

Handwriting Transformers

Supplementary Material

Ankan Kumar Bhunia¹ Salman Khan² Hisham Cholakkal¹ Rao Muhammad Anwer¹
Fahad Shahbaz Khan^{1,3} Mubarak Shah⁴

¹Mohamed bin Zayed University of AI, UAE ²Australian National University, Australia
³Linköping University, Sweden ⁴University of Central Florida, USA

In this supplementary material, we present additional human study details, additional qualitative results, and additional ablation study results. In Sec. 1, we provide details of human study experiments. Sec. 2 presents the additional visualisation results of transformer encoder-decoder attention maps. Sec. 3 presents additional quantitative results on RIMES [4] and CVL [6] datasets. Sec. 4 shows qualitative comparison of our proposed HWT. Sec. 5 presents handwritten text recognition results. Sec. 6 shows the interpolations between two different calligraphic styles on the IAM dataset. Finally, Sec. 7 presents additional ablation results and Sec. 8 presents some additional experiment results.

1. Human Study Additional Details

Here, we present results of our two user studies on 100 human participants to evaluate the human plausibility in terms of style mimicry of our proposed HWT. In both these user studies, the forged samples are generated using *unseen writing styles* of test set writers of IAM dataset, and for textual content we use sentences from Stanford Sentiment Treebank [7] dataset.

User Preference Study: Fig. S1 shows the interface for the *User preference study* experiment, which compares styled text images. In this study, each participant is shown a real handwritten text image of a person and the synthesized handwriting text images of that person using our proposed HWT, Davis *et al.* [1] and GANwriting [5]. We randomly present generated results of these methods to the user. Then, the user can compare the real image and the generated images side by side on the same screen and without any time restriction to give the answer. Each participant is required to provide response for a total of ten questions. Overall, we have collected 1000 responses from 100 participants. Table S1 shows the results of *User preference study*. Davis *et al.* [1] and GANwriting [5] were preferred 9% (90 responses out of the total 1000) and 10% (100 responses out of the total 1000), respectively. Our proposed HWT was preferred 81% (810 responses out of the total 1000 re-

Table S1: *User preference study* in comparison to GANwriting [5] and Davis *et al.* [1]. The result shows that our proposed HWT was preferred 81% of the time over the other two compared methods.

	Total Responses	User Preferences
GANwriting [5]		100
Davis <i>et al.</i> [1]	1000	90
HWT (Ours)		810

Table S2: Confusion matrix (%) obtained from *User plausibility study*. Only 48.1% of the images were correctly classified, indicating an output comparable to a random choice in a two-class problem.

Actual	Predicted	
	Real	Fake
Real	24.9	25.1
Fake	26.8	23.2

sponses) over the other two existing methods.

User Plausibility Study: Fig. S2 shows the interface for the *User plausibility study*, which evaluates the proximity of the synthesized samples generated by our proposed HWT to the real samples. Here, each participant is shown a person’s actual handwriting, followed by six samples, where each of these samples is either genuine or synthesized handwriting of the same person. Each participant received equal number of real and forged samples. Participants are asked to identify whether a given handwritten sample is genuine or not (forged/synthesized) with no time limit restriction to answer the question. In total, we collect 6000 responses for 100 human participants as each one provides 60 responses. The study revealed that the generated images produced by our proposed HWT were deemed plausible. Table S2 shows the confusion matrix of the human assessments. For this particular study, only 48.1% of the images have been correctly classified, which indicates a comparable performance

Table S3: Comparison of our HWT with GANwriting [5] in terms of FID scores computed between generated text images and real text images of RIMES [4] and CVL [6] datasets, respectively. HWT achieves promising results on both datasets. Best results are in bold.

Dataset	Method	IV-S	IV-U	OOV-S	OOV-U
RIMES	GANwriting [5]	101.30	105.32	115.65	118.78
	HWT (Ours)	93.65	95.43	103.78	107.05
CVL	GANwriting [5]	145.56	157.53	157.89	165.45
	HWT (Ours)	134.78	139.72	143.09	145.76

Table S4: Handwritten text recognition (HTR) results on CVL [6] dataset. We compare our HWT with HiGAN [3] and ScrabbleGAN [2] under identical settings, following the protocol used in HiGAN paper. Our HWT outperforms both HiGAN [3] and ScrabbleGAN [2], suggesting that generated images from HWT helps to achieve improved recognition results.

Method	Training Data			CVL(%)		CVLoov(%)	
	GAN	CVL	IAM	WER	CER	WER	CER
—	✗	✓	✓	29.41	13.13	37.63	17.16
HiGAN [3]	✓	✓	✓	28.91	12.54	37.06	16.67
ScrabbleGAN [2]	✓	✓	✓	28.68	12.13	37.10	16.73
HWT (Ours)	✓	✓	✓	27.81	11.84	36.47	15.95

to random choice in a two-class problem.

2. Additional Visualizations of Transformer Encoder-Decoder Attention

Fig. S3 shows the visualization of attention maps obtained using encoder-decoder of our HWT at the last layer of the transformer decoder. We compute the attention matrices for four different words: ‘laughs’, ‘because’, ‘inside’, and ‘fashion’. Note that the attention maps generated by our model focus on the relevant regions of interest in the style examples for each query character. For instance, to infer character-specific style attributes of a given character ‘h’ in the query word ‘laughs’, the model gives priority to multiple image regions containing the character ‘h’. Note that if the query character isn’t found in the style examples, the model attempts to find similar characters. For example, to obtain character representation of ‘u’ in the query word ‘laughs’, the attention algorithm highlights image regions containing similar characters (*e.g.* ‘n’).

3. Additional Quantitative Results on RIMES and CVL Datasets

We present additional qualitative results on RIMES [4] and CVL [6] datasets for styled handwritten text generation. The RIMES dataset contains over 60k images written by 1300 distinct writers, all of which are in the French lan-

guage. The CVL dataset consists of seven handwritten documents. About 310 people contributed to these texts, resulting in approximately 83k word crops separated into train and test sets. Tab. S3 shows that our proposed HWT performs favorably against GANwriting [5] on both datasets.

4. Additional Qualitative Comparison

Figs. S4-S21 show qualitative comparison between our proposed HWT with [5, 1] for styled handwritten text generation. Note that we use the same textual content for all the examples figures for all the three methods to ensure a fair comparison. The first row in each figure presents the different writers example style images. The rest of the rows correspond to our HWT and [5, 1] respectively. The qualitative results suggest that our method is promising at imitating character-level patterns, while the other two methods struggle to retain character-specific details. The success of the other two methods is limited to capturing only the global patterns (*e.g.*, slant, ink widths). In some cases, these methods even struggle to capture global styles. In Fig. S6, Fig. S16 and Fig. S18, Davis *et al.* [1] suffer to capture the slant. Whereas, in Fig. S16 and Fig. S20, the ink width of the images generated by this method is not consistent with the style examples. On the other hand, since GANwriting [5] is limited to a fixed length query words, it is unable to complete few words that exceed the limit.

Figs. S22-S23 show qualitative results using the same text as in the style examples to compare our proposed HWT with [5, 1]. Figs. S24-S26 show examples, where we aim to generate arbitrarily long words. The results show that our model is capable of consistently imitating the styles of the given style example, even for arbitrarily long words. Note that GANwriting [5] struggles to generate long words.

5. Improving Handwritten Text Recognition

We utilize our generated samples for Handwritten Text Recognition (HTR) training to validate if the generated images can help improve recognition performance. Tab. S4 shows recognition results on CVL dataset. We compare with HiGAN [3] and ScrabbleGAN [2] under identical settings, following HTR protocol used in HiGAN. Our HWT performs favorably against both HiGAN and ScrabbleGAN for HTR task.

6. Latent Space Interpolations

Fig. S27 shows interpolations between two different calligraphic styles on the IAM dataset. To interpolate by λ between two writers A and B , we compute the weighted average $Z_{AB} = \lambda Z_A + (1 - \lambda) Z_B$, while keeping the textual contents fixed. Here, Z_A and Z_B are the style feature sequences obtained from encoder T_E . It is worth mentioning that our models produce images seamlessly by adjust-

ing from one style to other, which indicates that our model generalizes in the latent space rather than memorizing any trivial writing patterns.

7. Additional Ablation Results

Fig. S28 presents additional qualitative results that show the impact of integrating transformer encoder (Enc), transformer decoder (Dec) and cycle loss (CL) to the baseline (Base). Fig. S29 shows additional qualitative comparisons between word-level and character-level conditioning.

8. Additional Experiments

Source image requirements: We vary #source style examples ($P = [5:5:25]$) and compare with GANwriting [14]. We observe favorable results against [5] at all P values with consistent gains of over 12% in FID score.

Comparison with HiGAN [3]: For fair comparison with [1, 2], we follow their experimental setup [1] in Tab. 2 of our main paper, which is agnostic to writing style mimicry. HiGAN uses a slightly different evaluation setup, where both writing style mimicry and content consistency are considered. For fair comparison with HiGAN [3], we perform three experiments and report FID scores. We evaluate (i) HWT in HiGAN evaluation setting (HiGAN: 17.28 vs. **Ours: 14.33**), (ii) HiGAN in our Tab. 2 settings (HiGAN: 21.56 vs. **Ours: 19.40**) (iii) HiGAN in one of the most challenging settings (OOV-U) in our Tab. 1 (HiGAN: 129.89 vs. **Ours: 114.10**). In *all* three experiments, we observe that our proposed HWT performs favourably against HiGAN.

Line-level FID Comparison: In Tab. 2 of our main paper, similar to [2], we compute the FID scores at word-level. Also, we perform line-level evaluation by concatenating word images and obtain promising (Davis *et al.* [1]: 20.65 vs. **Ours: 19.48**).

References

- [1] Brian Davis, Chris Tensmeyer, Brian Price, Curtis Wigington, Bryan Morse, and Rajiv Jain. Text and style conditioned gan for generation of offline handwriting lines. *BMVC*, 2020.
- [2] Sharon Fogel, Hadar Averbuch-Elor, Sarel Cohen, Shai Mazor, and Roee Litman. Scrabblegan: semi-supervised varying length handwritten text generation. In *CVPR*, pages 4324–4333, 2020.
- [3] Ji Gan and Weiqiang Wang. Higan: Handwriting imitation conditioned on arbitrary-length texts and disentangled styles. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 7484–7492, 2021.
- [4] Emmanuele Grosicki and Haikal El Abed. Icdar 2009 handwriting recognition competition. In *2009 10th International Conference on Document Analysis and Recognition*, pages 1398–1402. IEEE, 2009.
- [5] Lei Kang, Pau Riba, Yaxing Wang, Marçal Rusiñol, Alixia Fornés, and Mauricio Villegas. Ganwriting: Content-conditioned generation of styled handwritten word images. In *ECCV*, pages 273–289. Springer, 2020.
- [6] Florian Kleber, Stefan Fiel, Markus Diem, and Robert Sablatnig. Cvl-database: An off-line database for writer retrieval, writer identification and word spotting. In *2013 12th international conference on document analysis and recognition*, pages 560–564. IEEE, 2013.
- [7] Richard Socher, Alex Perelygin, Jean Wu, Jason Chuang, Christopher D Manning, Andrew Y Ng, and Christopher Potts. Recursive deep models for semantic compositionality over a sentiment treebank. In *EMNLP*, pages 1631–1642, 2013.

Figure S1: Screenshot of the Interface used in *User preference study* experiment. Each participant is shown the real handwritten text image (on the left side) of a person and synthesized handwriting text images (on the right side) of that person generated using three different methods. Participants have to mark the best method for mimicking the real handwriting style.

User preference study: Instructions

1. The image on the left side shows a real example of handwriting style of a person. 2. The images on the right side are generated by computer using three different methods (randomly ordered). Please note that the textual content in the right images is different from the left image. Question: Which one is better at mimicking the handwriting style of the left image? Please give your response by clicking on the checkboxes.

Cecil searched in his pockets once more and came
paper-clips after a few seconds of hunting
a bent wire loop into the lock and jangled it

- Shaky closeups of turkeyonrolls stubby chins liver spots red noses and the filmmakers new bobbed do draw easy chuckles but lead
- Shaky closeups of turkeyonrolls stubby chins liver spots red noses and the filmmakers new bobbed do draw easy chuckles but lead
- Shaky closeup of turkeyo stubby chins spots red nose and the filmak new bobbed do draw easy chuckle but lead

Next [0/10]

Figure S2: Screenshot of the Interface used in *User plausibility study* experiment. Each participant is shown a person's actual handwriting (on the left side), followed by six samples (on the right side), where three out of these samples are genuine and the rest are synthesized. Participants have to classify each sample as genuine or forgery by looking at the real image.

User plausibility study: Instructions

1. The image on the left side shows a real example of handwriting style of a person. 2. The images on the right side are either real handwriting or forged handwriting of that person. In total there are three real images and three forged images. You have to identify the forged samles out of these images.

that sinister yarn in which half the cast try to
persuade

- heroine that she is out of her mind Despite flagrant cheating
- Director Andrew Niccol ... demonstrates a way understanding of the quirks of fame
- Happy Times maintains an appealing veneer without becoming too cute about it
- He has earned his break The film is a well-made variation
- eerie atmosphere is built up nearly Susan Strasberg is the crippled
- A yarn that respects the Marvel version without becoming ensnared by it

Next [0/10]

Figure S3: Additional visualization results of encoder-decoder attention maps at the last layer of the transformer decoder. The attention maps are computed for four different query words: ‘laughs’, ‘because’, ‘inside’, and ‘fashion’. Heat maps corresponding to all characters (including repetitions, as the letter ‘i’ appears twice in ‘inside’) of these query words are shown in the figure.

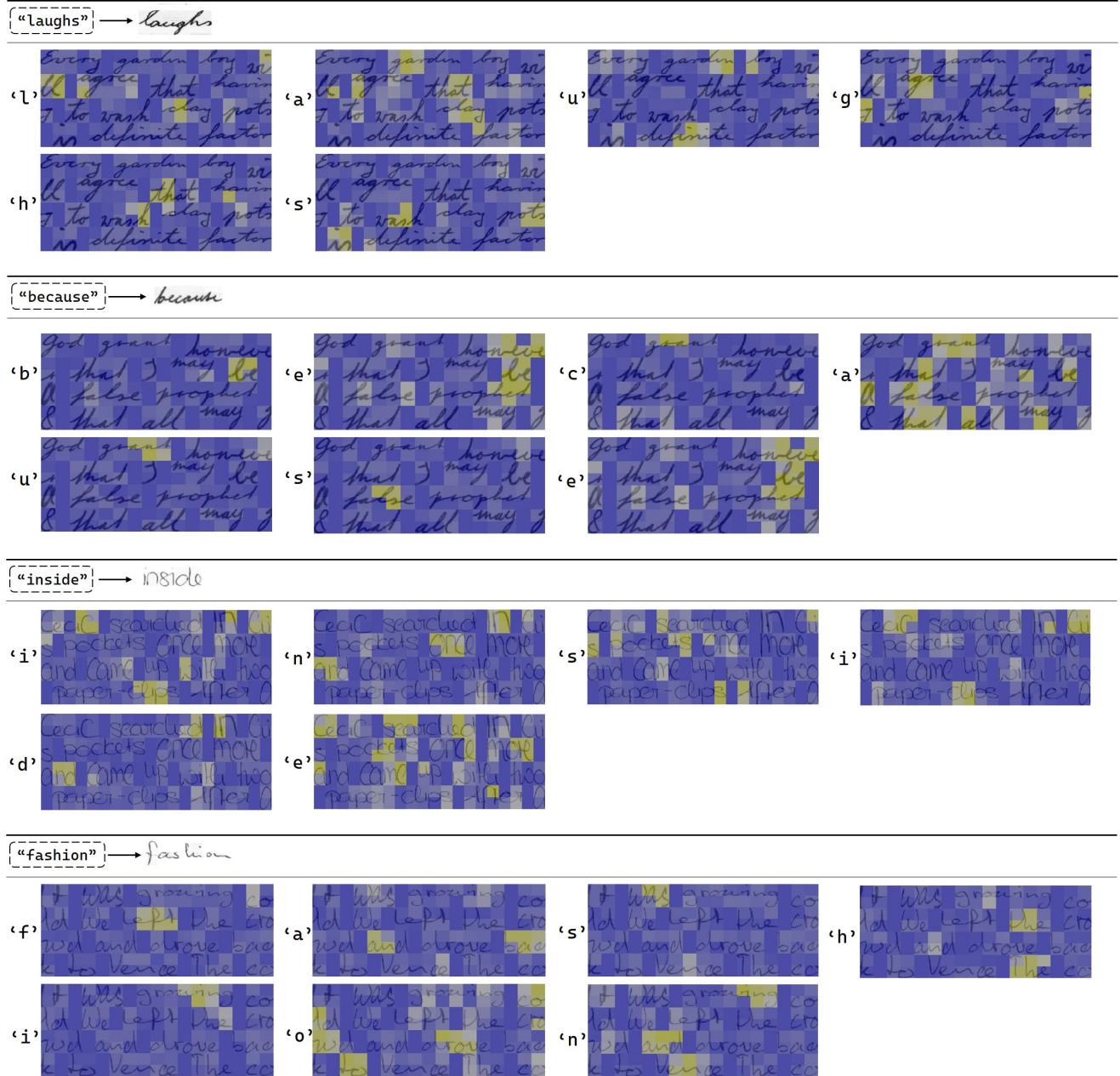


Figure S4: Additional qualitative comparisons of our proposed HWT with GANwriting [5] and Davis *et al.* [1], when generating the same text ‘With more character development this might have been an eerie thriller with better payoffs it could have been a thinking’.

Style	sorts of reasons or you might “leave him” But how utterly societies for that writer will allow for the wife to
HWT (Ours)	With more character development this might have been an eerie thriller with better payoffs it could have been a thinking
GANwriting	With more character develop this might have been an eerie thriller with better payoffs it could have been a thinkin
Davis et al.	With more character development this might have been an eerie thriller with better payoffs it could have been a thinking

Figure S5: Additional qualitative comparisons of our proposed HWT with GANwriting [5] and Davis *et al.* [1], when generating the same text ‘It’s not helpful to listen to extremist namecalling regardless of whether you think Kissinger was a calculating’.

Style	between the do-it-yourself cupboard and the polished brass fourteen-pounder shell-case which sorted respectively to one
HWT (Ours)	It’s not helpful to listen to extremist namecalling regardless of whether you think Kissinger was a calculating
GANwriting	It’s not helpful to listen to extremi namecall regardl of whether you think Kissing was a calcula
Davis et al.	It’s not helpful to listen to extremist namecalling regardless of whether you think Kissinger was a calculating

Figure S6: Additional qualitative comparisons of our proposed HWT with GANwriting [5] and Davis *et al.* [1], when generating the same text ‘Shaky closeups of turkeyonrolls stubby chins liver spots red noses and the filmmakers new bobbed do draw easy chuckles but’.

Style	Cecil searched in his pockets once more and came up with two paper-clips after a few seconds of hoisting he roughly thrust
HWT (Ours)	Shaky closeups of turkeyonrolls stubby chins liver spots red noses and the filmmakers new bobbed do draw easy chuckles but
GANwriting	Shaky closeup of turkeyo stubby chins liver spots red noses and the filmmak new bobbed do draw easy chuckle but
Davis et al.	Shaky closeups of turkeyonrolls stubby chins liver spots red noses and the filmmakers new bobbed do draw easy chuckles but

Figure S7: Additional qualitative comparisons of our proposed HWT with GANwriting [5] and Davis *et al.* [1], when generating the same text '*This film was made by and for those folks who collect the serial killer cards and are fascinated by the mere suggestion*'.

Style	He read the film star's sorry story and frowned at the provisions of Schedule D taxation which not only allowed his
HWT (Ours)	<i>This film was made by and for those folks who collect the serial killer cards and are fascinated by the mere suggestion</i>
GANwriting	<i>This film was made by and for those folks who collect the serial killer cards and are fascina by the mere suggest</i>
Davis et al.	<i>This film was made by and for those folks who collect the serial killer cards and are fascinated by the mere suggestion</i>

Figure S8: Additional qualitative comparisons of our proposed HWT with GANwriting [5] and Davis *et al.* [1], when generating the same text '*Its a drawling slobbering lovable runon sentence of a film a Southern Gothic with the emotional arc of its raw blues*'.

Style	Compared with 1958 the expenditure index for 1959 showed increases of 4 to 6 per cent for couples without children and
HWT (Ours)	<i>It a drawling slobbering lovable runon sentence of a film Southern Gothic with the emotional arc of its raw blues</i>
GANwriting	<i>It is a drawlin slobber lovable runon sentence of a film Southern Gothic with the emotion arc of its raw blues</i>
Davis et al.	<i>It s a drawling slobbering lovable runon sentence of a film a Southern Gothic with the emotional arc of its raw blues</i>

Figure S9: Additional qualitative comparisons of our proposed HWT with GANwriting [5] and Davis *et al.* [1], when generating the same text '*LRB W RRB hile long on amiable monkeys and worthy environmentalism Jane Goodalls Wild Chimpanzees is short*'.

Style	In April of that year his first wife's brother-in-law the diplomatist Lord Ponsonby had written to advise Chafee
HWT (Ours)	<i>LRB W RRB hile long on amiable monkeys and worthy environmentalism Jane Goodall i Wild Chimpanzees is short</i>
GANwriting	<i>LRB W RRB hile long on amiable monkeys and worthy environ Jane Goodall & Wild Chimpas is short</i>
Davis et al.	<i>LRB W RRB hile long on amiable monkeys and worthy environmentalism Jane Goodall s Wild Chimpanzees is short</i>

Figure S10: Additional qualitative comparisons of our proposed HWT with GANwriting [5] and Davis *et al.* [1], when generating the same text '*For close to two hours the audience is forced to endure three terminally depressed mostly inarticulate hyper dysfunctional*'

Style	give the system a trial adding that it was being cultivated with extraordinary success in France and Italy and that he
HWT (Ours)	<i>For close to two hours the audience is forced to endure three terminally depressed mostly inarticulate hyper dysfunctional</i>
GANwriting	<i>For close to two hours the audience is forced to endure three terminally depressed mostly inarticulate hyper dysfunctional</i>
Davis et al.	<i>For close to two hours the audience is forced to endure three terminally depressed mostly inarticulate hyper dysfunctional</i>

Figure S11: Additional qualitative comparisons of our proposed HWT with GANwriting [5] and Davis *et al.* [1], when generating the same text '*Claude Chabrol's camera has a way of gently swaying back and forth as it cradles its characters veiling tension beneath*'.

Style	Are there differences in adjustment to ageing and refinement according to the occupational level of employees
HWT (Ours)	<i>Claude Chabrol's camera has a way of gently swaying back and forth as it cradles its characters veiling tension beneath</i>
GANwriting	<i>Claude Chabrol's camera has a way of gently swaying back and forth as it cradles its characters veiling tension beneath</i>
Davis et al.	<i>Claude Chabrol's camera has a way of gently swaying back and forth as it cradles its characters veiling tension beneath</i>

Figure S12: Additional qualitative comparisons of our proposed HWT with GANwriting [5] and Davis *et al.* [1], when generating the same text '*Though the plot is predictable the movie never feels formulaic because the attention is on the nuances of the*'.

Style	None of the numerous conventional remedies to which he had been subjected till since the symptoms had first shown
HWT (Ours)	<i>Though the plot is predictable the movie never feels formulaic because the attention is on the nuances of the</i>
GANwriting	<i>Though the plot is predictable the movie never feels formulaic because the attention is on the nuances of the</i>
Davis et al.	<i>Though the plot is predictable the movie never feels formulaic because the attention is on the nuances of the</i>

Figure S13: Additional qualitative comparisons of our proposed HWT with GANwriting [5] and Davis *et al.* [1], when generating the same text ‘*A comingofage tale from New Zealand whose boozy languid air is balanced by a rich visual clarity and deeply felt*’.

Style	As a result the Glasgow Retirement Council came into being in April 1958 with Dr. Andrew Hood as chairman and Mr. Andrew
HWT (Ours)	A comingofage tale from New Zealand whose boozy languid air is balanced by a rich visual clarity and deeply felt
GANwriting	A comingo tale from New Zealand whose boozy languid air is balance by a rich visual clarity and deeply felt
Davis <i>et al.</i>	A comingofage tale from New Zealand whose boozy languid air is balanced by a rich visual clarity and deeply felt

Figure S14: Additional qualitative comparisons of our proposed HWT with GANwriting [5] and Davis *et al.* [1], when generating the same text ‘*Unfortunately Kapur modernizes AEW. Masons story to suit the sensibilities of a young American a decision that plucks The*’.

Style	The problems of old ageing & already in retirement but also bring desire the part of all concerned for concerted action
HWT (Ours)	Unfortunately Kapur modernizes AEW Mason's story to suit the sensibilities of a young American a decision that plucks The.
GANwriting	Unfortunately Kapur moderni AEW Mason's story to suit the sensibi of a young America a decisio that plucks The
Davis <i>et al.</i>	Unfortunately Kapur modernizes AEW Mason's story to suit the sensibilities of a young American a decision that plucks The

Figure S15: Additional qualitative comparisons of our proposed HWT with GANwriting [5] and Davis *et al.* [1], when generating the same text ‘*Unless Bob Crane is someone of particular interest to you this films impressive performances and adept direction are*’.

Style	The Rubery Owen scheme is MW in its fourth year and opportunity has been taken to revise the course in the light
HWT (Ours)	Unless Bob Crane is someone of particular interest to you this film's impressive performances and adept direction are
GANwriting	Unless Bob Crane is someone of particluar interes to you this film's impress perform and adept directi are
Davis <i>et al.</i>	Unless Bob Crane is someone of particular interest to you this film's impressive performances and adept direction are

Figure S16: Additional qualitative comparisons of our proposed HWT with GANwriting [5] and Davis *et al.* [1], when generating the same text ‘Affirms the gifts of all involved starting with Spielberg and going right through the ranks of the players oncamera and off’.

Style	nevertheless identify it. Similarly the psychologist has to be prepared to observe and make inferences about all kinds of
HWT (Ours)	Affirms the gifts of all involved starting with Spielberg and going right through the ranks of the players oncamera and off
GANwriting	Affirms the gifts of all involve startin with Spielberg and going right through the ranks of the players oncamer and off
Davis <i>et al.</i>	Affirms the gifts of all involved starting with Spielberg and going right through the ranks of the players oncamera and off

Figure S17: Additional qualitative comparisons of our proposed HWT with GANwriting [5] and Davis *et al.* [1], when generating the same text ‘Though this rude and crude film does deliver a few gut-busting laughs its digs at modern society are all things we ve seen’.

Style	headlines for having smashed large number + to relieve his feelings On the small-to-medium establishment it is
HWT (Ours)	Though this rude and crude film does deliver a few gutbusting laughs its digs at modern society are all things we ve seen
GANwriting	Though this rude and crude film does deliver a few gutburst laughs its digs at modern society are all things we ve seen
Davis <i>et al.</i>	Though this rude and crude film does deliver a few gutbusting laughs its digs at modern society are all things we ve seen

Figure S18: Additional qualitative comparisons of our proposed HWT with GANwriting [5] and Davis *et al.* [1], when generating the same text ‘You ll laugh at either the obviousness of it all or its stupidity or maybe even its inventiveness but the point is’.

Style	He had long sensed injustice in the distinctions drawn between ordinary wage-earners and those self-employed By
HWT (Ours)	You ll laugh at either the obviousness of it all or its stupidity or maybe even its inventiveness but the point is
GANwriting	You ll laugh at either the obvious of it all or its stupiditi or maybe even its inventi but the point is
Davis <i>et al.</i>	You ll laugh at either the obviousness of it all or its stupidity or maybe even its inventiveness but the point is

Figure S19: Additional qualitative comparisons of our proposed HWT with GANwriting [5] and Davis *et al.* [1], when generating the same text ‘*Writerdirector s Mehta s effort has tons of charm and the whimsy is in the mixture the intoxicating masala of cultures*’.

Style	avoided by the installation of a power tool As an elementary example one can enjoy rapping out parts from hardship text
HWT (Ours)	Writerdirector s Mehta s effort has tons of charm and the whimsy is in the mixture the intoxicating masala of cultures
GANwriting	Writer s Mehta s effort has tons of charm and the whimsy is in the mixture the intoxicating masala of culture
Davis <i>et al.</i>	Writerdirector s Mehta s effort has tons of charm and the whimsy is in the mixture the intoxicating masala of cultures

Figure S20: Additional qualitative comparisons of our proposed HWT with GANwriting [5] and Davis *et al.* [1], when generating the same text ‘*While easier to sit through than most of Jaglom s selfconscious and gratingly irritating films it s*’.

Style	mystically returned to the shed their books tidied and even their shoes cleaned ? How showed their this morning with box
HWT (Ours)	While easier to sit through than most of Jaglom s selfconscious and gratingly irritating films it s
GANwriting	While easier to sit through than most of Jaglom s selfcon and grating irritat films it s
Davis <i>et al.</i>	While easier to sit through than most of Jaglom s selfconscious and gratingly irritating films it s

Figure S21: Additional qualitative comparisons of our proposed HWT with GANwriting [5] and Davis *et al.* [1], when generating the same text ‘*The connected stories of Breitbart and Hanussen are actually fascinating but the filmmaking in Invincible is such that the*’.

Style	To carry brought Tory policy Gaitshill's stupid hope The tragedy is that enormous inroads could already have been made
HWT (Ours)	The connected stories Of Breitbart And Hanussen are actually fascinating but the filmmaking in Invincible is such that the
GANwriting	The connect stories of Breitba and Hanusee are actuall fascina but the filmmak in Invinci is such that the
Davis <i>et al.</i>	The connected stories of Breitbart and Hanussen are actually fascinating but the filmmaking in Invincible is such that the

Figure S22: Reconstruction results using the proposed HWT in comparison to GANwriting [5] and Davis *et al.* [1]. We use the same text as in the style examples to generate handwritten images.

Style	few people in the bar elderly well-off artistic who you felt had made
HWT (Ours)	few people in the bar elderly well-off artistic who you felt had made
GANwriting	few people in the bar elderly welloff artisti who you felt had made
Davis et al.	few people in the bar elderly well-off artistic who you felt had made
Style	Fortunately, however the fashion for Victorian architecture which Mr. John Betjeman had started several decades
HWT (Ours)	Fortunately however the fashion for Victorian architecture which Mr. John Betjeman had started several decades
GANwriting	Fortuna however the fashion for Victori archite which Mr John Betjema had started several decades
Davis et al.	Fortunately however the fashion for Victorian architecture which Mr. John Betjeman had started several decades
Style	families with three child for whom the index declined from index which compares the prices
HWT (Ours)	families with three child for whom the index declined from index which compares the -rices
GANwriting	familie with three children for whom the index decline from index which compare the prices
Davis et al.	families with three child for whom the index declined from index which compares the prices
Style	They closed on a single bundle and fumbling with nervous excitement he pulled it out.
HWT (Ours)	They closed on a single bundle and fumbling with nervous excitement he pulled it out
GANwriting	They closed on a single bundle and fumbling with nervous escitem he pulled it out
Davis et al.	They closed on a single bundle and fumbling with nervous excitement he pulled it out
Style	Mauro's first action was to write to his revered master at Leipzig asking for
HWT (Ours)	Mauro's first action was to write to his revered master at Leipzig asking for
GANwriting	Mauros first action was to write to his revered master at Leipzig asking for
Davis et al.	Mauro's first action was to write to his revered master at Leipzig asking for
Style	described his distinguished patient and his symptoms He told Hahnemann that he
HWT (Ours)	described his distinguished patient and his symptoms He told Hahnemann that he
GANwriting	describ his disting patient and his symptom He told Hahnema that he
Davis et al.	described his distinguished patient and his symptoms He told Hahnemann that he
Style	This remarkable of medicine whom Sir Francis Burdett described to Anglesey
HWT (Ours)	This remarkable of medicine whom Sir Francis Burdett described to Anglesey
GANwriting	This remaika of medicin whom Sir Francis Burdett describ to Anglese
Davis et al.	This remarkable of medicine whom Sir Francis Burdett described to Anglesey

Figure S23: Reconstruction results using the proposed HWT in comparison to GANwriting [5] and Davis *et al.* [1].

Style	Mauro's first action was to write to his revered master at Leipzig asking for
HWT (Ours)	Mauro's first action was to write to his revered master at Leipzig asking for
GANwriting	Mauro's first action was to write to his revered master at Leipzig asking for
Davis et al.	Mauro's first action was to write to his revered master at Leipzig asking for
Style	In April of that year his first wife's brother-in-law the diplomatist
HWT (Ours)	In April of that year his first wife's brother-in-law the diplomatist
GANwriting	In April of that year his first wife's brother the diploma
Davis et al.	In April of that year his first wife's brother-in-law the diplomatist
Style	What a frightful event he wrote I tremble indifferent but I really tremble
HWT (Ours)	What a frightful event he wrote I tremble indifferent but I really tremble
GANwriting	What a frightful event he wrote to tremble indifferent but I really tremble
Davis et al.	What a frightful event he wrote I tremble indifferent but I really tremble
Style	In those early years life was very full both in the parish and in the wider war activities
HWT (Ours)	In those early years life was very full both in the parish and in the wider war activities
GANwriting	In those early years life was very full both in the parish and in the wider war activit
Davis et al.	In those early years life was very full both in the parish and in the wider war activities
Style	He may never have had the disease himself but he can nevertheless identify it
HWT (Ours)	He may never have had the disease himself but he can nevertheless identify it
GANwriting	He may never have had the disease himself but he can neverth identif it
Davis et al.	He may never have had the disease himself but he can nevertheless identify it
Style	The large attendance and atmosphere of this Conference held in October
HWT (Ours)	The large attendance and atmosphere of this Conference held in October
GANwriting	The large attenda and atmosph of this Confere held in october
Davis et al.	The large attendance and atmosphere of this Conference held in October
Style	They are slightly more difficult to manage however until a little
HWT (Ours)	They are slightly more difficult to manage however until a little
GANwriting	They are slightl more difficu to manage however until a little
Davis et al.	They are slightly more difficult to manage however until a little

Figure S24: Handwritten text image generation of arbitrarily long words. We generate the 21-letter word ‘Incomprehensibilities’ in three different styles and compare the results with Davis *et al.* [1].

Style	Method	Generated Images
atmosphere	Davis <i>et al.</i>	Incomprehensibilities
	HWT (Ours)	Incomprehensibilities
affiliations	Davis <i>et al.</i>	Incomprehensibilities
	HWT (Ours)	Incomprehensibilities
individuality	Davis <i>et al.</i>	Incomprehensibilities
	HWT (Ours)	Incomprehensibilities

Figure S25: Handwritten text image generation of arbitrarily long words. We generate the 30-letter word ‘Pseudopseudohypoparathyroidism’ in three different styles and compare the results with Davis *et al.* [1].

Style	Method	Generated Images
overdrafts	Davis <i>et al.</i>	Pseudopseudohypoparathyroidism
	HWT (Ours)	Pseudopseudohypoparathyroidism
betrayed	Davis <i>et al.</i>	Pseudopseudohypoparathyroidism
	HWT (Ours)	Pseudopseudohypoparathyroidism
carefully	Davis <i>et al.</i>	Pseudopseudohypoparathyroidism
	HWT (Ours)	Pseudopseudohypoparathyroidism

Figure S26: Handwritten text image generation of arbitrarily long words. We generate the 28-letter word ‘Antidisestablishmentarianism’ in three different styles and compare the results with Davis *et al.* [1].

Style	Method	Generated Images
congratulation	Davis <i>et al.</i>	Antidisestablishmentarianism
	HWT (Ours)	Antidisestablishmentarianism
delighted	Davis <i>et al.</i>	Antidisestablishmentarianism
	HWT (Ours)	Antidisestablishmentarianism
Inspector	Davis <i>et al.</i>	Antidisestablishmentarianism
	HWT (Ours)	Antidisestablishmentarianism

Figure S27: Latent space interpolations between calligraphic styles on the IAM dataset. The first and last image in each column correspond to writing styles of two different writers. Total we have shown five sets of interpolation results. We observe how the generated images seamlessly adjust from one style to another. This result shows that our model can generalize in the latent space rather than memorizing any trivial writing patterns.



Figure S28: Additional qualitative ablation of integrating transformer encoder (Enc), transformer decoder (Dec) and cycle loss (CL) to the baseline (Base) on the IAM dataset. We show the effect of each component when generating six different words ‘especially’, ‘ethereal’, ‘emotional’, ‘standard’, ‘resorts’, and ‘under’.

Style examples →	outwardly	adult	expectations	discovered	revered	content
Base	especially	ethereal	emotional	standard	resorts	under
Base + Enc	especially	ethereal	emotional	standard	resorts	under
Base + Dec	especially	ethereal	emotional	standard	resorts	under
Base + Enc + Dec	especially	ethereal	emotional	standard	resorts	under
Base + Enc + Dec + CL	especially	ethereal	emotional	standard	resorts	under

Figure S29: Additional qualitative comparisons between word and character-level conditioning on IAM dataset. We show the comparison between word and character-level conditioning when generating six different words ‘engaging’, ‘actually’, ‘movie’, ‘rhythms’, ‘what’, and ‘evocative’.

Style examples →	fragile	comically	within	attacked	what	responding
Word-level	engaging	actually	movie	rhythms	what	evocative
Character-level	engaging	actually	movie	rhythms	what	evocative