Recognizing English Cursive Using Generative Adversarial Networks

Xinrui Yu and Jafar Saniie
Embedded Computing and Signal Processing Research Laboratory (http://ecasp.ece.iit.edu/)
Department of Electrical and Computer Engineering
Illinois Institute of Technology, Chicago, IL, U.S.A.

***Abstract*— English cursive is present in many historical documents, and there is a need to recognize such writing and convert them to electronic documents. However, automated English cursive recognition is still considered difficult, mainly because of the segmentation difficulty presented by connected characters in a word. To resolve this issue, we propose a method to recognize whole words using GAN (Generative Adversarial Networks). This is done by using trained GAN to generate a recognizable font from cursive words, and then using conventional OCR (Optical Character Recognition) method on this font. The performance of this method is evaluated using English cursive datasets.**

Keywords—optical character recognition, generative adversarial networks, cursive English script

I. Introduction

Among various styles of penmanship, cursive stands out as a common choice for many historical documents until the early 20th century; famous documents written in cursive include the United States Declaration of Independence, and the Gettysburg Address. Therefore, it would be vastly beneficial to use certain optical character recognition (OCR) techniques to recognize the aforementioned documents and store their contents in digital format for ease of viewing and searching.

However, while current OCR techniques achieves good results for general straight and italic scripts, cursive recognition still poses a challenge to different OCR techniques [1]. The main reason for the setback is considered to be the difference in script connection. Straight scripts mostly have simple written texts where characters are not connected/jointed, while most cursive script joint the characters in a word, making the handwriting continuous and greatly increases the difficulty of character segmentation. In addition, although recent OCR systems integrated deep learning techniques and vastly improved their performance, they were still insufficient to recognize cursive script with good accuracy.

In this paper, we developed a character recognition technique using generative adversarial networks (GAN), which is a deep learning method used in image stylization, recognition and classification [2]. In general, GAN consists of two competing neural networks, one network generates results similar to the samples in the database, while the other network discriminates the generated images from the samples within the same database. After some iterations in this competition, the GAN method performs better compared to the design of a single neural network without competition. Therefore, this competing neural networks, GAN, can be trained to recognize cursive script images with acceptable levels of accuracy.

II. Related Work

A. Implementations Using GAN

With its concept being developed in 2014 [2], GAN can be considered as relatively new in the fields of deep learning and neural networks. By contrast, it enjoyed rapid development and maturing in the last few years, with many astonishing implementations achieving unseen and unforeseen results. Such implementations are mostly related to image processing [3], and the most successful ones include image morphing, image stylization, and up-scaling of image textures. Among these implementations, the image stylization ones have the most similarity comparing with the implementation in this paper, as they are able to generate an image with a different style without changing the fundamental contents of the original image.

B. Accuracy of Current OCR Algorithms

An example is provided in Figure 1, showing that although current OCR algorithms are quite capable of recognizing italics, they are insufficient in recognizing English cursive [4]. This is true even for neural network based algorithms [5]. The first text image in the Figure is taken from a facsimile of the Declaration of Independence, and the recognition technique in this paper aims to perfectly recognize such script; the second text image is of the same sentence typed out with a similar font.

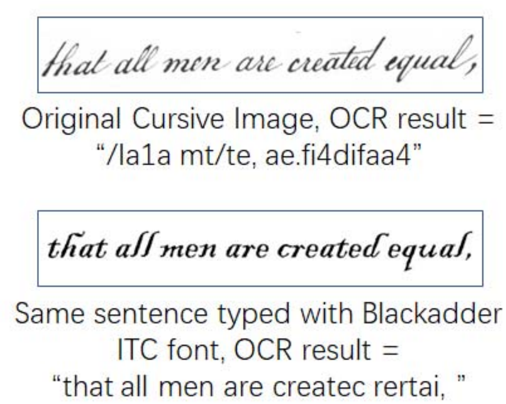


Figure 1. Example of Original Cursive Script and Generated Script

It can be seen that this OCR algorithm from onlineocr.net is almost completely useless on this cursive script. However, the

978-1-7281-5317-9/20/$31.00 ©2020 IEEE

same OCR algorithm have an 87% accuracy for the typed-out script. This shows that there is still much to improve for recognition of English cursive; also, this shows that if we can somehow convert original cursive script to a similar-looking computer generated script, we have a greatly improved accuracy for recognition.

III. Proposed Method for Recognizing Cursive

A. Principles of Operation for GAN

The model of the neural networks used in this paper can be described as a multilayer perceptron. While this is true for most image processing related neural network, the most important difference between GAN and other neural network is the adversarial nature of GAN. A diagram of an image processing GAN is shown in Figure 2. In this case, we choose to use Multi-Content GAN (MC-GAN) [6] as the basis of our study. MC-GAN is chosen as it performs the task of font style transfer with great results, and is able to generate scripts with a different font according to the original characters in the training images.

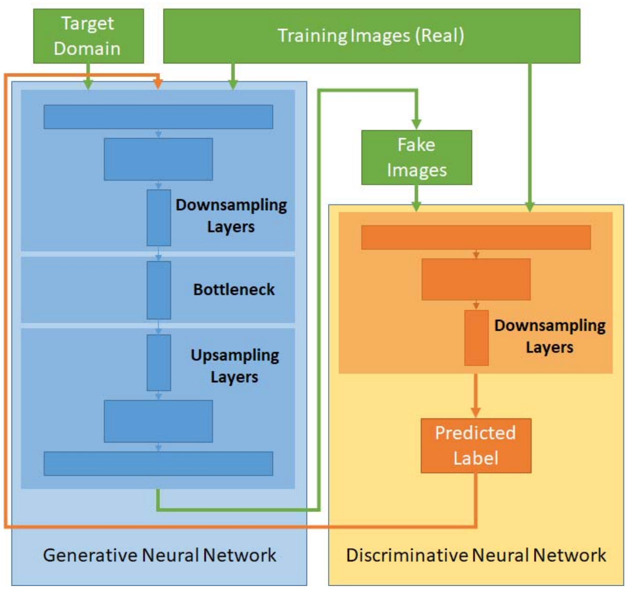


Figure 2. Structure Diagram of Image Processing GAN

An issue with the architecture of GAN in our study is that GAN tends to generate one-to-one results according to the original training image, which means each unique input will generate a unique output. While this is a good feature for general image style transfer implementations, it is different from the goal we have at hand, as we would like the GAN to generate same image for the same word with a slightly different font. Still, the results shown in the test [6] is promising, and it should be possible to generate recognizable output text image based on the input image. Also, while there is other research [7] aiming at performing OCR on cursive using GAN, their GAN generates ASCII codes directly as the output corresponding to the input cursive script image, which is different from our method below.

B. Cursive Script Recognition using GAN

The promise of recognizing English cursive is based on the decision of using an intermediate font between the original script and the final recognition result. The style transfer from the original script to the intermediate font is done by GAN; the recognition result is obtained by using conventional OCR algorithm on the intermediate font. This process is shown in Figure 3. The first line shows the original script; the second line shows the intermediate font generated by GAN; The last line in the figure represents the obtained OCR recognition result. In the actual process, only the original script and the intermediate script are processed as images; the final recognition result is encoded text, the machine font in the third line is only a representation of this encoded text.

The selection of the intermediate font is of great importance in this method. On the one hand, it has to be close related to the original script, so GAN will not have great difficulty performing style transfer from the original script; on the other hand, it need to be a font with good accuracy using conventional OCR algorithms, thus the result will be meaningful.

It is worthy of note that similar method is used in [1]; still, by using a different GAN architecture that showed success in font transformation, and by choosing a suitable intermediate font, it is possible to achieve better results.

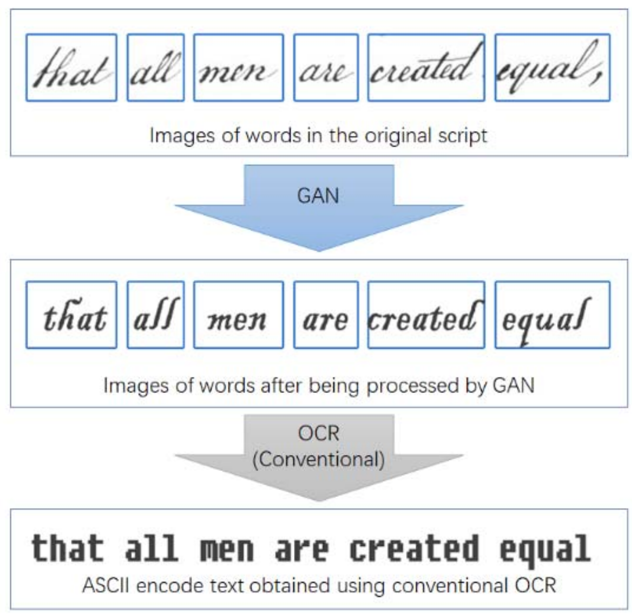


Figure 3. Flowchart for Cursive Script Recognition

A quantitative evaluation for the first step performing style transfer can be done by calculating reconstruction error compared with the ground truth of the intermediate script. The formula for calculating reconstruction error is:

$$E_{Rec} = E_{x,c}[\|x - G(x,c)\|] \quad (1)$$

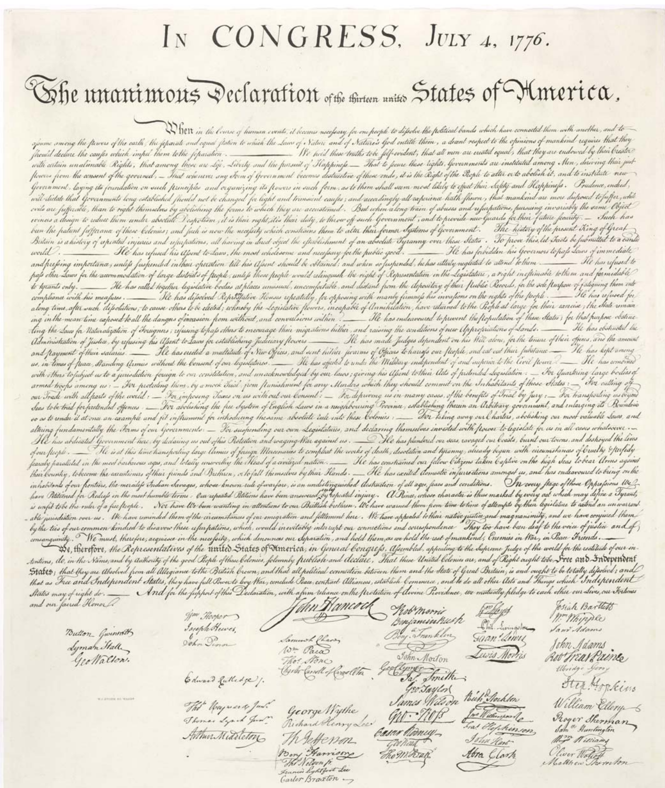
where x is the image of the exact computer-generated intermediate script corresponding to the original script, $G(x,c)$ is the intermediate script generated by GAN according to the original script.

It is interesting to compare this reconstruction error with the loss in OCR accuracy. As the GAN is unable to generate perfect computer scripts from the original script, a loss in OCR accuracy is expected. If it is possible to obtain the relationship between the reconstruction error and the loss in OCR accuracy, we can use the relationship to further optimize the selection of intermediate font.

IV. Implementation and Results

A. GAN Training and Intermediate Script Generation

For the purpose of generating intermediate script according to the original script, the GAN need to be trained with a dataset containing a large number of sample English cursive scripts and their corresponding computer-generated text, in image format. We were unable to find existing datasets matching such requirements; thus, a small dataset is created per the requirements. The sample English cursive Scripts is taken from a version of the Declaration of Independence, with each word taken as one sample. The original image of the entire Declaration of Independence is shown in Figure 4. The title, signatures of the congress members, and some texts with a different font is excluded from the dataset, as to not interfere the GAN in fitting to a specific font.


In CONGRESS, July 4, 1776.

The unanimous Declaration of the thirteen united States of America,

Figure 4. Original Image of the Declaration of Independence used in The Dataset

The reason to use one single word in each sample is because most documents written in cursive only connects characters within each word, instead of connecting characters within an entire sentence. Thus, while incapable of recognizing characters in a word, conventional OCR algorithms are still able to separate words in a document written in cursive; the conclusion is there is no need to use entire sentence for every sample in the dataset. This simplified the process of creating a large enough dataset, and also made the design of the GAN considerably simpler.

There are 1,458 words in the Declaration of Independence; 900 of these words are taken as 900 samples in the training set, and among the remaining words, 300 are used as test set. As the encoded text of the Declaration is readily available, the corresponding encoded text for every sample is used to generate computer-font text images. After training, the test set is given to the trained GAN, and the GAN generates intermediate scripts for the corresponding sample. An example of the result is shown in Figure 5.

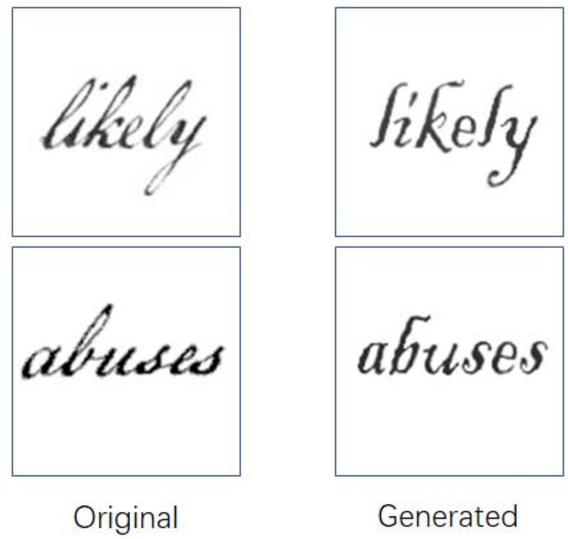


Figure 5. Original and Generated Script of GAN

B. Recognition

After generating intermediate scripts for each test sample, these scripts in image format are given to readily available conventional OCR algorithm to perform the final recognition step. Recognition is also performed on the original samples, so it can be used as a reference group.

Theoretically, we would expect the text images generated by GAN to have the highest OCR accuracy, and almost nil accuracy for the original text. A half sentence is chosen to show the image result in Figure 6. The text for the half sentence is: “it is the Right of the People to alter or to abolish it,”. There are 13 words in this half sentence, a total of 44 characters excluding space, and a total of 56 characters including space.

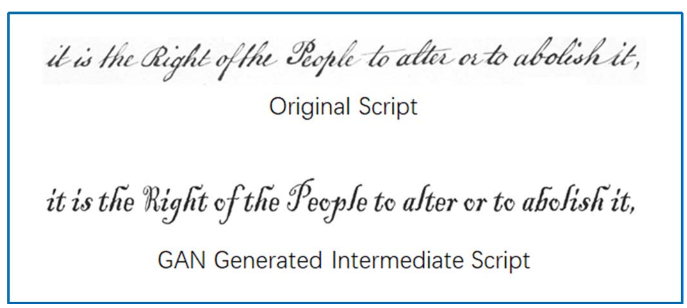


Figure 6. Original, Generated and Desired Script of GAN

Both group of images are given to a conventional OCR algorithm to perform character recognition and check the accuracy for OCR. The respective recognized texts and accuracies for the entire test set are shown in Table I. The accuracy is calculated by dividing the number of total characters including space by the number of correctly recognized characters including space. In the case of the recognized text and accuracy of the original text, it is not possible to determine the number of correctly recognized characters as the characters in each word is incorrectly separated, or not separated at all. Thus, the accuracy for OCR of original text is given as ~0%, or essentially nil.

TABLE I. OCR ACCURACY FOR TEST SET

Type	# of Words	# of Characters	Recognized Characters	Accuracy
Original	300	1786	N/A (unable to check)	~0%
GAN Generated	300	1786	1475	82.59%

As we can see from the table, by using the intermediate script for conventional OCR, we successfully recognized otherwise unrecognizable original text. The performance of this technique is also in accordance with our expectations, with the accuracy of the GAN generated text image much higher than that of the original text image.

To create an unbiased and simple test, some features common in everyday OCR software is not present in the OCR algorithm used in this paper, such as autocorrection based on meaningful words. As a good portion of recognition errors are one or two characters within a word, it can be said that the OCR accuracy can be further improved for our method should these features are introduced.

V. CONCLUSION

According to the test results mentioned above, we are able to recognize previously unrecognizable documents written in cursive and achieve an acceptable OCR accuracy. It is worthy of note that other GAN structures that have success in image morphing [8] and image style transfer [9] may also have good prospect in this field. Such success would be very beneficial to the field of handwritten script OCR, as the method of generating intermediate script using GAN is not restricted for English cursive. It is highly probable that other languages will also make use of this method. The recognition technique developed can be further expand to other field of operation that may be more difficult, such as recognizing texts that are generally written in hard-to-recognize scripts [10]. A cloud structure can be used to determine a good corresponding font, simplifying the task of network training and intermediate font selection [11]. Also, as the architecture of GAN suggests, it is not limited to generate intermediate scripts using original script; the opposite should also be possible, making it be able to generate cursive handwriting images with encoded text input.

REFERENCES

[1] E. Eman, S. S. Bukhari, M. Jenckel and A. Dengel, "Cursive Script Textline Image Transformation for Improving OCR Accuracy," *2019 International Conference on Document Analysis and Recognition Workshops (ICDARW)*, Sydney, Australia, 2019, pp. 59-64.

[2] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. "Generative Adversarial Nets," *Advances in Neural Information Processing Systems*, 2014, pp. 2672-2680.

[3] X. Zhang, X. Zhu, 3. X. Zhang, N. Zhang, P. Li and L. Wang, "SegGAN: Semantic Segmentation with Generative Adversarial Network," *2018 IEEE Fourth International Conference on Multimedia Big Data (BigMM)*, Xi'an, 2018, pp. 1-5.

[4] U. Choudhary, S. Bhosale, S. Bhise and P. Chilveri, "A survey: Cursive handwriting recognition techniques," *2017 2nd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT)*, Bangalore, 2017, pp. 1712-1716.

[5] P. V. Bhagyasree, A. James and C. Saravanan, "A Proposed Framework for Recognition of Handwritten Cursive English Characters using DAG-CNN," *2019 1st International Conference on Innovations in Information and Communication Technology (ICIICT)*, CHENNAI, India, 2019, pp. 1-4.

[6] S. Azadi, M. Fisher, V. Kim, Z. Wang, E. Shechtman and T. Darrell, "Multi-content GAN for Few-Shot Font Style Transfer," *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, 2018, pp. 7564-7573.

[7] G. Arna, B. Biswarup and C, Somnath, "Handwriting Profiling using Generative Adversarial Networks," *arXiv*:1611.08789v1, 2016

[8] R. Raghavendra, K. B. Raja, S. Venkatesh and C. Busch, "Transferable Deep-CNN Features for Detecting Digital and Print-Scanned Morphed Face Images," *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Honolulu, HI, 2017, pp. 1822-1830.

[9] T. Karras, S. Laine and T. Aila, "A style-based generator architecture for generative adversarial networks." *arXiv preprint arXiv*:1812.04948, 2018

[10] Kutzner, Tobias; Pazmiño-Zapatier, Carlos F.; Gebhard, Matthias; Bönninger, Ingrid; Plath, Wolf-Dietrich; Travieso, Carlos M. "Writer Identification Using Handwritten Cursive Texts and Single Character Words." *Electronics* 8, no. 4, 2019, pp. 391.

[11] P. S. Dhande and R. Kharat, "Character Recognition for Cursive English Handwriting to Recognize Medicine Name from Doctor's Prescription," *2017 International Conference on Computing, Communication, Control and Automation (ICCUBEA)*, Pune, 2017, pp. 1-5.