Supplementary File: Exploring the Boundaries of Diffusion Models for Offline Writer Identification with Sparse and Intra-Variable Data

Aritra Dey[1], Chandranath Adak[1], Kumari Priya[1], Soumi Chattopadhyay[2], Sukalpa Chanda[3]
[1]Dept. of CSE, Indian Institute of Technology Patna, India-801106
[2]Dept. of CSE, Indian Institute of Technology Indore, India-453552
[3]Dept. of CSC, Østfold University College, Norway-1757
aritradey1993@gmail.com, chandranath@iitp.ac.in

1. Analysis of Style Embeddings in V1-TD

In our results section, we have already provided and discussed the performance of our model on the V1-TD dataset under close vocabulary conditions. In this supplement, we provide a deeper analysis of our model using a variety of experiments. The analysis will include the following:

- Qualitative analysis of handwriting style embeddings for the testing set in the form of t-SNE maps.
- Qualitative and quantitative investigation of shape perturbations using skeleton warping.
- Investigate style embeddings under two sub-conditions:
  - Same word written by different writers.
  - Different word samples written by the same writer
- Model analysis by word length

1.1. Qualitative Analysis of Handwriting Style Embeddings for the Test Set

In Figure 1, we present the style embeddings in the form of t-SNE maps. Since our proposed dataset contains a total of 100 writers, we have intentionally segregated them into 10 sub-figures to showcase the style embeddings of 10 consecutive writers in one figure. So, the top left figure indicates embeddings for writer IDs 0-9; 10-19 in the immediate right, and so on till the bottom right, which has writers 90-99.

A brief overall look at all the sub-figures shows that there is some amount of outliers in most of the sub-figures, but in the t-SNE map for writers 0-9 and writers 80-89, two clusters have completely submerged into one another. Embeddings with writer ID 3 and writer ID 4 have no clear boundary between their clusters. Similarly, embeddings from writer IDs 81 and 89 have no visible segregation between their clusters.

1.2. Qualitative and Quantitative Investigation of Shape Perturbations

In handwriting analysis, specifically, writer identification, the overall objective is to model the intrinsic writing patterns of individual authors. Shape perturbations are necessary litmus tests to evaluate the robustness of our model. In this analysis, we have specifically considered the skeleton warping-based shape perturbation, as it distorts the strokes of the handwritten characters. Figure 2 provides a visual example of how skeleton warping can distort natural strokes in handwritten texts. We considered 4 random writers and 4 random words from their samples. The upper row shows unperturbed handwriting, while the lower row depicts warped skeleton modifications of their former state.

For the qualitative results of shape perturbation, we again refer to Figure 1. Each subfigure clearly indicates that the handwriting style embedding from all the authors in our dataset has a clear and recognizable boundary under closed vocabulary conditions.

Our method for shape perturbation is inspired by [2]. The following steps define the process of shape perturbation used in our experiments.

Displacement Field Generation: Two random displacement maps are constructed:

$$\begin{aligned} dx(x,y) &= \alpha \cdot \big(G_\sigma * \eta_x\big)(x,y)\ ; \\ dy(x,y) &= \alpha \cdot \big(G_\sigma * \eta_y\big)(x,y), \end{aligned} \tag{1}$$

where, $\eta_x, \eta_y \sim \mathcal{U}(-1,1)$ are IID random fields sampled from continuous uniform distribution $\mathcal{U}$, G_σ is a Gaussian kernel with standard deviation σ, $\circledast$ denotes convolution, α controls the distortion strength.

Warp Mapping: The warp function is defined as:

$$\varphi(x,y) = \big(x + dx(x,y),\ y + dy(x,y)\big) \tag{2}$$

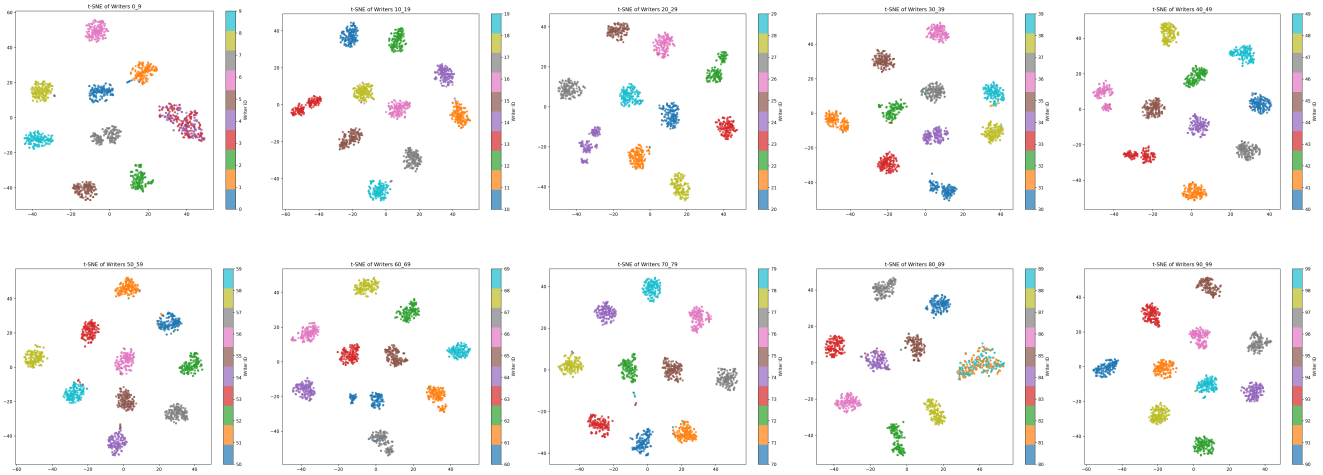


Figure 1. t-SNE map for the dataset V1-TD under shape perturbation modification. Each subplot shows a group of 10 writers, e.g., the first subplot presents Writer IDs 0–9, the last subplot presents Writer IDs 90–99.

Image Remapping: The distorted image I' is obtained from the original image I by backward mapping:

$$I'(x, y) = I(x + dx(x, y),\ y + dy(x, y)) \tag{3}$$

Final Form: Substituting the displacement fields:

$$I'(x, y) = I\big(x+\alpha(G_\sigma \circledast \eta_x)(x, y),\ y+\alpha(G_\sigma \circledast \eta_y)(x, y)\big) \tag{4}$$

Table 1. Shape perturbation analysis on Dataset V1-TD

Shape Perturbation	α	σ	Top-1 Accuracy (%)
No	N.A.	N.A.	90.77
Yes	30	5	90.25
Yes	36	7	90.20

Table 1 indicates the quantitative results from the shape perturbation based on skeleton warping. In this table, alpha and sigma are the skeleton warping parameters. From the table, it is evident that skeleton warping perturbation has little to no effect on modeling the intricate handwriting patterns or style of individual writers in the V1-TD dataset.

1.3. Investigation of Style Embeddings Under Inter-Content and Inter-Writer Variations

In this section, we analyze the style embeddings under two different conditions. The inter-writer variability on the same word and the inter-word variability for the same writer.

1.3.1. Same Word Written by Different Writers

We now investigate the embeddings of the word images encoded by the model from different writers. The experiment is strictly performed using a single word label written by different authors. In our experiment, we have chosen words of varied length. The word 'an' is of 2 characters, while 'before' is of 6 characters in length. We have already mentioned in our dataset description that in our testing set, each writer has 2 samples for each word available in the dataset. This is evident in Figure 3. Each writer in the experiment is color-coded, and each writer has exactly two datapoints in the scatter plot. Here, each datapoints is a style embedding of the corresponding writer of the chosen word label.

It is interesting to note that no two embeddings from the same writer overlap with each other, yet they are closer to one another as they are far from the embeddings of other writers corresponding to the same word label. This clearly indicates that the embedding clusters are well separated from each other even though the given writers have written the same word.

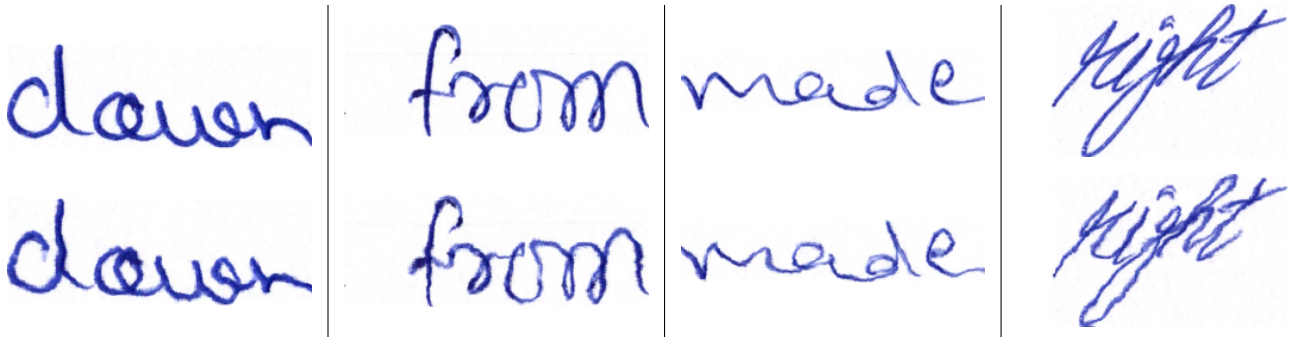
1.3.2. Different Word Samples Written by the Same Writer

Although this experiment requires analyzing different word embeddings from a single writer, we have used 4 random writers, each writing a set of 6 random words from the dataset. This approach was required to clarify that a set of distinct word embeddings from the same writer cluster together, and they are well separated from the same set of word embeddings of a different writer.

Figure 4 shows that a particular author writing a unique set of words can very well be identified from another writer writing the same set, which is the original claim of our work.

1.4. Analysis of Model Performance using Word Length Variations

Since our custom word dataset consists of words of variable length, it is essential to verify the model's ability to iden-



(a) Writer 31 (b) Writer 52 (c) Writer 20 (d) Writer 90

Figure 2. Comparison of normal vs. skeleton warped ($\alpha = 36$, $\sigma = 7$) handwriting samples from four random writers.

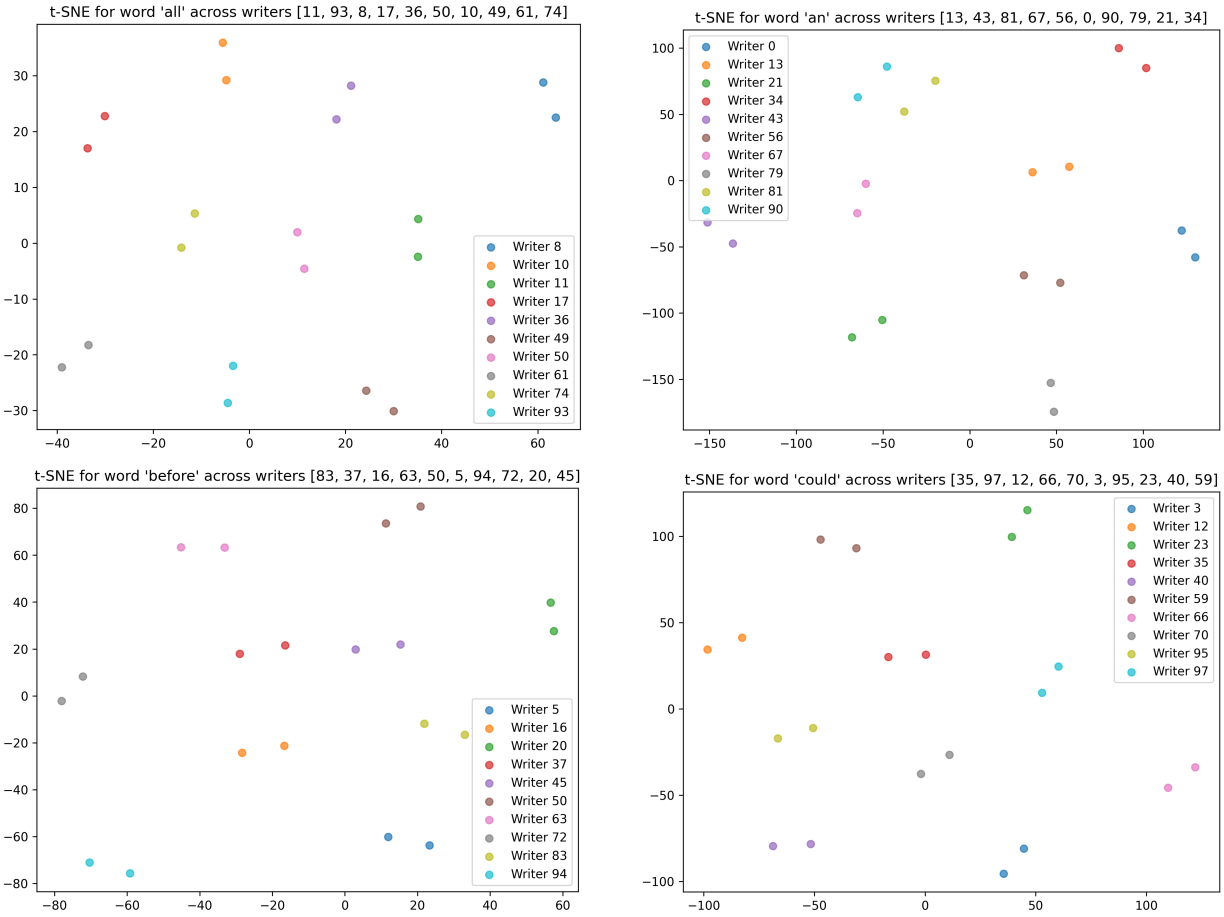


Figure 3. t-SNE map for analyzing the same word with different writers on the dataset V1-TD.

tify authors based on the length of the word data. Figure 5 provides a comprehensive view of the writer identification accuracy purely based on the length of the given word. The word length in our dataset varies from 1 to 6.

Although Top-5 accuracy has been consistently above 98%, Top-1 accuracy exhibits a significant variation with respect to the length of the word. In the domain of handwriting analysis and writer identification, the relevant information is available in the non-background pixels, specifically in the strokes and loops of handwritten characters. This information are building block essential for obtaining the feature vectors used in writer identification. The more loops and strokes a word image has, the better the identification of the model.

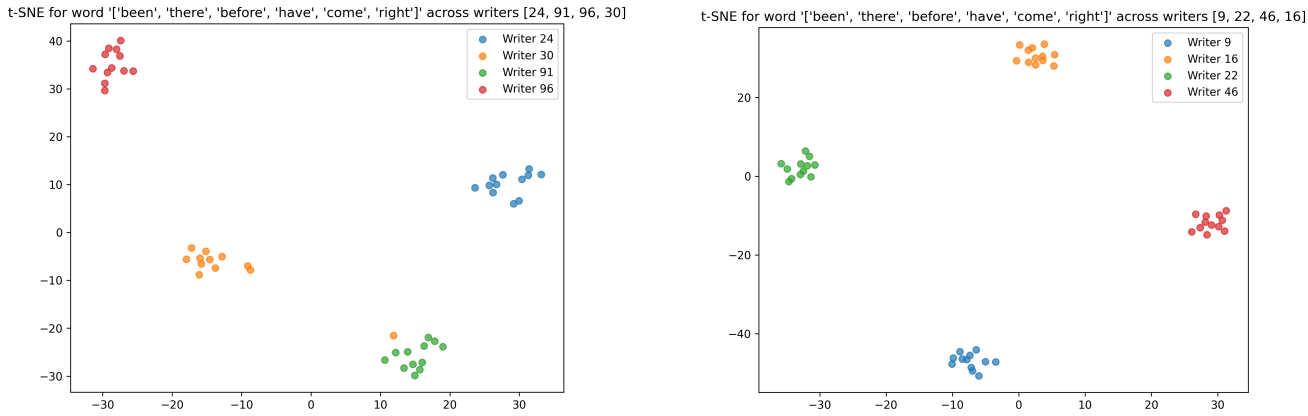


Figure 4. t-SNE map for analyzing the same writer with different words on the dataset V1-TD.

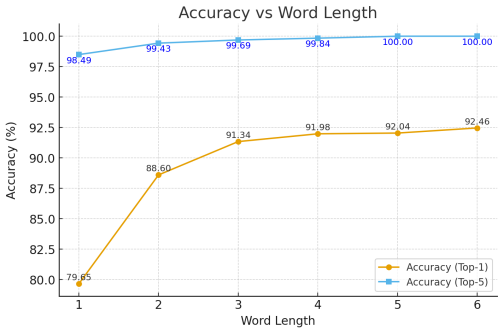


Figure 5. Accuracy (*Top-1* and *Top-5*) with respect to word length. The plot shows that Top-1 accuracy improves as word length increases and saturates around 92%, while Top-5 accuracy quickly reaches 100%.

This experiment highlights an important aspect of information content in handwritten text images. Single alphabet words like 'a', 'i', 'I', and words with simple alphabets are more prone to misclassification.

2. Cross-Dataset Analysis of Our Model

In our Solution Architecture, we have already mentioned that our model uses a pretrained mobilenet-based style encoder. Although it worked considerably well for text-dependent data, for text-independent data it has shown downgraded performance. To explore the cross-dataset applicability of our trained model trained on V1-TD, we have finetuned our model on the IAM and CERUG-EN datasets using transfer learning.

Table 2 shows the performance of our model in a cross-domain scenario. We have already trained our model on V1-TD, which has achieved a Top-1 accuracy of 90.77% under closed vocabulary constraints. The IAM and CERUG-EN datasets initially achieved Top-1 accuracies of 81.91% and 71.58%, respectively. However, finetuning these datasets with pre-trained weights from V1-TD further improved their identification accuracy to 84.55% and 75.64%, respectively.

Table 2. Cross-domain analysis on IAM and CERUG-EN

Dataset	Top-1 Accuracy (%)
IAM	84.55
CERUG-EN	75.64

3. Choice of Handcrafted Features

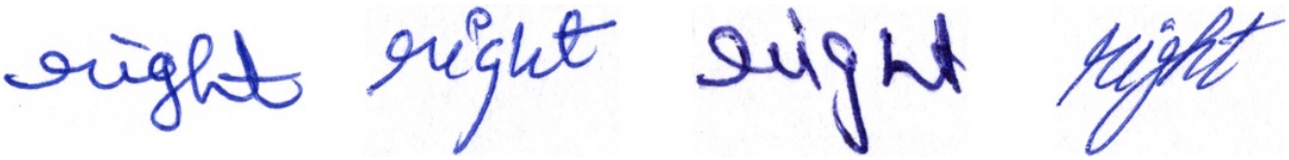
In this section, we discuss the handcrafted feature selection used in this study.

Contour-Hinge Features: From any given handwritten text image, contour-hinge extracts a joint distribution of pairs of contour directions at a fixed distance along the boundary (essentially encoding hinge angles). The main advantage of contour-hinge is its ability to encode hinge points, which are essentially curves in a handwritten stroke or loop. They carry crucial information about the handwriting pattern of an individual, providing rich representations.

Contour-Direction Features: On the other hand, contour-direction extracts a distribution of contour directions (angles of boundary pixels). It essentially encodes orientation information like slant, stroke flow. Although it provides writer-specific traits like preferred stroke direction, it misses out stroke transitions as it only considers local direction, but not how directions connect.

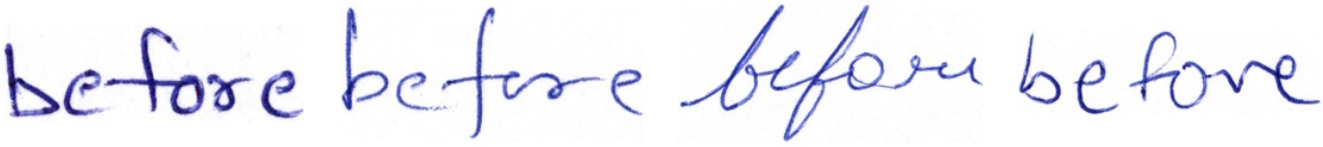
Run-Length Features: These features capture consecutive runs of black pixels (foreground) in rows or columns, but ignore stroke curvature, angles, or detailed contour shape.

We now present an analysis of hand-crafted features corresponding to individual handwriting styles in Table 3, and Figures 8, 9, 10, 11. Table 3 provides an overview of the



(a) Writer 6 (b) Writer 21 (c) Writer 70 (d) Writer 90

Figure 6. The word ‘right’ from writers 6, 21, 70, 90.



(a) Writer 14 (b) Writer 48 (c) Writer 77 (d) Writer 98

Figure 7. The word ‘before’ from writers 14, 48, 77, 99.

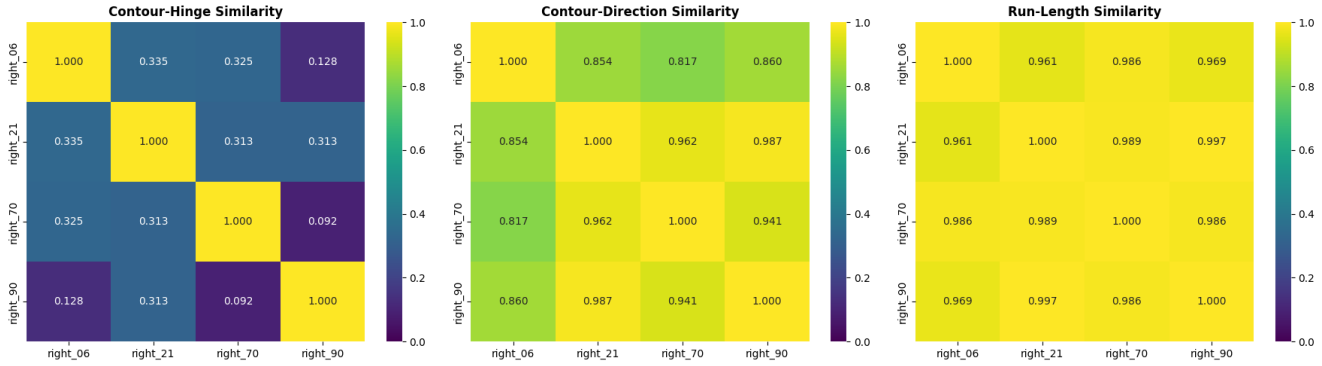


Figure 8. Cosine similarity for the word ‘right’ written by writers 06, 21, 70, 90 using handcrafted features.

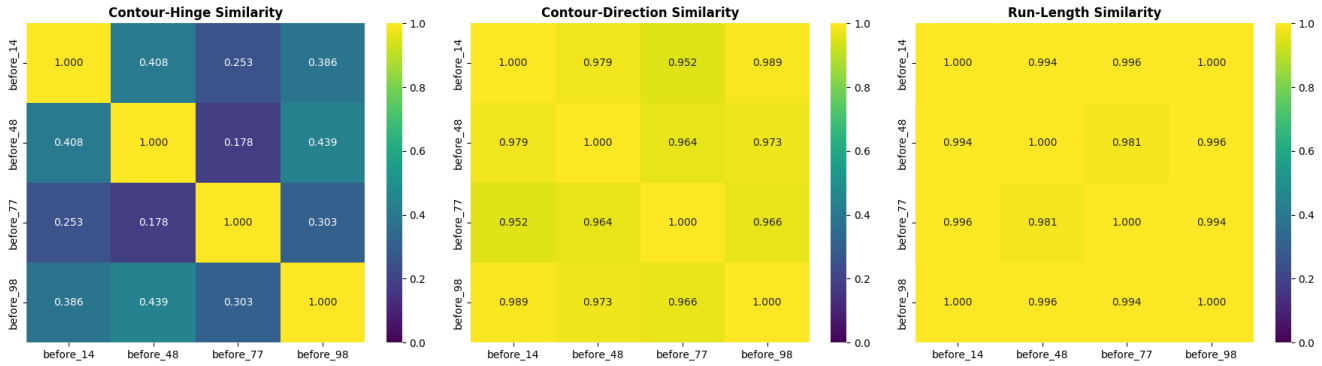


Figure 9. Cosine similarity for the word ‘before’ written by writers 14, 48, 77, 98 using handcrafted features.

handcrafted features with respect to dimension, computation methodology, etc. This overview table is inspired by [1], which has provided a comprehensive comparison of textural and allographic features used for writer identification and verification.

We now present the qualitative analysis of handcrafted feature selection. For a fair comparison, we use lexicographically identical words from four different writers. This enables evaluation of the discriminative power of the chosen handcrafted features. To indicate the word label and its corresponding writer ID in the respective figures, we use the template *<word-label>_<writer-id>*. Figure 6 shows the word ‘right’ written by writer IDs 6, 21, 70, and 90, respectively. Similarly, Figure 7 shows the word ‘before’,

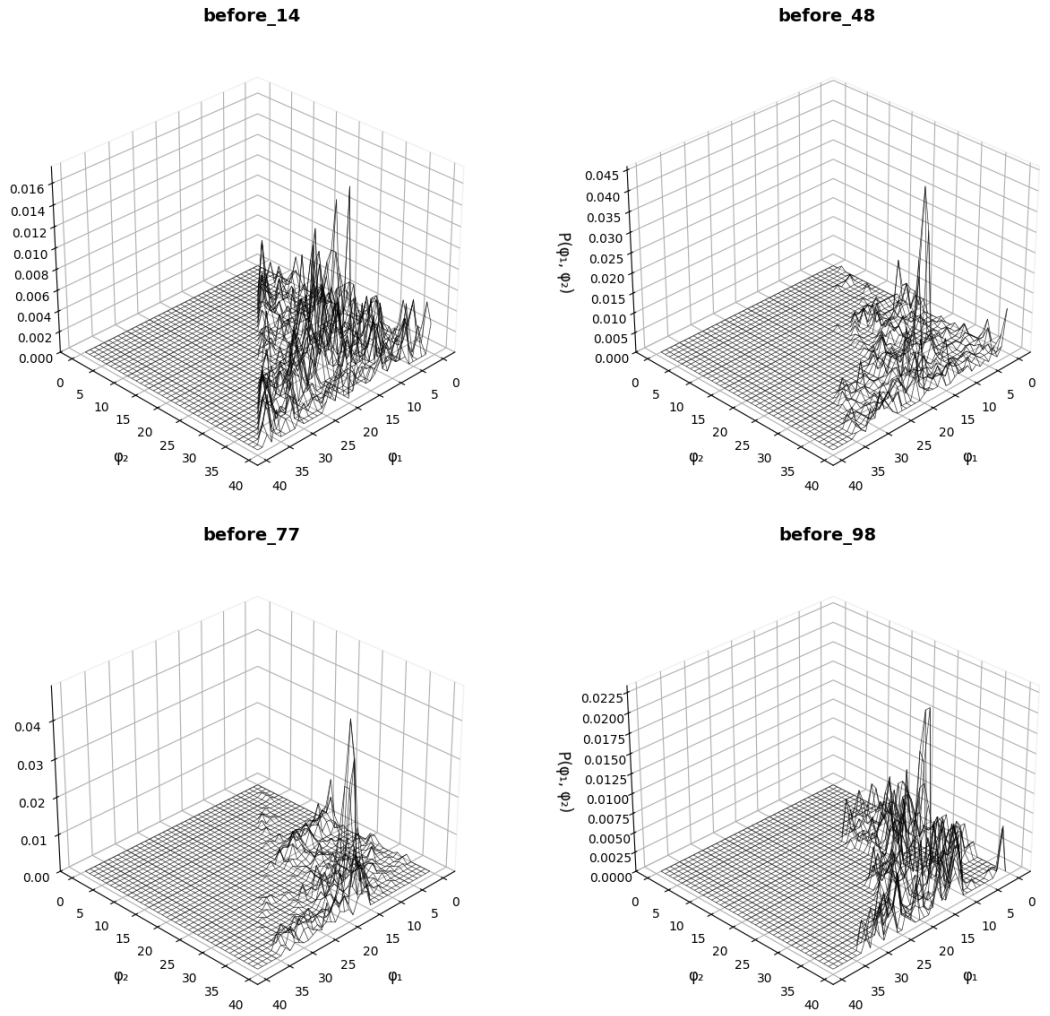


Figure 10. Surface plots of the contour-hinge PDF for the word 'before' written by Writers 14, 48, 77, and 98.

Table 3. Overview of engaged handcrafted features

Feature	Notation	Dimension	Computation
Contour-Hinge	$p(\phi_1, \phi_2)$	780	contours
Contour-Direction	$p(\phi)$	36	contours
Run-length	$p(rl)$	120	binary image

scribbled by writer IDs 14, 48, 77, and 99. The corresponding cosine similarity metrics are shown in Figures 8 and 9, respectively. The similarity matrix justifies the choice of contour-hinge as a handcrafted feature that corresponds to individual writer styles.

Figure 10 shows a 3D surface plot of the joint contour hinge probability density function $p(\phi_1, \phi_2)$ for the word *'before'* across writers 14, 48, 77, 98, illustrating how hinge angles capture variations in word shape and structure among writers.

References

[1] Marius Bulacu and Lambert Schomaker. Text-independent writer identification and verification using textural and allographic features. *IEEE Trans. on PAMI*, 29(4):701–717, 2007. 5

[2] P.Y. Simard, D. Steinkraus, and J.C. Platt. Best practices for convolutional neural networks applied to visual document analysis. In *ICDAR*, pages 958–963, 2003. 1

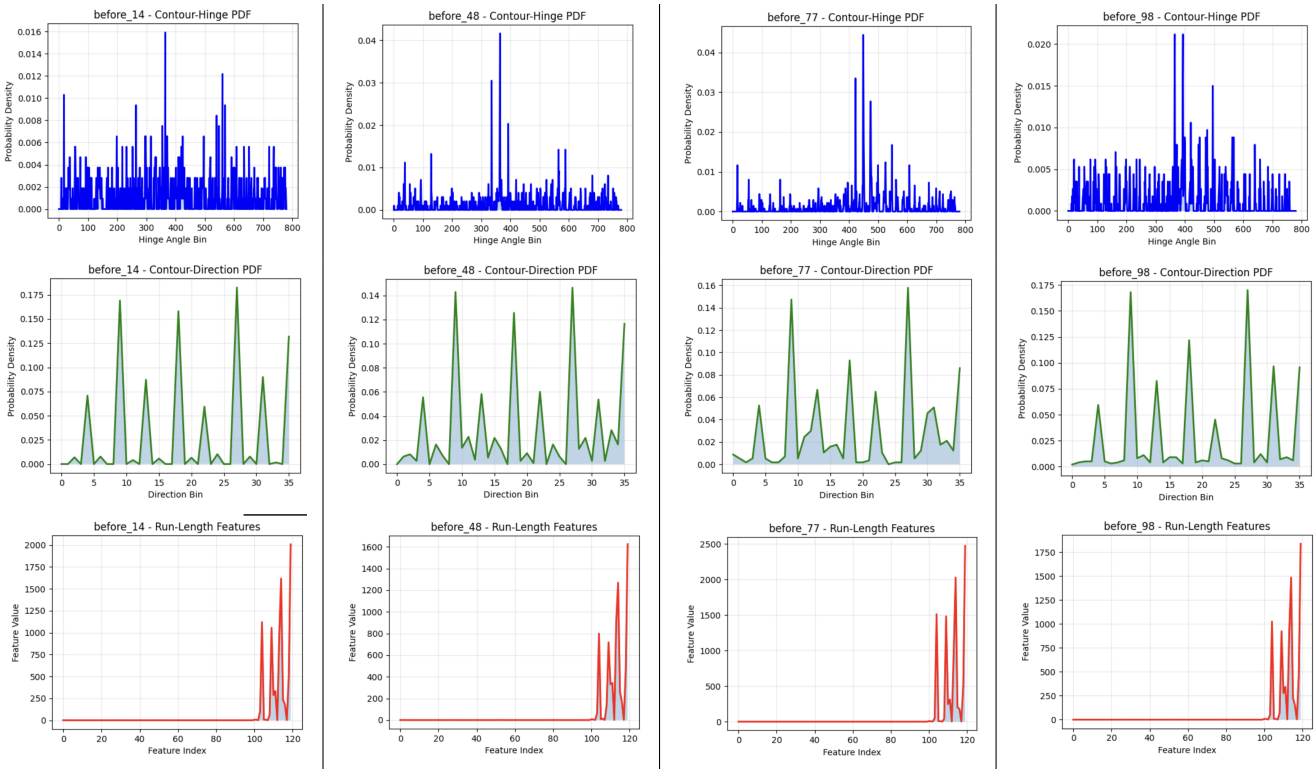


Figure 11. Variations observed in handcrafted features across words of Figure7 with respect to Writers 14, 48, 77, 98. *First row*: contour-hinge, *second row*: contour-direction, *third row*: run-length features.