_0 = 2$	$T_0 = 5$	$T_0 = 2$	$T_0 = 5$	$T_0 = 2$	$T_0 = 5$	$T_0 = 2$
line-level DTW (Archibald et al., 2021)		0.0090	0.0258	0.3720	0.4745	0.0175	0.2790	0.0625	0.5581
word-level DTW		0.0049	0.0217	**0.1758**	0.3001	0.0108	0.1497	0.0395	0.7240
word-level DTW_G	$\sigma = 2$	0.0046	0.0183	0.1840	**0.2780**	0.0018	0.1442	0.0418	**0.3109**
	$\sigma = 5$	**0.0040**	**0.0167**	0.1828	0.2943	**0.0010**	**0.1429**	**0.0113**	0.4064

Table 1: Quantitative comparison of line-level, word-level and word-level DTW with Gaussian weighted cost matrix.

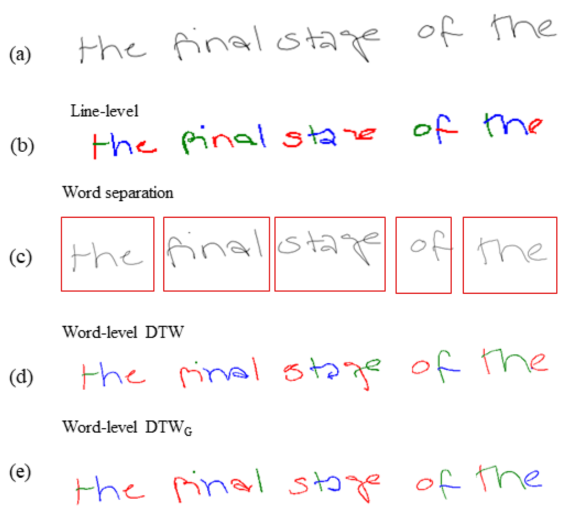

Figure 4: The visual quality of stroke recovery: (a) original handwriting, (b) line-level DTW recovery, (c) word level separation, (d) word-level DTW recovery, (e) proposed word-level DTW_G recovery for $\sigma = 5$. Each stroke is shown in different color (red, blue or green).

3.3 Results

The previous methods on IAM-online datasets work with line-level text for stroke trajectory recovery (Archibald et al., 2021). We initialize with the pre-trained model on the line-level text and fine-tune it for a word-level dataset with the Gaussian cost matrix in DTW.

Table 1 presents quantitative comparison, where bold numbers show the best results (lowest value). We first observe that most metrics are much lower for word-level handwriting than the line-level input. These results show that separating the strokes from line-level text into words using flexible criteria for different handwriting improves the results compared to the line-level datasets. Furthermore, the addition of Gaussian weighting in the cost matrix in DTW loss (*word-level* DTW_G) gives the lowest values for $\%N_{p,t}$ and $dist_{p,t}$, which mean that the predicted strokes are better imitating the GT strokes.

Method	line-level	word level	word-level DTW_G $\sigma = 2$	word-level DTW_G $\sigma = 5$
D_{DTW}	1.4058	1.1394	1.1258	**1.0987**

Table 2: DTW distance (D_{DTW}) between GT and predicted strokes

We also validated our method for different values of variance ($\sigma = 2$ and $\sigma = 5$) in Gaussian function as shown in Figure 3. We do not go beyond $\sigma = 5$, since $\sigma = 5$ incurs sufficiently large penalty for far-off points as shown in Figure 3. Table 1 shows the quantitative results for $\sigma = 2$ and $\sigma = 5$. Gaussian functions with variance $\sigma = 2$ and $\sigma = 5$ have very close performance but $\sigma = 5$ have slightly better results.

We also evaluate our method for DTW distance metric (D_{DTW}) as it gives us the holistic view on the resemblance of predicted and GT sequence. As given in Table 2, for line-level D_{DTW} is 1.4058 and for word-level D_{DTW} is 1.1394 respectively. Whereas for Gaussian cost matrix in DTW, the D_{DTW} is 1.1258 and 1.0987 for $\sigma = 2$ and $\sigma = 5$, respectively.

All the results demonstrate that the proposed Gaussian weighed cost matrix for DTW on word-level datasets outperforms the DTW loss for handwriting stroke recovery.

The visualization of recovered strokes in Figure 4 also shows that the proposed method (*word-level* DTW_G) gives better results than the line-level and word-level DTW.

4 Conclusion

The proposed method for word-level stroke trajectory learning with Gaussian weighted DTW loss improves quantitative and qualitative results.

References

Ibrahim SI Abuhaiba, Murray JJ Holt, and S Datta. 1998. Recognition of off-line cursive handwriting. *Computer Vision and Image Understanding*, 71(1):19–38.

Taylor Archibald, Mason Poggemann, Aaron Chan, and Tony Martinez. 2021. Trace: A differentiable approach to line-level stroke recovery for offline handwritten text. *arXiv preprint arXiv:2105.11559*.

Donald J Berndt and James Clifford. 1994. Using dynamic time warping to find patterns in time series. In *KDD workshop*, volume 10, pages 359–370. Seattle, WA, USA:.

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

Wonyoung Choi, Jaechan Cho, Seongjoo Lee, and Yunho Jung. 2020. Fast constrained dynamic time warping for similarity measure of time series data. *IEEE Access*, 8:222841–222858.

David S Doermann and Azriel Rosenfeld. 1995. Recovery of temporal information from static images of handwriting. *International Journal of Computer Vision*, 15(1-2):143–164.

Young-Seon Jeong, Myong K Jeong, and Olufemi A Omitaomu. 2011. Weighted dynamic time warping for time series classification. *Pattern recognition*, 44(9):2231–2240.

U-V Marti and Horst Bunke. 2002. The iam-database: an english sentence database for offline handwriting recognition. *International Journal on Document Analysis and Recognition*, 5(1):39–46.

Elmokhtar Moussa, Thibault Lelore, and Harold Mouchère. 2021. Applying end-to-end trainable approach on stroke extraction in handwritten math expressions images. In *ICDAR 2021: 16th International Conference on Document Analysis and Recognition*.

Hung Tuan Nguyen, Tsubasa Nakamura, Cuong Tuan Nguyen, and Masaki Nakawaga. 2021. Online trajectory recovery from offline handwritten japanese kanji characters of multiple strokes. In *2020 25th International Conference on Pattern Recognition (ICPR)*, pages 8320–8327. IEEE.

Claudio M Privitera and Réjean Plamondon. 1995. A system for scanning and segmenting cursively handwritten words into basic strokes. In *Proceedings of 3rd International Conference on Document Analysis and Recognition*, volume 2, pages 1047–1050. IEEE.

Christian Viard-Gaudin, Pierre-Michel Lallican, and Stefan Knerr. 2005a. Recognition-directed recovering of temporal information from handwriting images. *Pattern Recognition Letters*, 26(16):2537–2548.

Christian Viard-Gaudin, Pierre-Michel Lallican, and Stefan Knerr. 2005b. Recognition-directed recovering of temporal information from handwriting images. *Pattern Recognition Letters*, 26(16):2537–2548.