Application of General Adversarial Networks in Smoothing Handwriting

John Santerre, YuMei Bennett, Edward Fry, Muchigi Kimari, Ikenna Nwaogu

Master of Science in Data Science, Southern Methodist University,  
Dallas, TX 75275 USA

{ybennett, edwardf, mkimari, inwaogu, jsanterre}@smu.edu

**Abstract.** – dysgraphia is a mental disorder that causes poor writing skills. Application of Generative Adversarial Networks can be used to improve this skill.

1 Introduction

**Problem Statement:** There are not many studies about dysgraphia, so the prevalence of this writing disability is not known exactly. However, it is estimated that between 5-20% population has a varied degree of dysgraphia. dysgraphia is largely unrecognized and misunderstood because we lose meaningful input from many people of high intelligence who simply communicate a little differently. dysgraphia is real and can have a significant negative impact.

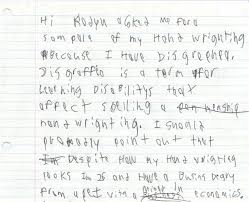
Because of this fine motor skill disorder, people with dysgraphia have poor writing skills. In addition to that, since so much focus is put on writing, people with dysgraphia also have a hard time putting their thoughts together on paper. However, recent developments in machine learning may present a solution. New algorithms, such as Generative Adversarial Networks, may be able to take unreadable handwriting and present it in a realistic handwritten text.

2 Literature Review

**Dysgraphia:** This is a motor disorder which affects skills such as writing and spelling. Since people with dysgraphia focus so much on writing, they tend to also find it difficult to express their ideas on paper. In addition to difficulty in expressing their ideas on paper, the academics for children with dysgraphia can be vastly affected [25]. dysgraphia is normally diagnosed early at the ages of five and six. The symptoms are:[25]

* Unreadable and slow handwriting
* Poor spelling
* Poor formation of letters
* And inconsistent spacing of letters effectively

Below is an example of handwriting from a child with dysgraphia.



**Project Goal:** This project seeks to produce a handwriting smoother that can be used by either of two audiences. The first audience would be normal (non-data scientist) people, who would simply scan a sample of their handwriting into an application and have smoothed results returned. The second audience would be the machine learning community, who would have access to the project code (open source with an MIT license) and to a full software package that can be imported and used in code.

**Technical Machine Learning (ML) Options**: To meet these goals, the smoother will need to be a trained machine learning model, so the next step is to evaluate possible candidates through the literature research and expert guidance. Five potential machine learning options emerge: Generative Adversarial Networks, Neural Style Transfer, Auto-encoders, Singular Value Decomposition (SVD), and Recurrent Neural Networks (RNNs). Of these, Generative Adversarial Networks and Style Transfer seem to show the most promise, so the project begins its focus on those two.

**Generative Adversarial Networks (GANS)**: This is a Neural Network used for unsupervised machine learning. Developed by Ian Goodfellow in 2014, GANs can be simplified into two parts, the generator and the discriminator. The main idea is that the generator creates fake data, and the discriminator validates to see if its real or fake. The goal for the generator is to fool the discriminator [3]. Since GANs are typically used in processing many images, it can be applied to improving bad handwriting.

**Neural Style Transfer**: This is an optimization technique used to copy the style from the style image and apply it to the content image [26]. “Style” is an indication of the patterns, the brushstrokes, and so on of handwriting. Neural style transfer takes three images, a content image, a style reference image, and an input image to be styled and then blend them together such that the input image is transformed to look like the content image but “painted” in the style of the style image.

The principle of neural style transfer is to define two distance functions, one that describes how different the content of two images are and one that describes the difference between the two images in terms of their style. Then, given three images, a desired style image, a desired content image, and the input image (initialized with the content image), the input image is transformed to minimize the content distance with the content image and its style distance with the style image.

Deﬁning and extracting handwriting styles alone is a challenging problem since there are no formal deﬁnitions for these styles. Smoothing dysgraphia handwriting and extracting a style is a new area of research.

**Data Sources**: In the interest of proper scale for a project of this size, the scope has a few limitations. First, only black and white training images are used. Second, only printed characters are supported. Finally, the character set is limited to upper and lower case English letters and numeric digits.

A total of 8 sources for the data were considered.

1. *UCI Letter Recognition Data Set* [17] – This repository is maintained by the University of California at Irvine and includes about 20000 capital letters in 20 different fonts.
2. *The Chars74K dataset* [18] – This collection of data includes about 74000 letters and numbers, but only 3410 of them are handwritten.
3. *EMNIST Dataset* [19] – These handwritten characters have been converted to the same 28x28 structure as the famous MNIST dataset. It includes 145,600 letters and 280,000 digits.
4. *MNIST DATABASE of handwritten digits* [20] – This famous dataset includes 70,000 handwritten numeric digits in 28x28 format.
5. *CVL-DATABASE* [21] – 310 writers wrote 7 texts out by hand in cursive script in this dataset.
6. *IAM Handwriting Top50* [22] – This collection of handwritten passages is broken out by sentences, lines of text, and individual words. It is mostly written in cursive.
7. *Yu Qiao* – This site is a collection of links to various handwriting samples and resources.
8. *Novel dataset creation* – In the event that no existing datasets contain the quality, type, or number of samples needed, a new dataset can be commissioned using a service such as Mechanical Turk.

The most promising datasets are *EMNIST*, *MNIST*, and *UCI* since they meet the requirements of individual printed numbers and letters in black and white. These are the data sources used in the construction of the training models.

**Machine Learning Platform**: Python will be used for this project. It is widely regarded as a suitable and standard programming language for machine learning and data science. Python is preferred because it supports object-oriented, and it has a wide variety of packages for cleaning data and machine learning. In addition, PyTorch has similar syntax to NumPy. The only difference is an option to use Graphic Processing Units (GPUs) [28]. The use of GPUs in PyTorch speeds the training of deep learning models.

**Distributed Computing**: Images by their very nature often have a lot of features since the individual pixels are typically used, so even a 28x28 image would have at least 784 features. Because of this, algorithm performance becomes a concern. To address this, we need to implement a distributed Python computing platform using Spark [27]. This platform automatically splits data and computation amongst a potentially unlimited number of computing nodes known as clusters. Thus, additional computers can be added to the cluster at any time to increase the speed of processing.

In addition, Spark supports the use of graphical processing units (GPUs), which are highly optimized for the sorts of linear algebra operations that most machine learning algorithms employ, thus adding both a greater quantity and a more performant set of computation resources to the execution of these machine learning algorithms.

**Hypothesis:** Having dysgraphia does not affect someone's mental capabilities; they just need extra help in handwriting skills. Consequently, the solution to challenges with having dysgraphia can be solved by Generative Adversarial Networks. GANs will be able to take bad handwriting from a document and present it in a better realistic handwritten text. Solving this problem will help people with dysgraphia focus more on expressing their ideas rather than how to write.

3 Methods

* Methods utilized for analysis
  + Interviews with Dr John Santerre and David Josephs
  + Decomposition of final desired product
* Initial methods we may use based on preliminary analysis
  + Generative Adversarial Network (GAN)
    - Two neural networks compete
    - Could merge samples
  + Style transfer
    - Famous example of applying artist style to photos
    - Would be used to merge samples
  + Auto encoder
    - Similar to principal components analysis (PCA)
    - More aggressive dimensional reduction
  + Singular value decomposition (SVD)
    - Linear algebra method
    - Sometimes helpful with image manipulation
  + Recurrent neural network (RNN)
    - Known applications in handwriting recognition
    - More difficult than GAN; will explore only as a backup option
  + Distributed computing
    - Used to minimize processing time
    - Explore Spark and Dask
    - Choose one to use
* Data sources that were used. Currently investigating:
  + MNIST – collection of handwritten digits in image format
  + UCI Machine Learning Repository - Letter Recognition Data Set
  + Chars74K dataset - Character Recognition in Natural Images
  + Data Augmentation – technique to expand existing datasets if needed

4 Results

* Demonstrate effectiveness of handwriting smoothing
* Identify performant models
* Produce a Python package other developers can use
* Release a command line utility any one can use to smooth handwriting
* Scope restricted to:
  + Printed handwriting
  + Numbers, upper, and lower case letters
  + Black and white samples
  + Texture of a standard ball-point ink pen
* Additional scope if time permits:
  + Punctuation
  + International characters from other languages
  + Color samples
  + Textures such as pencil, marker, paint, or crayon

5 Discussion

As explained in the introduction, dysgraphia is a learning disability which strictly affects the ability to write properly. This project seeks to provide a better way to help people with this condition to communicate when writing. In this project, an application will take an image input of a handwriting sample, and the application will replace the writing with smoother writing that is similar to the original input’s handwriting style.

Two different approaches are promising. The first is Generative Adversarial Networks (GANs), which is a generative deep learning neural network model that generates realistic new samples of the input. The second is the use of autoencoders, which are also generative deep learning models that would, instead of generating new samples, just reconstruct the input. GAN will be explored more because its unique functionality seems well-tailored to a smoothing problem. would make a better prediction model than the autoencoders. A GAN model has 2 main components: the generator and the discriminator. A discriminator is a classifier that tries to classify if a data is a real data or if the data was created by the generator. The generator learns to create a fake data and gets feedback from the discriminator.

Unfortunately, there no objective/loss functions used in training a GAN model. This means that there are no ways of checking the progress or efficiency of the model using a traditional loss function. Therefore, models will be evaluated by manually inspecting the generated image or using quantitative measures like inception score and boundary distortion. The evaluation process is replicated in other evaluated models in order to validate results evenly. Other options may be available to evaluate GANs, and this project explores the best options for both smoothing and for evaluation.

At the conclusion of this Capstone project, the application will be released as open source so that anyone that needs it would be able to have access to it. Finally, since the core goal of the project is to assist with a disability, there does not seem to be any negative impact to ethics.

6 Conclusion

Due to the impacts of dysgraphia on children and adults, this application is meant to help people who suffer from this condition to communicate better. The application would create handwriting that is more consistent and easier to read from an original image of a person’s handwriting. The application will not only help people with dysgraphia but also hopefully promote additional research into this area with the planned Python package to drive improvement in the years ahead.

Acknowledgments. Jacquelyn Cheun, PhD. – Capstone Professor

References

1. Renzhuo Wan, Shuping Mei, Jun Wang, Min Liu and Fan Yang. Multivariate Temporal Convolutional Network: A Deep Neural Networks Approach for Multivariate Time Series Forecasting. Electronics 2019, 8(8), 876; <https://doi.org/10.3390/electronics8080876>
2. Shengyu Zhao IIIS, Tsinghua University and MIT Zhijian Liu MIT Ji Lin MIT Jun-Yan Zhu Adobe Research Song Han MIT. Differentiable Augmentation for Data-Efficient GAN Training. <https://arxiv.org/pdf/2006.10738.pdf>
3. Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler and Sepp Hochreiter. GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium. <https://arxiv.org/pdf/1706.08500.pdf>
4. Phillip Isola Jun-Yan Zhu Tinghui Zhou Alexei A. Efros. Berkeley AI Research (BAIR) Laboratory, UC Berkeley. Image-to-Image Translation with Conditional Adversarial Networks. <https://arxiv.org/pdf/1611.07004.pdf>
5. Andrew Brock Heriot Watt University, Jeff Donahue DeepMind, Karen Simonyan DeepMind. LARGE SCALE GAN TRAINING FOR HIGH FIDELITY NATURAL IMAGE SYNTHESIS. <https://arxiv.org/pdf/1809.11096.pdf>
6. Yasin Yazıcı Nanyang Technological University (NTU), Chuan-Sheng Foo Institute for Infocomm Research, Stefan Winkler National University of Singapore (NUS), Kim-Hui Yap Nanyang Technological University (NTU), Georgios Piliouras Singapore University of Technology and Design, Vijay Chandrasekhar Institute for Infocomm Research. UNUSUAL EFFECTIVENESS OF AVERAGING IN GAN TRAINING. <https://arxiv.org/pdf/1806.04498.pdf>
7. Han Zhang, Zizhao Zhang, Augustus Odena, Honglak Lee Google Research. CONSISTENCY REGULARIZATION FOR GENERATIVE ADVERSARIAL NETWORKS. <https://arxiv.org/pdf/1910.12027.pdf>
8. Hao-Wen Dong · Yi-Hsuan Yang. May 2020. Towards a Deeper Understanding of Adversarial Losses under a Discriminative Adversarial Network Setting. <https://arxiv.org/pdf/1901.08753.pdf>
9. David Ha, Douglas Eck. A Neural Representation of Sketch Drawings. <https://arxiv.org/abs/1704.03477>
10. Adrian Rosebrock on August 24, 2020. Handwriting recognition with OpenCV, Keras, and TensorFlow. <https://www.pyimagesearch.com/2020/08/24/ocr-handwriting-recognition-with-opencv-keras-and-tensorflow/>
11. Omar Mohammed, Gerard Bailly, Damien Pellier, Univ. Grenoble-Alpes. Style Transfer and Extraction for the Handwritten Letters Using Deep Learning. <https://arxiv.org/pdf/1812.07103.pdf>
12. Bo Chang∗ Qiong Zhang∗ Shenyi Pan Lili Meng, University of British Columbia. Generating Handwritten Chinese Characters using CycleGAN. <https://arxiv.org/pdf/1801.08624.pdf>
13. Bharath Narasimhan University of Massachusetts Amherst. Calligraphy Style Transfer using Generative Adversarial Networks. <https://people.cs.umass.edu/~bnarasimhan/Cast.pdf>
14. Alex Graves Department of Computer Science University of Toronto. Generating Sequences With Recurrent Neural Networks. <https://arxiv.org/pdf/1308.0850>)
15. Piotr Teterwak Aaron Sarna Dilip Krishnan Aaron Maschinot
16. David Belanger, Ce Liu, William T. Freeman Google Research. Boundless: Generative Adversarial Networks for Image Extension. <https://arxiv.org/pdf/1908.07007.pdf>
17. Dua, D. and Graff, C. (2019). UCI Machine Learning Repository [<http://archive.ics.uci.edu/ml>]. Irvine, CA: University of California, School of Information and Computer Science.
18. De Campos, T.E., Babu, B.R. and Varma, M. (2009). The Chars74K dataset: Character Recognition in Natural Images [<http://www.ee.surrey.ac.uk/CVSSP/demos/chars74k/>]. Microsoft India
19. Cohen, G., Afshar, S., Tapson, J., & van Schaik, A. (2017). EMNIST: an extension of MNIST to handwritten letters. Retrieved from <http://arxiv.org/abs/1702.05373>
20. Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. "Gradient-based learning applied to document recognition." Proceedings of the IEEE, 86(11):2278-2324, November 1998.
21. Florian Kleber, Stefan Fiel, Markus Diem and Robert Sablatnig, CVL-Database: An Off-line Database for Writer Retrieval, Writer Identification and Word Spotting, In Proc. of the 12th Int. Conference on Document Analysis and Recognition (ICDAR) 2013, pp. 560-564, 2013.
22. U. Marti and H. Bunke. The IAM-database: An English Sentence Database for Off-line Handwriting Recognition. Int. Journal on Document Analysis and Recognition, Volume 5, pages 39 - 46, 2002.
23. Edgard Chammas and Chafic Mokel. Handwriting Recognition of Historical Documents with few labeled data. Nov 2018. March 2019. <https://arxiv.org/pdf/1811.07768.pdf>
24. Eloi Alonso, Ecole des Ponts and Bastien Moysset. Adversarial Generation of Handwritten Text Images Conditioned on Sequences. Nov 2018. <https://arxiv.org/pdf/1903.00277.pdf>
25. Biotteau, Maëlle et al. “Developmental Coordination Disorder and dysgraphia: Signs and Symptoms, Diagnosis, and Rehabilitation.” Neuropsychiatric disease and treatment 15 (2019): 1873–1885. Web.
26. Y.-L. Chen and C.-T. Hsu. Towards deep style transfer: A content-aware perspective. In Proceedings of the British Machine Vision Conference, 2016.
27. Matei Zaharia, Mosharaf Chowdhury, Michael J. Franklin, Scott Shenker, Ion Stoica. Spark: Cluster Computing with Working Sets. Amplab, University of California at Berkely, 2010.
28. Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Köpf, Edward Yang, Zach DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, Soumith Chintala. PyTorch: An Imperative Style, High-Performance Deep Learning Library. 33rd Conference on Neural Information Processing Systems, NeurIPS, 2019.

Appendix:

Leave here if needed for additional information.