Proposal for Research

Title:

Error-corrected deep learning approach to handwritten text recognition of Gregg shorthand

Investigator:

Alexander Weimer

Minnetonka Research
Minnetonka High School
18301 Highway 7
Minnetonka, MN 55345

E-Mail Address:

019476@mtka.org

September 12, 2024

1 Project Summary

Shorthand, also known as pen stenography, is a family of writing systems for English and other languages that emerged out of a need for a fast and efficient writing system in a pre-digital age. Of the many English shorthand systems, Gregg shorthand is the most prevalent [Zhai et al., 2018]. While largely made obsolete by typewriters and later general-purpose computers, the value within old shorthand documents mean that being able to efficiently scan shorthand documents into modern computer systems holds significant value.

The proposed research will leverage a deep-learning-based approach to handwritten text recognition (HTR) with several methods of error correction to enable the transliteration of full lines of Grgg shorthand into digital typed English. A physical hardware device will be developed to explore the possibility of low-cost on-site scanning of shorthand documents.

2 Background information

In the modern day, people can depend on computer systems for relatively fast text input, but this was not always the case. Shorthand, also known as pen stenography, is a family of writing systems for English and other languages that emerged out of a need for a fast and efficient writing system in a pre-digital age.

Of the many English shorthand systems, Gregg shorthand—first developed in 1888—is the most prevalent [Zhai et al., 2018]. With the advent of digital text input and storage, shorthand has largely fallen out of use in favor of standard typing and digital stenography [Rajasekaran and Ramar, 2012].

Optical Character Recognition (OCR) is the transliteration of (often handwritten) text samples from photographs or scans of physical material into digitally encoded text.

Handwritten text recognition (HTR) is the transliteration of full handwritten words or sentences, from both digital (i.e. touchscreen or stylus) and physical (i.e. pen and paper, captured through photographs or scans) input methods into digitally encoded text.

Transliteration of shorthand scripts into longhand ("regular) script presents a unique challenge for several reasons. Shorthand characters often lack distinct features, sometimes varying only in length or degree of curvature. Furthermore, shorthand lexicons are often simplified, often missing vowels or other defining features of words. While the human mind can accommodate well for these kinds of omissions, creating a digital system that can do the same poses a challenge.

Reading of manuscripts written in shorthand is vital to understanding documents in a wide variety of fields where fast handwriting was necessary in the past, such as law and medicine. The digitization and thereby preservation of shorthand documents, therefore, presents possible benefits in preservation of history and culture.

Further, the development of extensible HTR and OCR systems, in this case with a focus on English shorthand, feasibly opens avenues for the creation of HTR and OCR systems for other written languages—thus presenting possible benefits for the preservation of world languages and cultures.

Digital applications of Gregg shorthand (2) have been investigated as far back as 1990 by Agarwal1 [Agarwal, 1990]. In 2004, HTR of Pitman shorthand—a script related to gregg—was investigated as a means of rapid text entry into mobile devices by Higgins, who also presents important data on the feasibility of transliteration based on the simplified lexicon of many shorthand systems (2).

More recently, publications by Rajasekaran and Ramar in 2012 as well as Zhai et al. in 2018 have focused on machine learning approaches to HTR of Gregg shorthand [Rajasekaran and Ramar, 2012] [Zhai et al., 2018]. Zhai et al. release a dataset Gregg-1916, consisting of 16,000 common English words written in Gregg shorthand as part of their research [Zhai et al., 2018].

In June 2024, Heil and Nauwerck present a deep-learning approach to HTR of the Swedish-language Melin shorthand with fairly impressive results. As part of their publication, they include the LION dataset of shorthand manuscripts written in the Melin system [Heil and Nauwerck, 2024].

Recent shorthand recognition models still struggle with accurately identifying entire words correctly, with error rates near or above the 50% mark [Heil and Nauwerck, 2024]. This presents an opportunity to develop shorthand text recognition models further.

Connecting Heil and Nauwerck's modern deep learning approach to shorthand HTR in the Swedish Merlin system with past research into HTR of the English-language Gregg system, as well as other modern deep learning techniques would enable broader, more effective HTR within an English-language context. This would be beneficial for preservation, understanding and accessibility of historical texts and records written in shorthand.

3 Question to be Studied

The focus of this study is to develop a novel deep learning neural approach to text recognition of Gregg shorthand script. Based on the results of a similar investigation into Swedish Melin shorthand text recognition by Heil and Nauwerck, an accuracy of near 50% is expected.

4 Research Plan

Objective 1: Develop pytorch-based deep learning model for letter-by-letter Gregg shorthand
recognition.

Rationale: A deep learning model forms the core of the proposed research, and is the first link in a chain of software to convert physical, written shorthand to digitally encoded standard typed English. This step is prerequisite to all other objectives in the proposed research plan.

Objective 2: Develop & integrate word-by-word dictionary-based error correction algorithm

Rationale: This layer will contribute to making transliterated text from the letter-by-letter recognition model significantly more legible by cross-referencing with a list of words. This will improve accuracy and success rates of the proposed research overall.

Objective 3: Integrate error correction with paragraph-level context through the Llama 3.2 large language model

Rationale: Using an existing large language model to increase the accuracy of the proposed research by taking advantage of the large context windows of such LLMs would likely drastically improve the success rate of the proposed research. While not strictly necessary, this layer would ultimately provide benefits to the efficacy of the proposed research.

Objective 4: Develop software tests, benchmarks and visualizers for all tool layers

Rationale: Having software tests is considered industry good practice and implementing them would be beneficial for quality control of the proposed research. Similarly, software benchmarks will provide important data that identifies the efficacy of the models produced in the proposed research in a way that can be understood by an audience without a background related to the proposed research. Similarly, visualizers will produce images and other media to illustrate details of the proposed research to a wider audience. Both benchmarks and visualizers will also allow for direct comparison between my model and models published by others.

Objective 5: Create hardware device to deploy model in consistent environment; benchmark model in field environment.
Rationale: Creating a dedicated hardware device to integrate with the proposed research would mean having a physical, demonstrable product that would increase the physicality and legitimacy of the proposed research. It would also allow for a consistent environment for image scanning & recognition that could feasibly improve efficacy of the models produced in earlier steps in the proposed research.

5 Budget

No.	Vendor Name	Item # (SKU, etc.)	Item	Item Price	Qty.
1a	Sparkfun https://www.sparkfun.com/products/26355?src=raspberrypi	DEV-26355	RPi AI Camera	$70	1
1b	Adafruit https://www.adafruit.com/product/6009?src=raspberrypi	6009		$70	1
2a	Sparkfun https://www.sparkfun.com/products/23551?src=raspberrypi	DEV-23551	RPi 5 8g	$80	1
2b	Micro Center https://www.microcenter.com/product/673711/raspberry-pi-5-8gb?src=raspberrypi	635649		$70	1
3a	Adafruit https://www.adafruit.com/product/6011	6011	64gb MicroSD Card	$12	1
3b	Sparkfun https://www.sparkfun.com/products/26389	COM-2638		$15	1
4	N/A	N/A	Soldering Station	$0	1
			High Total	**$165**	
			Low Total	**$152**	

Figure 1: Proposed Project Budget - Also Available Online

6 Experimental Design

The Gregg-shorthand HTR machine learning model of Research Objective 1 was developed using the `pytorch` deep learning framework, itself based on the `torch` ML library.

The `pytorch` ecosystem was chosen for its advantages in flexibility and extensibility compared to popular alternative models within the deep learning field, most significantly `TensorFlow`.

To allow processing of image data, `torchvision`, a package that extends `pytorch` with common image transformations for data augmentation and preprocessing was implemented in the realization of a Gregg shorthand recognition model. `PIL`, the Python Imaging Library, is one of many more minor industry-standard libraries that was aksi used in the training of a model in line with Research Objective 1.

To allow for the effective processing of sequentially ordered Gregg shorthand characters, the aforementioned `pytorch` model was based on the Gated Recurrent Unit (GRU) neural network architecture. It was trained on the Gregg-1916 dataset introduced by Zhai et. al. in 2018, comprising greater than 16,000 Gregg shorthand words.

Dropout Regularization was implemented as protection against overfitting to the training dataset.

An interpreter layer to correct errors in the text output produced by the Gregg text recognition model based on the Levenshtein Distance metric was extended from the existing MIT-licensed `SymSpell` algorithm and implemented sequentially after the Gregg recognition model to correct errors with a dictionary-level context.

Both the pre-correction and post-correction recognized strings are, in pursuit of Research Objective 3, supplied to the lightweight open-source 3B-parameter & text-only version of the Large Language Model `Llama 3.2`, ran locally. This step was implemented to correct for possible errors in text recognition with full context of concepts too advanced for the previous spell-check error correction approach, such as logic.

Software tests were written directly into the codebase for the proposed research. The pytorch `benchmark` library was used to benchmark model performance on the development machine. Simple code execution time benchmarks were used in all other software components of the proposed research.

The `matplotlib` package was used to create graphical representations of benchmark data, as well as accuracy data. The `TensorBoard` package was used in conjunction with `pytorch` to create image visualizations of the core Gregg shorthand recognition models.

A hardware device in line with Research Objective 5 was developed using a Raspberry Pi 5 8gb single-board computer. A camera leveraging the on-module AI processor of the Sony IMX500 Intelligent Vision Sensor Raspberry Pi AI camera was used to optimize the execution of visual neural tasks. The physical enclosure of the hardware device was modeled in Solidworks and printed in ABS plastic.

All programming work was conducted within the MIT-Licensed VSCodium application (cross-platform). All development work was tracked using the Git distributed version control system and released to GitHub under an MIT license.

7 Timeline

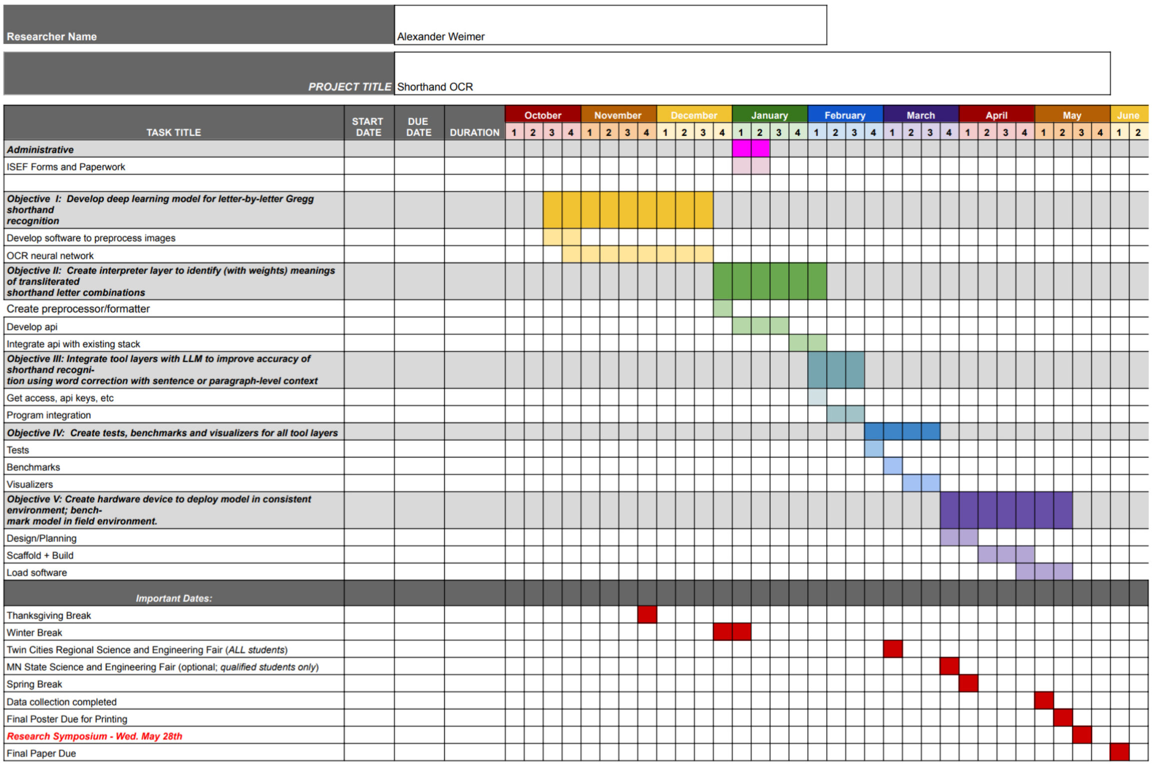

Figure 2: Proposed Project Timeline

8 Risk & Safety

There have been no significant risks identified to the researcher or those involved in this project.

References

[Agarwal, 1990] Agarwal, A. (1990). Off-line Shorthand Recognition System. [Online; accessed 9. Sep. 2024].

[Heil and Nauwerck, 2024] Heil, R. and Nauwerck, M. (2024). Handwritten stenography recognition and the LION dataset. *IJDAR*, pages 1–16.

[Rajasekaran and Ramar, 2012] Rajasekaran, R. and Ramar, K. (2012). Handwritten Gregg Shorthand Recognition. *International Journal of Computer Applications*.

[Zhai et al., 2018] Zhai, F., Fan, Y., Verma, T., Sinha, R., and Klakow, D. (2018). A Dataset and a Novel Neural Approach for Optical Gregg Shorthand Recognition. In *Text, Speech, and Dialogue*, pages 222–230. Springer, Cham, Switzerland.