

Strokes Trajectory Recovery for Unconstrained Handwritten Documents with Automatic Evaluation

Sidra Hanif^a and Longin Jan Latecki^b

Temple University, Philadelphia, PA, U.S.A.

Keywords: Handwriting Strokes Trajectory Recovery, Dynamic Time Warping, Chamfer Distance, LSTM, Word Detection, Automatic Evaluation.


Abstract: The focus of this paper is offline handwriting Stroke Trajectory Recovery (STR), which facilitates the tasks such as handwriting recognition and synthesis. The input is an image of handwritten text, and the output is a stroke trajectory, where each stroke is a sequence of 2D point coordinates. Usually, Dynamic Time Warping (DTW) or Euclidean distance-based loss function is employed to train the STR network. In DTW loss calculation, the predicted and ground-truth stroke sequences are aligned, and their differences are accumulated. The DTW loss penalizes the alignment of far-off points proportional to their distance. As a result, DTW loss incurs a small penalty if the predicted stroke sequence is aligned to the ground truth stroke sequence but includes stray points/ artifacts away from ground truth points. To address this issue, we propose to compute a marginal Chamfer distance between the predicted and the ground truth point sets to penalize the stray points more heavily. Our experiments show that the loss penalty incurred by complementing DTW with the marginal Chamfer distance gives better results for learning STR. We also propose an evaluation method for STR cases where ground truth stroke points are unavailable. We digitalize the predicted stroke points by rendering the stroke trajectory as an image and measuring the image similarity between the input handwriting image and the rendered digital image. We further evaluate the readability of recovered strokes. By employing an OCR system, we determine whether the input image and recovered strokes represent the same words.


1 INTRODUCTION

For a long time, handwriting analysis, such as handwriting recognition (Faundez-Zanuy et al., 2020) and signature verification (Diaz et al., 2022), has been an active research area. There are two categories of handwriting, online and offline. Online handwriting is captured in real-time on a digital device such as a tablet screen with a stylus pen. In contrast, the handwritten text scanned or captured by a camera from a physical medium such as paper is referred to as offline handwriting (Liu et al., 2011; Plamondon and Srihari, 2000). The handwriting inscribed on a digital device or captured from a physical medium is often unconstrained with varying orientations. The availability of temporal movements of a stylus pen for online handwriting makes the handwriting analysis task easier. However, for offline handwriting, the input is limited to handwritten images, making handwriting analysis much more difficult.

The current STR architectures for English hand-

writing use lines of text (Archibald et al., 2021; Bhunia et al., 2018) or characters of alphabets (Rabhi et al., 2021; Rabhi et al., 2022) as input. One of the recent datasets for STR, IAM-online (Marti and Bunke, 2002), includes only line-level annotations. For English handwritten documents, line detection is a prominent topic for processing historical document images (Boillet et al., 2021; Alberti et al., 2019). A learning-free mechanism for detecting lines in a historical handwritten document is presented in (Kurar Barakat et al., 2020). Another algorithm for line segmentation in handwritten documents is proposed in (Surinta et al., 2014b). This work is limited to separating the overlapping words only in horizontal lines (Surinta et al., 2014a). For more complex Arabic handwriting, (Gader and Echi, 2022; Gader and Echi, 2020) segment the curved text lines with overlapping words. A generative model to segment lines for Arabic handwriting is proposed in (Demir et al., 2021). This method can segment slanting and curved lines but is not accurately enclosing a complete word in lines because they consider a binary mask to train the generative model. We have tried a few line segmenta-

^a  <https://orcid.org/0000-0001-6531-7656>

^b  <https://orcid.org/0000-0002-5102-8244>

tion (Surinta et al., 2014b; Alberti et al., 2019) methods. They are either proposed for non-English text or do not work well for unconstrained non-horizontal text in handwritten document images. Even when line detection works well, performing STR on words yields better performance than on lines, as this paper demonstrates (the first two rows of Table 2).

Non-English handwriting datasets (Barakat et al., 2019; Yavariabdi et al., 2022) emphasize word-level annotations and provide word, line, and page-level annotations for historical handwritten documents. The Chinese, Japanese, Arabic, and Tamil (Bhunia et al., 2018; Nguyen et al., 2021; Viard-Gaudin et al., 2005; Privitera and Plamondon, 1995) datasets for stroke trajectory also provide word-level annotations. Older STR datasets such as IRONOFF (Viard-Gaudin et al., 1999) consists of words/characters/digits of English handwriting. Unipen dataset (Guyon et al., 1994) is available with character-level annotation, whereas IRONOFF consists of words. But we were unable to obtain any copies of IRONOFF datasets. In contrast, the publically available English handwriting dataset IAM-online (Marti and Bunke, 2002) for STR includes line-level annotations with missing word-level annotations. Therefore, we propose constructing a large-scale word-level annotation for the IAM-online dataset in our work. In recent years conventional methods (Diaz et al., 2022; Senatore et al., 2022) devised rule-based algorithms for signature/word trajectory recovery. Moreover, stroke trajectory recovery has progressed through deep neural networks. For stroke trajectory recovery, (Bhunia et al., 2018) introduces a first trainable convolutional network. This LSTM architecture learns strokes from Tamil scripts with Euclidean distance loss, making it hard to apply to long words with multiple strokes, such as English handwriting. Recently, (Rabhi et al., 2021; Rabhi et al., 2022) introduced an attention mechanism to train the writing order recovery for characters. These attention networks are trained on characters with L1-loss, which is, again difficult to train for words. Similarly, (Nguyen et al., 2021) employs an LSTM architecture with an attention layer and Gaussian mixture model trained with cross-entropy loss. However, it is limited to encoding only a single Japanese character. Recently, (Archibald et al., 2021) presents the stroke trajectory recovery, where LSTM is trained with a Dynamic Time Waring (DTW) loss function. (Archibald et al., 2021) has a disadvantage that it can work only for a line of text. Apart from the restriction of the input to the lines of text, DTW loss has a drawback for long sequence matching: it sums the loss function for all the points when finding the best alignment between two sequences. Hence order pre-

serving stray points, i.e., predicted stroke points far apart from their matching originals, have a minor influence on the DTW loss. However, they result in noticeable artifacts in the predicted strokes. To circumvent this issue, we propose to add the Chamfer distance (Akmal Butt and Maragos, 1998) between predicted and ground truth point sets to the loss function. In order to prevent penalizing stroke points with small deviations, we augment the Chamfer distance to a marginal Chamfer distance. Applying the marginal Chamfer distance yields a more significant penalty for stray points/artifacts.

Another challenge a stroke trajectory recovery system faces is the availability of ground truth strokes. Annotating ground-truth strokes for the STR system is a laborious process that demands time and resource allocation. Therefore, our work proposes an evaluation algorithm that does not require ground truth stroke points to evaluate the STR system. To the best of our knowledge, the evaluation of the STR system without ground truth stroke points has not been proposed before. The main contributions of this work are as follows: 1) We introduce large-scale word-level annotations for the English handwriting STR dataset sampled from the IAM-online dataset. Our version of the IAM-online dataset contains 62,000 words. 2) A word-level STR method estimates loss for each word rather than averaging DTW loss over the entire line of text. To avoid the stray points/artifacts in predicted stroke points, we employ a marginal Chamfer distance that penalizes large, easily noticeable deviations and artifacts. 3) We also introduce an algorithm for evaluating the STR system on images without ground truth stroke points. 4) Since our method works with words, we demonstrate that our method is scalable to unconstrained handwritten documents, i.e., full-page text. 5) The quantitative and qualitative results demonstrate the superior performance of our approach in comparison to the state-of-the-art (SOTA) methods.

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

2 METHOD

2.1 Network Architecture

We deploy a CNN with bidirectional LSTM for stroke trajectory learning. The input to the CNN is a word image resized to a fixed height with variable width to keep the same aspect ratio. A CNN branch consists of seven convolutional blocks with ReLU activation. The convolutional filters have a 3x3 kernel size with

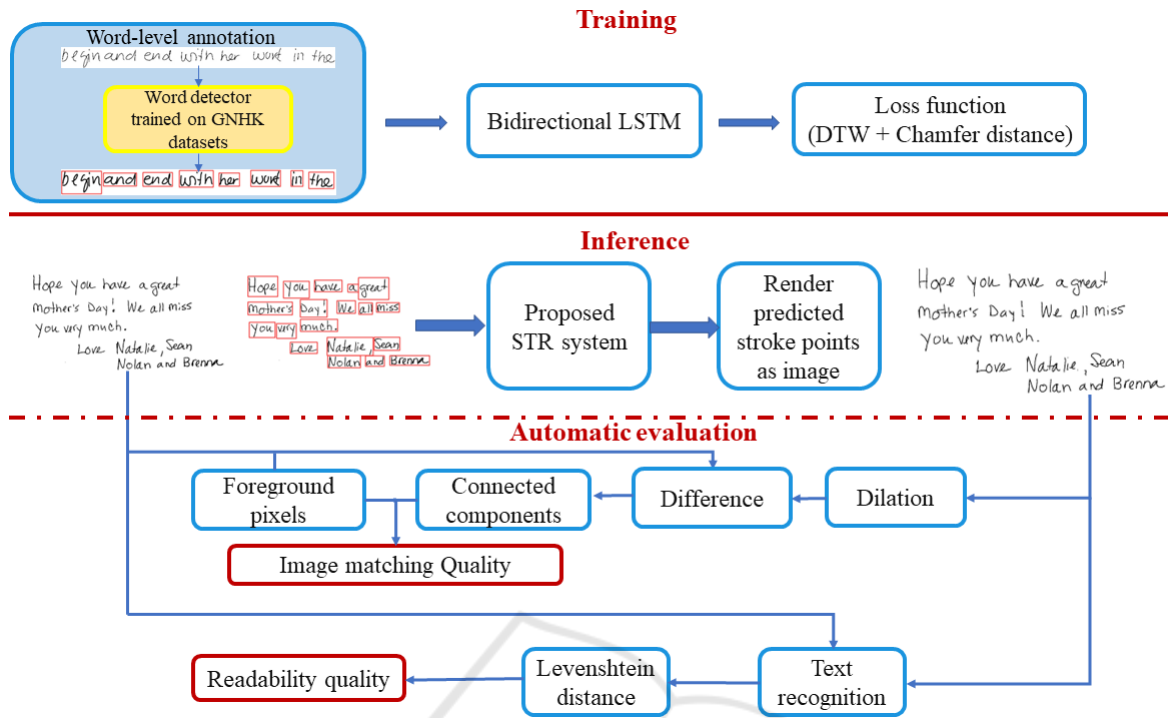


Figure 1: The block diagram of the overall system for training, inference, and automatic evaluation.

2x2 and 2x1 max pooling in each layer. The output of the CNN branch, a $W \times 1024$ dimensional feature vector, where W is the width of the image, is fed into eight bidirectional LSTM blocks. Each bidirectional LSTM block consists of 128 hidden units, so each LSTM block's input is $W \times 128$. Lastly, a bidirectional LSTM is followed by a 1-D convolutional block that predicts a $W \times 4$ -dimensional output. The number of output points is proportional to the width of the input image.

The first two dimensions of the output indicate the relative coordinates (x, y) with respect to the previous location. The last two dimensions indicate start-of-stroke (sos) and end-of-stroke (eos) tokens, respectively. Cross-entropy loss is employed to learn the start-of-stroke and end-of-stroke tokens. The overall architecture is shown in Fig. 1.

2.2 Loss Function

The ground truth in IAM-online dataset (Marti and Bunke, 2002) is a sequence of points defined as (x, y) coordinates with a time stamp. Since the number of predicted coordinates is proportional to the width of the input image, we re-sample the equidistant ground-truth coordinates such that the number of points is proportional to the image's width.

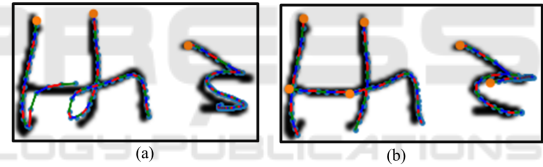


Figure 2: The impact of marginal Chamfer distance loss on the stroke trajectory recovery. (a) DTW loss only. (b) DTW + Chamfer distance loss.

2.2.1 Dynamic Time Warping (DTW) Loss

In general, DTW (Berndt and Clifford, 1994) computes the optimal match between ground-truth (GT) ($T = (t_1, t_2, t_3, \dots, t_m)$) and predicted sequences ($P = (p_1, p_2, p_3, \dots, p_n)$) of different lengths by finding the warping path between two sequences. In DTW loss, the cost matrix is calculated as:

$$\text{cost}(i, j) = \|p_i - t_j\|^2 \quad (1)$$

The accumulative cost matrix (A) is given as,

$$A(i, j) = \text{cost}(i, j) + \min[A(i-1, j), A(i-1, j-1), A(i, j-1)] \quad (2)$$

for $1 \leq i \leq n$ and $1 \leq j \leq m$.

Given matrix A , DTW computes the optimal warping path from $A(n, m)$ to $A(1, 1)$ as the alignment of points in P to points in T is expressed as index mapping $\alpha: \{1, \dots, n\} \rightarrow \{1, \dots, m\}$, where α is an onto

function.

$$\mathcal{L}_{DTW}(P, T) = \sum_{i=1}^n \|p_i - t_{\alpha(i)}\|. \quad (4)$$

2.2.2 Chamfer Distance Loss

The DTW gives promising results for training the stroke trajectory recovery systems (Archibald et al., 2021). It helps to match the point trajectory for ground truth and predicted strokes. However, the DTW loss function does not impose a sufficiently large penalty for predicted points following a similar trajectory as the ground truth points but far off compared to the ground truth at the pixel level. Fig. 2(a) shows that the predicted and ground truth points are far off, but the DTW loss is small since the predicted strokes follow the same trajectory as a ground truth stroke. Therefore, in this work, we propose to add a marginal Chamfer distance. Its effect is illustrated in Fig. 2(b). The proposed marginal Chamfer distance between the predicted and ground truth point sets is given by

$$d_{CD}(P, T) = \sum_{p \in P} \max(\min_{g \in T} \|p - g\|_2^2 - c^2, 0) + \quad (5)$$

$$\sum_{g \in T} \max(\min_{p \in P} \|g - p\|_2^2 - c^2, 0), \quad (6)$$

where P and T represent the point sets for predicted and ground truth strokes. The Chamfer distance is calculated on the pixel level. In the experimental section, we will discuss the setting of parameter c . The intuition behind the marginal Chamfer distance for STR is to consider the distance only if the predicted stroke point is at least a unit pixel apart from its nearest ground-truth stroke point. It leads to quantitative improvements in loss calculations as shown in Table 2. The proposed loss is simply a sum $\mathcal{L}_{DTW}(P, T) + d_{CD}(P, T)$

2.3 Automatic Evaluation

In the previous stroke trajectory recovery evaluation system, we computed the distance between ground-truth stroke points and predicted stroke points. For this purpose, we require the ground truth strokes' coordinate information to compute the difference. However, adding the coordinates information is extra work and requires extensive labor. Moreover, the existing handwriting datasets have limited availability of stroke coordinates information. Hence, although the proposed system can be applied to handwriting datasets without stroke coordinates information, it is impossible to evaluate the quality of its predicted strokes using the existing methods. Therefore,

we propose two measures for evaluating the quality of recovered stroke trajectories when ground truth stroke information is not given, namely image matching and readability.

2.3.1 Image Matching

The LSTM predicts the stroke trajectory recovery point coordinates (x, y) . We digitize the obtained strokes (X, Y) by plotting the points (x, y) and constructing a digital image from all the recovered stroke points (X, Y) . As we compare the original and reconstructed image, the dimensions of the reconstructed image are the same as the original image. However, since the text is plotted with unit thickness, the thickness of the text in the input and reconstructed image differs. To overcome this issue, we propose dilating the reconstructed images so that the thickness of the text in the input and the reconstructed image are the same. We dilate for the kernel size ranging from 0 to 10 and select the kernel size that yields the least number of pixels in the absolute difference between the input and dilated images. Let I_{input} be the input handwritten text image, and let $I_{predict}$ denote the image reconstructed from the coordinates of stroke points predicted by LSTM. $I_{predict}$ is reconstructed by digitizing the predicted stroke trajectory as described in Sec. 4.2. Hence all the words in $I_{predict}$ are one pixel thick.

Next, we dilate $I_{predict}$ with a dilation kernel $k \in (1, 10)$, and denote the dilated image with kernel k as $D(I_{predict}, k)$. We compare two digital images by computing their symmetric difference as

$$I_{Diff(k)} = |I_{input} - D(I_{predict}, k)|. \quad (7)$$

We define I_{Diff} as the image $I_{Diff(k)}$ with the minimum number of foreground pixels for $k \in (0, 10)$. This allows for estimating the thickness of the input text in the reconstructed image.

Next, we check the quality of the reconstructed image I_{Diff} by performing connected component analysis. Let C be the largest connected component in I_{Diff} image. The ratio of the number of foreground pixels in C to the total number of foreground pixels in I_{input} gives us the error in the stroke trajectory prediction, which is denoted as ϵ . This value can be used as a quantitative measure of the predicted strokes. Empirically, we observed that if the error ϵ is less than the threshold ($\mathcal{T} = 0.025$), then the quality of stroke trajectory recovery is good and vice versa.

The intuition of the proposed method is that the image I_{Diff} has small scattered connected components if stroke trajectory recovery is good. However, the $I_{Diff(k)}$ image has large and quite noticeable connected components if stroke trajectory recovery is of

poor quality. The example of good and poor stroke trajectory recovery validated by the proposed method is shown in Fig. 8.

2.3.2 Readability

The second part of the automatic evaluation checks the preservation of the readability of the input text and the text from the recovered handwriting trajectory. To verify that the recovered stroke trajectory is read the same as the input handwriting word, we recognize the characters in both images. Let I_i and I_r be the two images for word from the input handwriting image and the one recovered from the proposed stroke trajectory recovery method, respectively. The text recognition on I_i and I_r gives us the string of characters for input word denoted by $\mathcal{W}_i = [w_1, w_2, \dots, w_m]$ and the string characters from recovered stroke trajectory denoted by $\mathcal{W}_r = [w_1, w_2, \dots, w_n]$, where m and n are the total character recognized in I_i and I_r . We utilize the pre-trained text recognition network (Li et al., 2021) to compute \mathcal{W}_i and \mathcal{W}_r . The difference between the two recognized strings is computed by the edit distance between the two strings. In our work, we compute the edit distance between two strings \mathcal{W}_i and \mathcal{W}_r with Levenshtein distance. Let the Levenshtein distance between two strings be d_{lev} and the number of characters in input string \mathcal{W}_i is m . The readability error \mathcal{R} is defined as d_{lev}/m , that is, the ratio of incorrect string matching to the total number of characters in the input string. Ideally, the Levenshtein distance (d_{lev}) and readability error (\mathcal{R}) are expected to be zero for good stroke trajectory recovery. Empirically, we noticed that the \mathcal{R} less than $\mathcal{T} = 0.1$ results in satisfactory reconstruction, which we define as acceptable readability. This process is illustrated in Fig. 3.

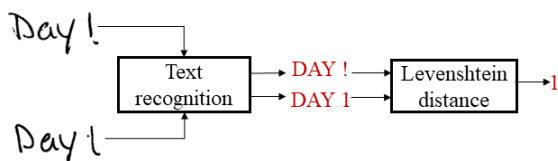


Figure 3: The readability module based on text recognition (Li et al., 2021) utilized for our automatic evaluation. The top-left text is the input and the bottom-left text is reconstructed from predicted stroke trajectory recovery.

3 EXPERIMENTS

We first present the construction of word-level annotations for IAM-online dataset (Sec. 3.1) and introduce the greeting-card handwritten messages (GHM) dataset (3.1.1). We discuss the evaluation metric and results of the proposed method on these datasets

(Secs. 3.2 and 4). Finally, we present the application of our proposed method for the GHM dataset (Sec. 4.2).

3.1 Word-Level Annotation

In the IAM-online dataset (Marti and Bunke, 2002), the stylus pen movement on an electronic device’s screen provides the coordinates of the ground-truth point for stroke trajectory recovery. However, the IAM-online dataset only provides the stroke’s ground truth for the line-level text. Therefore, we use word detection to construct stroke annotations for words. Word detection generally requires word bounding boxes to train the detection network, but the IAM-online dataset does not include word bounding boxes. So, we propose to train the word detection on GNHK (Lee et al., 2021) dataset and then applied to the IAM-online dataset. The images in the GNHK dataset are sourced from different regions of Europe, North America, Africa, and Asia containing 39k+ words, sufficient to train a word detector with data augmentation (Bochkovskiy et al., 2020). Hence, it is a diverse dataset regarding writing style and image quality as the penmanship varies in different parts of the world, and camera quality varies for each captured image.

We explore three state-of-the-art detectors; one scene text detector (Baek et al., 2019) and two object detectors for localizing the words in unconstrained handwritten text. A scene text detection network identifies character regions in natural images and detects the words based on character regions and affinity scores between them (Baek et al., 2019). However, the deterministic approach to constructing bounding boxes around character region scores is not well suited for handwriting word detection. Because in the handwritten text, adjacent lines may overlap, and characters may have high affinity scores in adjacent lines, which misleads the word detection results. This shortcoming of the scene text detector gives low detection accuracy for unconstrained handwritten text (as shown in Fig. 5).

On the other hand, the object detector attributes words as objects and is more efficient for detecting overlapping words in adjacent lines in a non-horizontal orientation. For our task, a single-stage detector YOLO (Bochkovskiy et al., 2020) performs better than a two-stage detector Faster R-CNN (Ren et al., 2015; Hanif et al., 2019; Hanif and Latecki,). Both are trained and evaluated on GNHK dataset (Lee et al., 2021). In Table 1, we list the quantitative results for word detection on GNHK dataset. Due to its best performance, we select the single-stage YOLO word detector (Bochkovskiy et al., 2020) and apply

Table 1: The accuracy (mAP) of word detection with state-of-the-art scene text detector (Baek et al., 2019) and single and two-stage object detectors (Ren et al., 2015; Bochkovskiy et al., 2020).

Method	mAP@0.5	mAP@0.5:0.95
Scene text detector (Baek et al., 2019)	0.603	0.565
Two-stage word detector (Ren et al., 2015)	0.780	0.565
Single-stage word detector (Bochkovskiy et al., 2020)	0.913	0.619

it to IAM-online dataset (Marti and Bunke, 2002) for word detection.

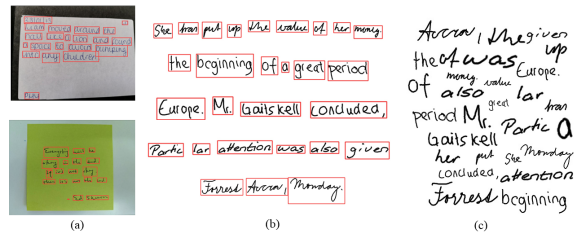


Figure 4: (a) Word detection for the GNHK dataset (Lee et al., 2021), (b) word detection for IAM-online dataset (Marti and Bunke, 2002), (c) Words from IAM-online dataset used to train our network.

In Fig. 4(a,b), we show the word detection visualization for sample images from GNHK dataset (Lee et al., 2021) and IAM-online dataset (Marti and Bunke, 2002) respectively. The images in GNHK (Lee et al., 2021) vary in handwriting style, background, and camera conditions. Therefore, the word detection trained in GNHK dataset (Lee et al., 2021) has acceptable performance for the IAM-online dataset (Marti and Bunke, 2002). We used the word-level annotations for the IAM-online dataset (Marti and Bunke, 2002) to finetune and evaluate our system. Word detection on the IAM-online dataset gives us 41,665 and 19,496 words for train and test sets, respectively. Fig. 4(b) shows the word detection from lines from IAM-online, and Fig. 4(c) shows the words used in our work to train the network.

3.1.1 Greeting Card Messages Dataset

In our work, we propose a word-level stroke trajectory recovery.

To evaluate our system on unlabelled handwritten documents, we acquire approximately 2,000 greeting-cards handwritten messages (GHM) dataset from greeting cards company (Signed, 2022). The GHM dataset shared¹. Handwritten messages from the GHM dataset do not follow any fixed template because it consists of user-uploaded handwritten messages to greeting cards company (Signed, 2022). The

¹<https://drive.google.com/file/d/1G-EZBfEhsHThR9dR1YtJdPE3Mg0ay0w-/view?usp=sharing>

samples from the GHM dataset are shown in Fig. 5.

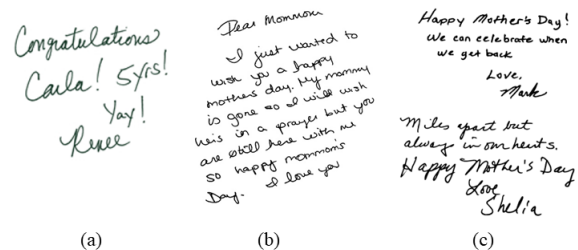


Figure 5: Sample documents form GHM dataset with unlabelled handwriting images.

3.2 Evaluation Metric

We used a distance-based evaluation metric to evaluate stroke trajectory recovery. The average distance of points in the ground-truth (T) stroke to its nearest predicted stroke (P) is denoted by $dist_{t,p}$. Similarly, the average distance of points in the predicted stroke (P) to its nearest ground-truth stroke (T) is denoted by $dist_{p,t}$. The metric $dist_{t,p}$ signifies that every ground-truth stroke point is close to the predicted point and vice versa for $dist_{p,t}$. $dist_{t,p}$ and $dist_{p,t}$ are the same evaluation metrics as used in (Archibald et al., 2021). However, apart from the mean (*mean*) of the distances between predicted and ground truth stroke points, we also compute the standard deviation (*std*) of the metrics. We also compute the loss for predicting the start-of-stroke token ϵ_{SOS} . ϵ_{SOS} should have the lowest value if the start-of-stroke token is predicted correctly.

4 RESULTS

DTW and Chamfer distance are complementary loss functions to train our system. DTW loss ensures that the predicted stroke sequences are similar to the ground truth stroke sequence. The Chamfer distance between the predicted and ground truth point set ensures that there are no spurious points/artifacts in predicted points; that is, no predicted points are far away from ground truth points.

Table 2: The quantitative comparison for training on DTW loss for lines and words. The third row lists the quantitative results for training with combined DTW and Chamfer distance loss between predicted and ground-truth points.

Loss function	$dist_{t,p}$ (mean)	$dist_{t,p}$ (std)	$dist_{p,t}$ (mean)	$dist_{p,t}$ (std)	ϵ_{sos}
DTW_{line} (Archibald et al., 2021)	0.01558	0.01402	0.02776	0.0353	0.1553
DTW_{word}	0.01653	0.014862	0.01287	0.00965	0.1427
DTW_{word} + Chamfer distance	0.01492	0.01042	0.01257	0.00946	0.1479

Table 3: The quantitative comparison of increasing the value of c from 1 to 3.

Loss function		$dist_{t,p}$ (mean)	$dist_{t,p}$ (std)	$dist_{p,t}$ (mean)	$dist_{p,t}$ (std)
DTW_{word} + Chamfer distance	$c = 1$	0.01492	0.01042	0.01257	0.00946
	$c = 3$	0.0177	0.0118	0.01859	0.0162

4.1 IAM-Online Dataset

Table 2 presents a quantitative comparison, where bold numbers show the best results (lowest value). The first and second rows of Table 2 show the evaluation with line-level and word-level input for DTW loss, respectively. We observe that both $dist_{t,p}$ and $dist_{p,t}$ metrics are much lower for word-level than the line-level input. These results show that training the stroke trajectory recovery with a word-level dataset, as we proposed, improves the results. We also validated that Chamfer distance loss for predicted and ground truth point sets improves the quantitative results. Both mean and standard deviation of $dist_{p,t}$ and $dist_{t,p}$ decrease by adding Chamfer distance to the loss function. It means that the predicted strokes are better at imitating the ground-truth strokes. The lower values of both $dist_{t,p}$ and $dist_{p,t}$ illustrate that every ground-truth stroke has a close predicted stroke and vice versa. So, we do not get spurious predicted strokes and yet do not miss to follow the shape of the ground-truth strokes. We noticed that chamfer distance loss has minimal influence on start-of-stroke (ϵ_{sos}) as shown in Table 2. In another experiment, we try the higher values of c (Eq. 6) as shown in Table 3, but by increasing the value of c increases the std of $dist_{t,p}$ and $dist_{p,t}$. Therefore, we keep the value of $c=1$ in our work.

The visualization of the elimination of spurious predicted points/artifacts from the predicted stroke trajectory by adding Chamfer distance loss is illustrated in Fig. 6. We can see that extra stroke points cause more artifacts in (a) than in (b).

4.2 Greeting Cards Messages Dataset

The previous methods on the IAM-online dataset work with line-level text for stroke trajectory recovery (Archibald et al., 2021), which is not scaleable to un-

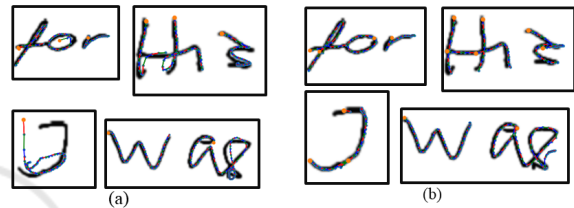


Figure 6: Samples of stroke trajectory recovery with (a) DTW loss and (b) DTW loss with Chamfer distance on the predicted and ground truth point sets. The recovered stroke trajectory is shown by red, blue, and green arrows, and the predicted start-of-stroke point is shown as an orange circle (best view in colored).

constrained handwritten text images without line detection. This is one of the main reasons why we work with word-level text.

In one of the applications of the proposed work, we show the stroke trajectory recovery for greeting-card handwritten messages (Signed, 2022). We applied the developed method trained for word-level annotation using DTW and Chamfer distance to the images containing greeting-card handwritten messages. Handwriting images from greeting card messages do not follow any fixed template because they consist of user-uploaded handwritten messages. Therefore, first, we detect the words in greeting card handwritten messages with word detector described in Sec. 3.1. Then we execute the trained STR model on each detected word. The recovery of stroke trajectory for greeting card handwritten messages is the word-level stroke trajectory recovery of each word.

We render the image from predicted stroke points as described in Fig. 7, where the proposed STR system predicts the stroke trajectory recovery for each word. Whereas the render image module converts stroke points into an image. Finally, we align the rendered image to the location of the detected word.

The visual results of the proposed word-level STR system on handwritten greeting card messages as

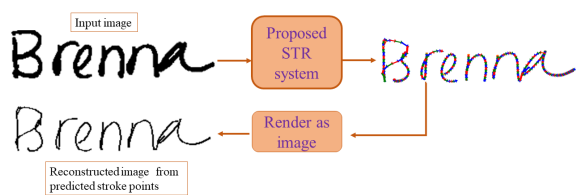


Figure 7: The mechanism of image rendering from predicted stroke points.

shown in Fig. 9 are rendered word-wise by the image render module illustrated in Fig. 7. Our proposed STR system can easily be applied to text with different orientations, as shown in Fig. 9. The left-hand images are input handwritten messages, and the right-hand images are rendered from the recovered stroke trajectory for each word.

4.2.1 Automatic Evaluation

We also applied the proposed automated evaluation method to the GHM dataset. In Fig. 8, we show the input handwritten messages, the dilated image, and connected components for I_{Diff} . The recovered image is dilated to match the width of the words in the input handwriting image as described in Sec. 2.3.1. According to the criteria defined in Sec. 2.3.1, the Fig. 8(top) example shows the *good* stroke trajectory recovery with small scattered connected components. Whereas, Fig. 8(bottom) shows the *poor* stroke trajectory recovery as there are larger connected components in the difference image I_{Diff} .

In Table 4, we listed the accuracy of the image matching and readability-based evaluation proposed in our work. The first column in Table 4 shows the percentage of documents with correctly recovered stroke trajectories according to our two proposed quality measures. According to the thresholds defined in Sec 2.3.1 and Sec. 2.3.2, the accuracy of stroke recovery from image matching and readability is 24.30% and 24.76%, respectively.

We manually verified the accuracy with user scoring.

If the threshold defined in image matching and readability evaluation classifies the image from the recovered stroke trajectory as satisfactory and the user also gives a satisfactory score to the recovered image, then a confidence score of 1 is assigned to that reconstruction. The average confidence score (*confidence*) for all the images is computed. The user scores the image in binary, scoring either 0 or 1. We applied this binary criterion to evaluate the robustness of the thresholds defined in Sec. 2.3.1 and Sec. 2.3.2. The accuracy of image matching and readability with the confidence score (*confidence*) are listed in Table 4. Our observation shows that the automatic evaluation

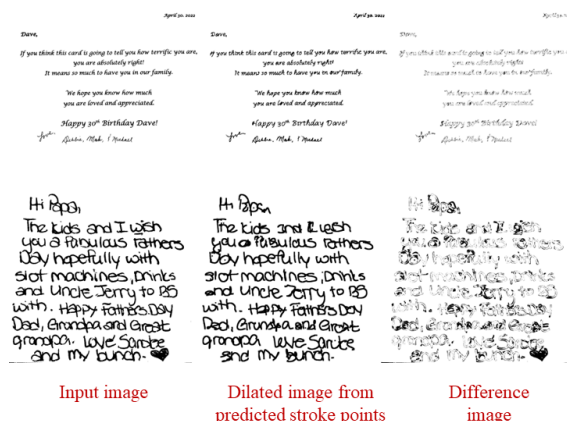


Figure 8: The input image, dilated image (after dilation is applied to the recovered image), and the difference of the input and dilated images.

based on image matching is a better evaluation measure than the readability evaluation. One of the reasons is that the text recognizer can correctly recognize words even if they are visually dissimilar to the input handwritten words.

Table 4: The quantitative analysis of automatic evaluation on greeting cards handwritten messages (Signed, 2022).

Method	Accuracy	Confidence
Image matching	24.30%	74.0%
Readability evaluation	24.76%	53.8%

5 CONCLUSIONS

In our proposed work, we trained a neural network with word-level annotations for the IAM-online dataset using DTW and Chamfer distance loss functions. We demonstrated that adding Chamfer distance loss to DTW is beneficial for removing artifacts and spurious stroke points for better stroke trajectory recovery. We also proposed automatic evaluation methods using image matching and readability consistency to evaluate the quality of stroke trajectory recovery for unlabeled datasets. Finally, we demonstrate the ability of our proposed work to work in unconstrained practical applications by applying and evaluating it on an unlabeled handwriting greeting card messages dataset.

ACKNOWLEDGEMENT

We would like to thank Signed Inc. for funding this work. The calculations were carried out on HPC resources at Temple University, supported in part

Input handwritten document

Will and Christina,
 We are so very sorry for your loss. We hope you are comforted by remembering all the good times you shared with your mother. May God grant you peace during this time.
 Much Love! From your JRSS (Engineering, Field and Staging) Team/Family at ByLight

Hope you have a great
 Mother's Day! We all miss
 you very much.
 Love Natalie, Sean
 Nolan and Brenna

Hi Alain!

Hope all is well!

I just wanted to send you a little something to express my continuing gratitude for referring couples my way. Your trust is so meaningful to me and I love working with you and your whole team!

I'm looking forward to working with you again soon but in the meantime, please check your email for a little something from me. I hope it makes a rough day a little easier or is the start of celebrating a great one!

Thanks again, see you soon!

*Best,
 Ralph*

Rendered image from predicted stroke trajectory

Will and Christina,
 We are so very sorry for your loss. We hope you are comforted by remembering all the good times you shared with your mother. May God grant you peace during this time.
 Much Love! From your JRSS (Engineering, Field and Staging) Team/Family at ByLight

Hope you have a great
 Mother's Day! We all miss
 you very much.
 Love Natalie, Sean
 Nolan and Brenna

Hi Alain!

Hope all is well!

I just wanted to send you a little something to express my continuing gratitude for referring couples my way. Your trust is so meaningful to me and I love working with you and your whole team!

I'm looking forward to working with you again soon but in the meantime, please check your email for a little something from me. I hope it makes a rough day a little easier or is the start of celebrating a great one!

Thanks again, see you soon!

*Best,
 Ralph*

Figure 9: The input image and the rendered image from predicted stroke trajectory recovery for GHM dataset.

by the NSF through grant number 1625061 and by the U.S. ARL under contract number W911NF-16-2-0189. We would also like to thank the undergraduate student Abdulrahman Samir Alshehri for cleaning the dataset for the handwritten documents

REFERENCES

- Akmal Butt, M. and Maragos, P. (1998). Optimum design of chamfer distance transforms. *IEEE Transactions on Image Processing*, 7(10):1477–1484.
- Alberti, M., Vögtlin, L., Pondenkandath, V., Seuret, M., Ingold, R., and Liwicki, M. (2019). Labeling, cutting, grouping: an efficient text line segmentation method

- for medieval manuscripts. In *2019 International Conference on Document Analysis and Recognition (ICDAR)*, pages 1200–1206. IEEE.
- Archibald, T., Poggemann, M., Chan, A., and Martinez, T. (2021). Trace: A differentiable approach to line-level stroke recovery for offline handwritten text. *arXiv preprint arXiv:2105.11559*.
- Baek, Y., Lee, B., Han, D., Yun, S., and Lee, H. (2019). Character region awareness for text detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9365–9374.
- Barakat, B. K., El-Sana, J., and Rabaev, I. (2019). The pinkas dataset. In *2019 International Conference on Document Analysis and Recognition (ICDAR)*, pages 732–737. IEEE.
- Berndt, D. J. and Clifford, J. (1994). Using dynamic time warping to find patterns in time series. In *KDD workshop*, volume 10, pages 359–370. Seattle, WA, USA:.
- Bhunias, A. K., Bhowmick, A., Bhunia, A. K., Konwer, A., Banerjee, P., Roy, P. P., and Pal, U. (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.
- Bochkovskiy, A., Wang, C.-Y., and Liao, H.-Y. M. (2020). Yolov4: Optimal speed and accuracy of object detection. *arXiv preprint arXiv:2004.10934*.
- Boillet, M., Kermorvant, C., and Paquet, T. (2021). Multiple document datasets pre-training improves text line detection with deep neural networks. In *2020 25th International Conference on Pattern Recognition (ICPR)*, pages 2134–2141. IEEE.
- Demir, A. A., ÖzŞeker, İ., and Özkaya, U. (2021). Text line segmentation in handwritten documents with generative adversarial networks. In *2021 International Conference on Innovations in Intelligent Systems and Applications (INISTA)*, pages 1–5. IEEE.
- Diaz, M., Crispo, G., Parziale, A., Marcelli, A., and Ferrer, M. A. (2022). Writing order recovery in complex and long static handwriting.
- Faundez-Zanuy, M., Fierrez, J., Ferrer, M. A., Diaz, M., Tolosana, R., and Plamondon, R. (2020). Handwriting biometrics: Applications and future trends in e-security and e-health. *Cognitive Computation*, 12(5):940–953.
- Gader, T. B. A. and Echi, A. K. (2020). Unconstrained handwritten arabic text-lines segmentation based on ar2u-net. In *2020 17th International Conference on Frontiers in Handwriting Recognition (ICFHR)*, pages 349–354. IEEE.
- Gader, T. B. A. and Echi, A. K. (2022). Deep learning-based segmentation of connected components in arabic handwritten documents. In *International Conference on Intelligent Systems and Pattern Recognition*, pages 93–106. Springer.
- Guyon, I., Schomaker, L., Plamondon, R., Liberman, M., and Janet, S. (1994). Unipen project of on-line data exchange and recognizer benchmarks. In *Proceedings of the 12th IAPR International Conference on Pattern Recognition, Vol. 3-Conference C: Signal Processing (Cat. No. 94CH3440-5)*, volume 2, pages 29–33. IEEE.
- Hanif, S. and Latecki, L. J. Autonomous character region score fusion for word detection in camera-captured handwriting documents.
- Hanif, S., Li, C., Alazzawe, A., and Latecki, L. J. (2019). Image retrieval with similar object detection and local similarity to detected objects. In *Pacific Rim International Conference on Artificial Intelligence*, pages 42–55. Springer.
- Kurar Barakat, B., Cohen, R., Droby, A., Rabaev, I., and El-Sana, J. (2020). Learning-free text line segmentation for historical handwritten documents. *Applied Sciences*, 10(22):8276.
- Lee, A. W., Chung, J., and Lee, M. (2021). Gnhk: A dataset for english handwriting in the wild. In *International Conference on Document Analysis and Recognition*, pages 399–412. Springer.
- Li, M., Lv, T., Cui, L., Lu, Y., Florencio, D., Zhang, C., Li, Z., and Wei, F. (2021). Trocr: Transformer-based optical character recognition with pre-trained models. *arXiv preprint arXiv:2109.10282*.
- Liu, C.-L., Yin, F., Wang, D.-H., and Wang, Q.-F. (2011). Casia online and offline chinese handwriting databases. In *2011 International Conference on Document Analysis and Recognition*, pages 37–41. IEEE.
- Marti, U.-V. and Bunke, H. (2002). The iam-database: an english sentence database for offline handwriting recognition. *International Journal on Document Analysis and Recognition*, 5(1):39–46.
- Nguyen, H. T., Nakamura, T., Nguyen, C. T., and Nakawaga, M. (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.
- Plamondon, R. and Srihari, S. N. (2000). Online and offline handwriting recognition: a comprehensive survey. *IEEE Transactions on pattern analysis and machine intelligence*, 22(1):63–84.
- Privitera, C. M. and Plamondon, R. (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.
- Rabhi, B., Elbaati, A., Boubaker, H., Hamdi, Y., Hussain, A., and Alimi, A. M. (2021). Multi-lingual character handwriting framework based on an integrated deep learning based sequence-to-sequence attention model. *Memetic Computing*, 13(4):459–475.
- Rabhi, B., Elbaati, A., Boubaker, H., Pal, U., and Alimi, A. (2022). Multi-lingual handwriting recovery framework based on convolutional denoising autoencoder with attention model.
- Ren, S., He, K., Girshick, R., and Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems*, 28.

- Senatore, R., Santoro, A., Parziale, A., and Marcelli, A. (2022). A biologically inspired approach for recovering the trajectory of off-line handwriting. Signed (2022).
- Surinta, O., Holtkamp, M., Karaaba, M. F., van Oosten, J., Schomaker, L. R. B., and Wiering, M. A. (2014a). A* path planning for line segmentation of handwritten documents. In *Frontiers in Handwriting Recognition (ICFHR), 2014 14th International Conference on*, pages 175–180. IEEE.
- Surinta, O., Holtkamp, M., Karabaa, F., Van Oosten, J.-P., Schomaker, L., and Wiering, M. (2014b). A path planning for line segmentation of handwritten documents. In *2014 14th international conference on frontiers in handwriting recognition*, pages 175–180. IEEE.
- Viard-Gaudin, C., Lallican, P.-M., and Knerr, S. (2005). Recognition-directed recovering of temporal information from handwriting images. *Pattern Recognition Letters*, 26(16):2537–2548.
- Viard-Gaudin, C., Lallican, P. M., Knerr, S., and Binter, P. (1999). The ireste on/off (ironoff) dual handwriting database. In *Proceedings of the Fifth International Conference on Document Analysis and Recognition. ICDAR'99 (Cat. No. PR00318)*, pages 455–458. IEEE.
- Yavariabdi, A., Kusetogullari, H., Celik, T., Thummanapally, S., Rijwan, S., and Hall, J. (2022). Cardis: A swedish historical handwritten character and word dataset. *IEEE Access*, 10:55338–55349.

SCITEPRESS
SCIENCE AND TECHNOLOGY PUBLICATIONS