Handwriting Generation and Animation with Deep Learning

Anita Dash
MA20BTECH11001

Dhruv Srikanth
EE20BTECH11014

Shambhu Prasad Kavir
CS20BTECH11045

Taha Adeel Mohammed
CS20BTECH11052

Abstract

Handwriting style transfer for personalized text generation is a fascinating and evolving field that sits at the intersection of computer vision and natural language processing. It has several applications in various domains, ranging from enhancing digital communication and marketing to fostering creativity in art and design, and from improving accessibility and education to preserving historical documents and providing unique customization in merchandise and entertainment. In this project, we aim to create a system that takes a written text prompt and a sample of a person's or a specific handwriting style's characteristics as input and generates text and a sequence of brush strokes that closely resembles the provided style while maintaining readability and coherence.

1. Introduction

Handwriting is a deeply personal and authentic form of expression. It holds a unique place in human communication and self-expression. In this regard, Handwriting Style Transfer can come in handy. A model that generates handwritten text in a particular style can be helpful in the following fields

- In education, personalized handwritten materials can improve student engagement and retention, offering educators a powerful tool to make learning more effective and enjoyable.
- Handwriting style transfer can improve accessibility for individuals with motor disabilities, allowing them to convey their thoughts and emotions through personalized handwriting styles, promoting inclusivity in communication.
- Handwritten image generation can also provide additional data to train more accurate general handwriting recognition models
- It can help in the field of forensic analysis such as forgery detection, document authentication etc.
- Businesses can leverage personalized handwriting in marketing materials, enhancing customer engagement, and creating memorable brand experiences.

2. Problem Statement

Implement a deep learning model that can take a text prompt and a sample image of a person's handwriting style as input, and generate the image of the text prompt, along with the temporal sequence of brush strokes, in the specified handwriting style as the output. We want the model to accurately mimic the unique characteristics, nuances, and idiosyncrasies of the provided handwriting style while maintaining the readability and coherence of the generated text.

3. Literature Review

3.1. GANwriting

3.2. Handwriting Transformers

Earlier Handwriting generative methods process style and features separately. It doesn't encode style content entangelement at a character level. In this paper [5]. the authors propose a transformer-based styled handwritten text image generation approach, HWT, that strives to learn both style-content entanglement as well as global (such as ink width, slant) and local (such as character style, ligatures) writing style patterns. The overall architecture has four components:

- Conditional Generator : Synthesize handwritten text.
- Discriminator: Ensures realistic generation of handwriting styles. It is designed to be convolutional in nature
- Recognizer: Aids in textual content preservation. It is inspired by CRNN

- Style Classifier: Ensures satisfactory transfer of calligraphic styles.

In the paper, the focus of the design is in the generator model. To imitate a handwritting style as realistically as possible, This model is designed to learn style content entaglement as well as local and global style patterns. It is a transformer-based generative network for unconstrained styled handwritten text image generation. It has two main components an encoder network and a decoder network. Both the encoder and decoder networks constitute a hybrid design based on convolution and multi-head self-attention networks.

HWT generates realistic styled handwritten text images and significantly outperforms other state-of-the-art models through extensive qualitative, quantitative and human based evaluations. The model also generalizes well to the challenging scenarios where both words and writing style are unseen during training, generating realistic styled handwritten text images

3.3. Decoupled Style Descriptors

3.4. Dynamically configurable CRNN

3.5. Handwriting Trajectory Recovery

3.6. Domain-Adversarial Neural Network

[**?**] DANN (Domain-Adversarial Neural Network) is an interesting architecture that we believe, could be utilized in the problem of specific-style handwriting generation. DANN is an improvement to any regular network structure due to its parallel branch.

Given a Neural Network performing a downstream task on a given dataset, the output of the penultimate layer is considered the final representation of the input data. On this representation, we apply our last layer which is the downstream task itself but the more interesting part is the learnt representation. We claim that with enough training, this penultimate layer output is the best possible latent space representation of the data distribution possessing the information required to perform a downstream task. The nature of Deep Learning makes it so that we can never truly know the kind of correlations the model has learnt from the input data and how the information is being represented but we know that the information required for the downstream task is somehow contained in the representation. If the existence of a downstream task and its loss function controls the information present in the representation, then if we want to include or exclude further information from the representation, all we need to do is have a parallel branch of the neural network performing that downstream task on the representation. When we consider the loss of this new branch downstream task while updating our representation-learning, the learnt representation now possesses or is independent of information regarding this task.

For example, if we had images of X-Rays of a human body part and we had to determine if the subject is fractured or not, our primary downstream task would be to classify the image as fractured or not. But if we have labels as to what part of the human body is present in the image, then we can ensure that our representation learns features of the image that are independent of whether the subject is a hand or a leg. This can be implemented by having a parallel downstream task on the representation that classifies the image as being an X-Ray of a hand or an X-Ray of a leg. Next, we subtract the loss of this classifier for the representation-learning layers. Thus, the model is forced to get worse at identifying whether the subject in the image is a hand or a leg making its learnt representation independent (but aware) of this information and only focussed on whether the subject is fractured or not.

Thus, this architecture allows us to hand-pick the kind of information we want to be depicted in the model's latent space representation of the data by making the model either improve or become worse at a parallel task involving correlated information. This architecture was originally proposed for use in the biomedical space but we believe that style transfer problems too could benefit from such an architecture due to the requirement of considering handwriting style information and character information distinctly. If we are able to isolate the style information from the character information completely giving the model a better empirical understanding of the handwriting style itself, this may improve our ability to generate better handwriting samples.

4. Introduction

Please follow the steps outlined below when submitting your manuscript to the IEEE Computer Society Press. This style guide now has several important modifications (for example, you are no longer warned against the use of sticky tape to attach your artwork to the paper), so all authors should read this new version.

4.1. Language

All manuscripts must be in English.

4.2. Dual submission

Please refer to the author guidelines on the CVPR 2023 web page for a discussion of the policy on dual submissions.

4.3. Paper length

Papers, excluding the references section, must be no longer than eight pages in length. The references section will not be included in the page count, and there is no limit on the length of the references section. For example, a paper of eight pages with two pages of references would have

a total length of 10 pages. **There will be no extra page charges for CVPR 2023.**

Overlength papers will simply not be reviewed. This includes papers where the margins and formatting are deemed to have been significantly altered from those laid down by this style guide. Note that this LaTeX guide already sets figure captions and references in a smaller font. The reason such papers will not be reviewed is that there is no provision for supervised revisions of manuscripts. The reviewing process cannot determine the suitability of the paper for presentation in eight pages if it is reviewed in eleven.

4.4. The ruler

The LaTeX style defines a printed ruler which should be present in the version submitted for review. The ruler is provided in order that reviewers may comment on particular lines in the paper without circumlocution. If you are preparing a document using a non-LaTeX document preparation system, please arrange for an equivalent ruler to appear on the final output pages. The presence or absence of the ruler should not change the appearance of any other content on the page. The camera-ready copy should not contain a ruler. (LaTeX users may use options of cvpr.sty to switch between different versions.)

Reviewers: note that the ruler measurements do not align well with lines in the paper — this turns out to be very difficult to do well when the paper contains many figures and equations, and, when done, looks ugly. Just use fractional references (*e.g.*, this line is 087.5), although in most cases one would expect that the approximate location will be adequate.

4.5. Paper ID

Make sure that the Paper ID from the submission system is visible in the version submitted for review (replacing the "*****" you see in this document). If you are using the LaTeX template, **make sure to update paper ID in the appropriate place in the tex file**.

4.6. Mathematics

Please number all of your sections and displayed equations as in these examples:

$$E = m \cdot c^2 \tag{1}$$

and

$$v = a \cdot t. \tag{2}$$

It is important for readers to be able to refer to any particular equation. Just because you did not refer to it in the text does not mean some future reader might not need to refer to it. It is cumbersome to have to use circumlocutions like "the equation second from the top of page 3 column 1". (Note that the ruler will not be present in the final copy, so is not an alternative to equation numbers). All authors will benefit from reading Mermin's description of how to write mathematics: http://www.pamitc.org/documents/mermin.pdf.

4.7. Blind review

Many authors misunderstand the concept of anonymizing for blind review. Blind review does not mean that one must remove citations to one's own work—in fact it is often impossible to review a paper unless the previous citations are known and available.

Blind review means that you do not use the words "my" or "our" when citing previous work. That is all. (But see below for tech reports.)

Saying "this builds on the work of Lucy Smith [1]" does not say that you are Lucy Smith; it says that you are building on her work. If you are Smith and Jones, do not say "as we show in [7]", say "as Smith and Jones show in [7]" and at the end of the paper, include reference 7 as you would any other cited work.

An example of a bad paper just asking to be rejected:

> An analysis of the frobnicatable foo filter.
>
> In this paper we present a performance analysis of our previous paper [1], and show it to be inferior to all previously known methods. Why the previous paper was accepted without this analysis is beyond me.
>
> [1] Removed for blind review

An example of an acceptable paper:

> An analysis of the frobnicatable foo filter.
>
> In this paper we present a performance analysis of the paper of Smith *et al.* [1], and show it to be inferior to all previously known methods. Why the previous paper was accepted without this analysis is beyond me.
>
> [1] Smith, L and Jones, C. "The frobnicatable foo filter, a fundamental contribution to human knowledge". Nature 381(12), 1-213.

If you are making a submission to another conference at the same time, which covers similar or overlapping material, you may need to refer to that submission in order to explain the differences, just as you would if you had previously published related work. In such cases, include the anonymized parallel submission [6] as supplemental material and cite it as

> [1] Authors. "The frobnicatable foo filter", F&G 2014 Submission ID 324, Supplied as supplemental material `fg324.pdf`.

Finally, you may feel you need to tell the reader that more details can be found elsewhere, and refer them to a technical report. For conference submissions, the paper must stand on its own, and not *require* the reviewer to go to a tech report for further details. Thus, you may say in the body of the paper "further details may be found in [7]". Then submit the tech report as supplemental material. Again, you may not assume the reviewers will read this material.

Sometimes your paper is about a problem which you tested using a tool that is widely known to be restricted to a single institution. For example, let's say it's 1969, you have solved a key problem on the Apollo lander, and you believe that the CVPR70 audience would like to hear about your solution. The work is a development of your celebrated 1968 paper entitled "Zero-g frobnication: How being the only people in the world with access to the Apollo lander source code makes us a wow at parties", by Zeus *et al*.

You can handle this paper like any other. Do not write "We show how to improve our previous work [Anonymous, 1968]. This time we tested the algorithm on a lunar lander [name of lander removed for blind review]". That would be silly, and would immediately identify the authors. Instead write the following:

> We describe a system for zero-g frobnication. This system is new because it handles the following cases: A, B. Previous systems [Zeus et al. 1968] did not handle case B properly. Ours handles it by including a foo term in the bar integral.
>
> ...
>
> The proposed system was integrated with the Apollo lunar lander, and went all the way to the moon, don't you know. It displayed the following behaviours, which show how well we solved cases A and B: ...

As you can see, the above text follows standard scientific convention, reads better than the first version, and does not explicitly name you as the authors. A reviewer might think it likely that the new paper was written by Zeus *et al*., but cannot make any decision based on that guess. He or she would have to be sure that no other authors could have been contracted to solve problem B.

FAQ

Q: Are acknowledgements OK?
A: No. Leave them for the final copy.

Q: How do I cite my results reported in open challenges?
A: To conform with the double-blind review policy, you can report results of other challenge participants together with your results in your paper. For your results, however, you should not identify yourself and should not mention your participation in the challenge. Instead present your results referring to the method proposed in your paper and draw conclusions based on the experimental comparison to other results.

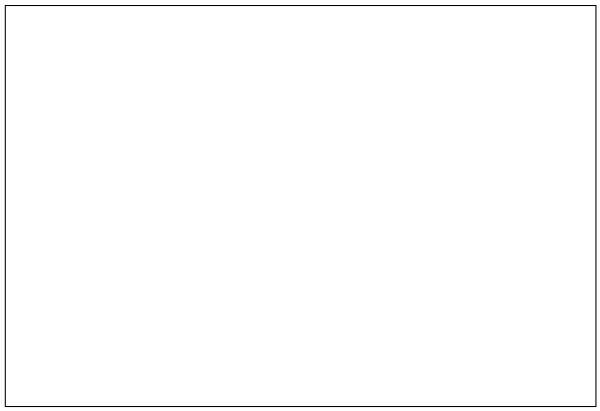


Figure 1. Example of caption. It is set in Roman so that mathematics (always set in Roman: $B \sin A = A \sin B$) may be included without an ugly clash.

4.8. Miscellaneous

Compare the following:
`$conf_a$` $conf_a$
`$\mathit{conf}_a$` conf_a
See The TEXbook, p165.

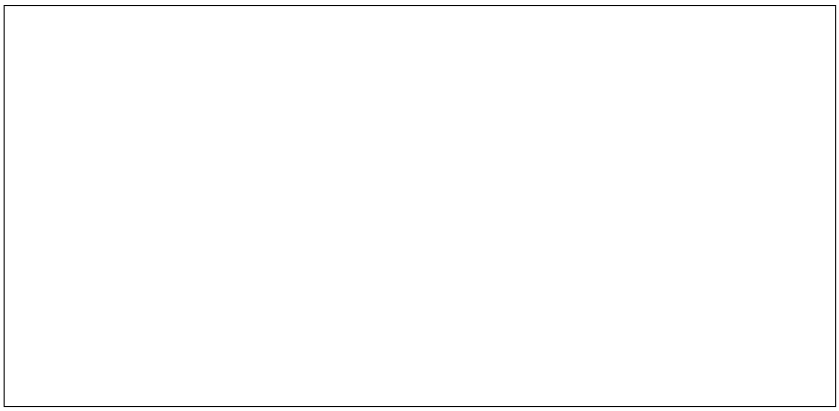
The space after *e.g.*, meaning "for example", should not be a sentence-ending space. So *e.g.* is correct, *e.g.* is not. The provided `\eg` macro takes care of this.

When citing a multi-author paper, you may save space by using "et alia", shortened to "*et al.*" (not "*et. al.*" as "*et*" is a complete word). If you use the `\etal` macro provided, then you need not worry about double periods when used at the end of a sentence as in Alpher *et al*. However, use it only when there are three or more authors. Thus, the following is correct: "Frobnication has been trendy lately. It was introduced by Alpher [1], and subsequently developed by Alpher and Fotheringham-Smythe [2], and Alpher *et al*. [3]."

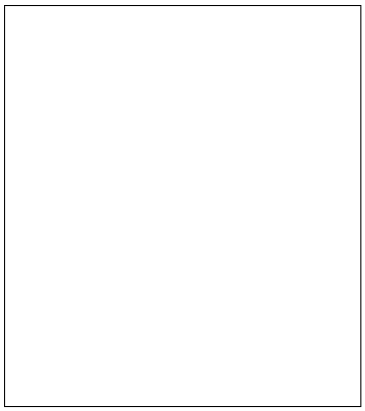
This is incorrect: "... subsequently developed by Alpher *et al*. [2] ..." because reference [2] has just two authors.

5. Formatting your paper

All text must be in a two-column format. The total allowable size of the text area is $6\frac{7}{8}$ inches (17.46 cm) wide by $8\frac{7}{8}$ inches (22.54 cm) high. Columns are to be $3\frac{1}{4}$ inches (8.25 cm) wide, with a $\frac{5}{16}$ inch (0.8 cm) space between them. The main title (on the first page) should begin 1 inch (2.54 cm) from the top edge of the page. The second and following pages should begin 1 inch (2.54 cm) from the top edge. On all pages, the bottom margin should be $1\frac{1}{8}$ inches (2.86 cm)



(a) An example of a subfigure.



(b) Another example of a subfigure.

Figure 2. Example of a short caption, which should be centered.

from the bottom edge of the page for 8.5×11-inch paper; for A4 paper, approximately $1\frac{5}{8}$ inches (4.13 cm) from the bottom edge of the page.

5.1. Margins and page numbering

All printed material, including text, illustrations, and charts, must be kept within a print area $6\frac{7}{8}$ inches (17.46 cm) wide by $8\frac{7}{8}$ inches (22.54 cm) high. Page numbers should be in the footer, centered and $\frac{3}{4}$ inches from the bottom of the page. The review version should have page numbers, yet the final version submitted as camera ready should not show any page numbers. The LaTeX template takes care of this when used properly.

5.2. Type style and fonts

Wherever Times is specified, Times Roman may also be used. If neither is available on your word processor, please use the font closest in appearance to Times to which you have access.

MAIN TITLE. Center the title $1\frac{3}{8}$ inches (3.49 cm) from the top edge of the first page. The title should be in Times 14-point, boldface type. Capitalize the first letter of nouns, pronouns, verbs, adjectives, and adverbs; do not capitalize articles, coordinate conjunctions, or prepositions (unless the title begins with such a word). Leave two blank lines after the title.

AUTHOR NAME(s) and AFFILIATION(s) are to be centered beneath the title and printed in Times 12-point, non-boldface type. This information is to be followed by two blank lines.

The ABSTRACT and MAIN TEXT are to be in a two-column format.

MAIN TEXT. Type main text in 10-point Times, single-spaced. Do NOT use double-spacing. All paragraphs should be indented 1 pica (approx. $\frac{1}{6}$ inch or 0.422 cm). Make sure your text is fully justified—that is, flush left and flush right. Please do not place any additional blank lines between paragraphs.

Figure and table captions should be 9-point Roman type as in Figs. 1 and 2. Short captions should be centred.

Callouts should be 9-point Helvetica, non-boldface type. Initially capitalize only the first word of section titles and first-, second-, and third-order headings.

FIRST-ORDER HEADINGS. (For example, **1. Introduction**) should be Times 12-point boldface, initially capitalized, flush left, with one blank line before, and one blank line after.

SECOND-ORDER HEADINGS. (For example, **1.1. Database elements**) should be Times 11-point boldface, initially capitalized, flush left, with one blank line before, and one after. If you require a third-order heading (we discourage it), use 10-point Times, boldface, initially capitalized, flush left, preceded by one blank line, followed by a period and your text on the same line.

5.3. Footnotes

Please use footnotes[1] sparingly. Indeed, try to avoid footnotes altogether and include necessary peripheral observations in the text (within parentheses, if you prefer, as in this sentence). If you wish to use a footnote, place it at the bottom of the column on the page on which it is referenced. Use Times 8-point type, single-spaced.

5.4. Cross-references

For the benefit of author(s) and readers, please use the

```
\cref{...}
```

command for cross-referencing to figures, tables, equations, or sections. This will automatically insert the appropriate label alongside the cross-reference as in this example:

[1] This is what a footnote looks like. It often distracts the reader from the main flow of the argument.

Method	Frobnability
Theirs	Frumpy
Yours	Frobbly
Ours	Makes one's heart Frob

Table 1. Results. Ours is better.

To see how our method outperforms previous work, please see Fig. 1 and Tab. 1. It is also possible to refer to multiple targets as once, *e.g.* to Figs. 1 and 2a. You may also return to Sec. 5 or look at Eq. (2).

If you do not wish to abbreviate the label, for example at the beginning of the sentence, you can use the

```
\Cref{...}
```

command. Here is an example:

Figure 1 is also quite important.

5.5. References

List and number all bibliographical references in 9-point Times, single-spaced, at the end of your paper. When referenced in the text, enclose the citation number in square brackets, for example [6]. Where appropriate, include page numbers and the name(s) of editors of referenced books. When you cite multiple papers at once, please make sure that you cite them in numerical order like this [1, 2, 4, 6, 7]. If you use the template as advised, this will be taken care of automatically.

5.6. Illustrations, graphs, and photographs

All graphics should be centered. In LaTeX, avoid using the `center` environment for this purpose, as this adds potentially unwanted whitespace. Instead use

```
\centering
```

at the beginning of your figure. Please ensure that any point you wish to make is resolvable in a printed copy of the paper. Resize fonts in figures to match the font in the body text, and choose line widths that render effectively in print. Readers (and reviewers), even of an electronic copy, may choose to print your paper in order to read it. You cannot insist that they do otherwise, and therefore must not assume that they can zoom in to see tiny details on a graphic.

When placing figures in LaTeX, it's almost always best to use `\includegraphics`, and to specify the figure width as a multiple of the line width as in the example below

```
\usepackage{graphicx} ...
\includegraphics[width=0.8\linewidth]
                {myfile.pdf}
```

5.7. Color

Please refer to the author guidelines on the CVPR 2023 web page for a discussion of the use of color in your document.

If you use color in your plots, please keep in mind that a significant subset of reviewers and readers may have a color vision deficiency; red-green blindness is the most frequent kind. Hence avoid relying only on color as the discriminative feature in plots (such as red *vs.* green lines), but add a second discriminative feature to ease disambiguation.

6. Final copy

You must include your signed IEEE copyright release form when you submit your finished paper. We MUST have this form before your paper can be published in the proceedings.

Please direct any questions to the production editor in charge of these proceedings at the IEEE Computer Society Press: https://www.computer.org/about/contact.

References

[1] FirstName Alpher. Frobnication. *IEEE TPAMI*, 12(1):234–778, 2002. 4, 6

[2] FirstName Alpher and FirstName Fotheringham-Smythe. Frobnication revisited. *Journal of Foo*, 13(1):234–778, 2003. 4, 6

[3] FirstName Alpher, FirstName Fotheringham-Smythe, and FirstName Gamow. Can a machine frobnicate? *Journal of Foo*, 14(1):234–778, 2004. 4

[4] FirstName Alpher and FirstName Gamow. Can a computer frobnicate? In *CVPR*, pages 234–778, 2005. 6

[5] Ankan Kumar Bhunia, Salman H. Khan, Hisham Cholakkal, Rao Muhammad Anwer, Fahad Shahbaz Khan, and Mubarak Shah. Handwriting transformers. *CoRR*, abs/2104.03964, 2021. 1

[6] FirstName LastName. The frobnicatable foo filter, 2014. Face and Gesture submission ID 324. Supplied as supplemental material `fg324.pdf`. 3, 6

[7] FirstName LastName. Frobnication tutorial, 2014. Supplied as supplemental material `tr.pdf`. 4, 6