**Historical style generator - Literature review:**

**Students – Leor Ariel Rose, Yahav Bar David  
Academic advisor - Dr. Irina Rabaev**

Historical documents can reveal a lot of information, such as form of writing, wording, content that did not exist and more. Since those documents are usually written in diaries, pages or letters that have a final lifespan, there are certain periods that lack historical documents. The lack of data restricts us from using the power of deep learning and machine learning model, because most of these models require a large amount of data (big data).  
In this study we want to test whether it is possible to synthesize historical documents using two sources: a historical document (as a style) and an image of modern handwritten text (as a content).

Transferring the style from one image into another can be considered a problem of texture transfer. In texture transfer the goal is to synthesize a texture from a source image while constraining the texture synthesis to preserve the semantic content of a target image.

In the papers of Gatys et.al. [1] and [2], the researchers discovered a new technique for texture synthesis, an artificial system based on a deep neural network that creates artistic images of high perceptual quality. In those studies, the researchers used a CNN (Convolutional Neural Network) called VGG19 from the Caffe deep learning framework explained in [3] which is known as a deep network to classify images. The researchers used the output of the middle layers in the CNN network to extract the representation of the images. In each layer a representation of the image is created, and it is possible by rebuilding this representation to see the content obtained from this layer and take the output from the layer in which the representation of content and style is ideal.  
As you can see in Figure 1 the researchers found that in the first layers the output of the layers is almost identical to the original image, while the content is in the higher layers (deeper in the neural network).  
In addition, the researchers found that the style of an image can be described by the means and correlations across the different feature maps. Therefore, they calculate a Gram matrix that includes this information by taking the outer product of the feature vector with itself at each location and averaging that outer product over all locations.  
To connect the content representation and the style representation, the researchers create a new image that matches the two representations while reducing the distance and loss of information until reaching an acceptable threshold.  
In this study, the researchers were able to come to the understanding that there is a clear separation between content and style and that images can be combined to produce a synthesis of a particular style on a source image, as can be seen in Figure 1.  
In addition, the researchers came to the understanding that there is a trade-off between content and style, which means that coefficients can and should give more importance to the content of the image or the style of the image in order to achieve desired results.

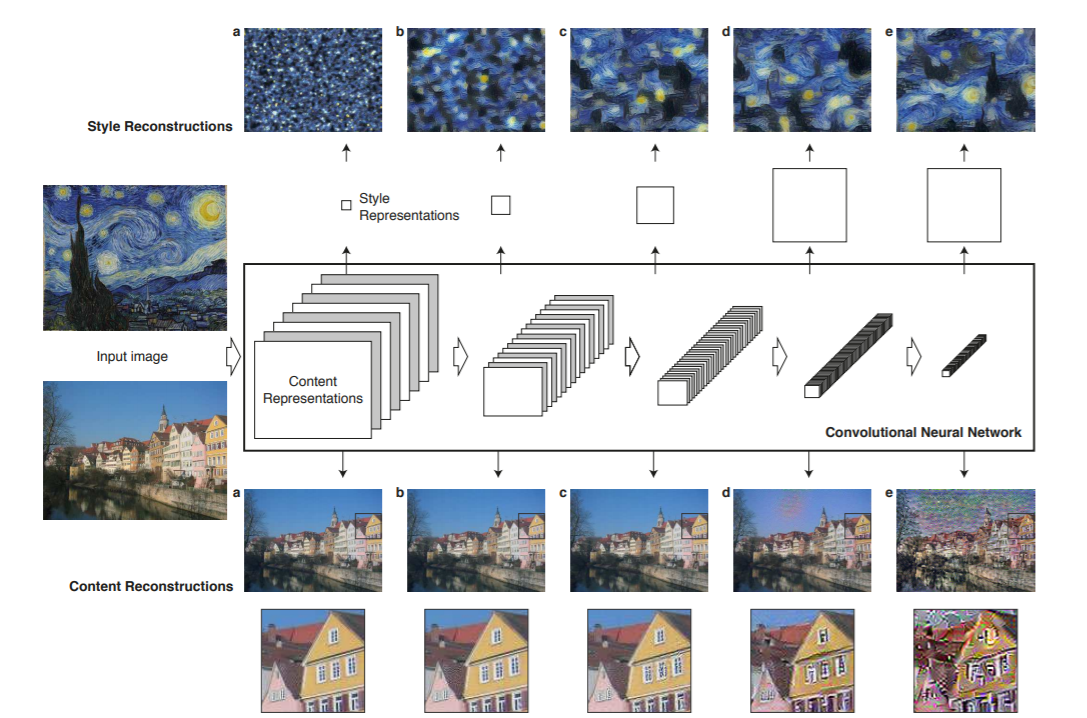


Figure 1 – Taken from “Image Style Transfer Using Convolutional Neural Networks”. The Figure shows the content and style in the VGG network.

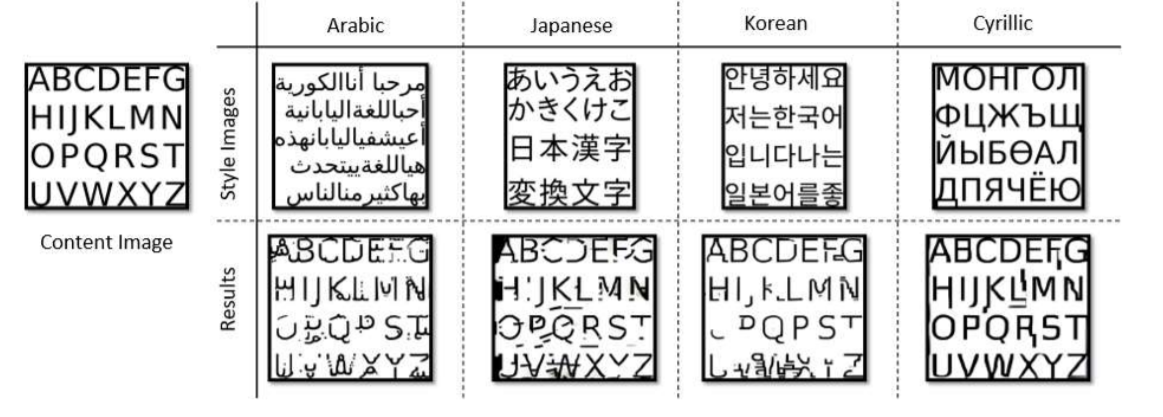
Basedon Gatys et.al papers [1] [2], a few papers were released about image style transfer, some of these papers use the original VGG19 architecture and some use additional techniques or models.

In the paper of Ter-Sarkisov [4] the researcher tries to transfer style from the logos of heavy metal bands onto corporate logos using a VGG16 network.  
The researcher used the VGG16 CNN architecture to extract style and content. Like the original paper [2], in this research he used the same technique to extract the style from heavy metal bands logos and apply them onto corporates logos technique (they performed experiments trying to find which layers are best for extracting content and style).  
It was found that there is a significant trade-off between content and style in order to maintain the logo readable. Since the researcher only wanted to transfer the font style, he suggests that this model is not suitable for this mission and that maybe it needs to be approached in a different way. Some of his results are visible in Figure 2.



Figure 2 – Taken from “Network of Steel: Neural Font Style Transfer from Heavy Metal to Corporate Logos“. The Figure shows the company logos with metal bands style.

Although Ter-Sarkisov [4] says this technique is not the best for the text, in the paper of Atarsaikhan et.al. [9] the researchers generate fonts by using neural style transfer. The VGG16 model was used to extract and determine the content representation of a source image and feature maps of a style image to synthesize them together.  
It was found that α (the coefficient of content representation) should be bigger than β (the coefficient of style representation) in order to give the content more importance.  
The researchers showed that it is possible to generate readable fonts using neural network approach. They observed that font style can only be transferable if style images contain multiple characters. Also, the researchers managed to transfer a non-font style to sources like icons and graphs. In addition, they found that if the difference between source and style is large the result characters will be illegible. Some of their results are illustrated in Figure 3.

Figure 3 – Taken from “Neural Font Style Transfer”. The Figure shows the results of transferring style to content image from English content and foreign languages.

The works of Ter-Sarkisov [4] and Atarsaikhan et.al. [5] study the style transfer on isolated words and letters and not full sentences or documents. In addition, they focus on font transformation and not the background and paper style that is very important in the field of historical documents. In the paper of Gomez et.al. [6], the researchers suggest a new architecture that learns the features that encode a certain text style and can transfer them to other text instances while preserving their content. In addition, they built a model that transfers style to only desired image pixels. The researchers used a VGG16 model for style transfer and two different models for selective style transfer (transfer style to only desired image pixels).  
They found out that the style transfer model is able to learn text styles as the characters shapes, line style, and colors, and to transfer it to an input text preserving the original characters. Some of their results shown in Figure 4 and Figure 5.

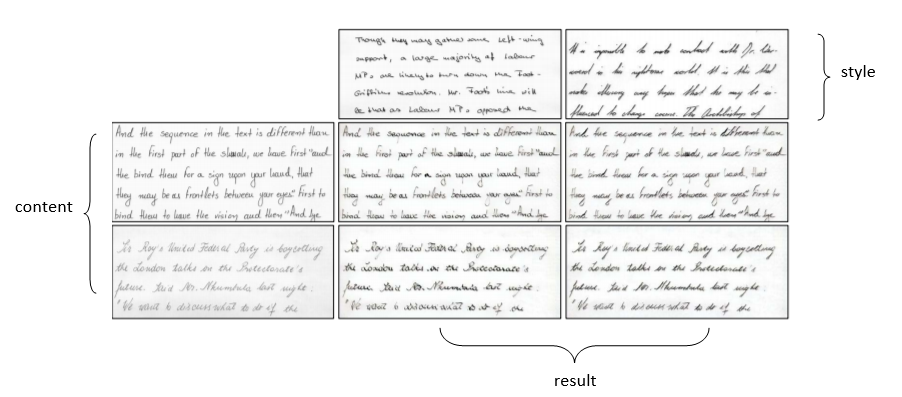


Figure 4 – Taken from “Selective Style Transfer for Text”. The Figure shows the result of their model to transfer content with given style.

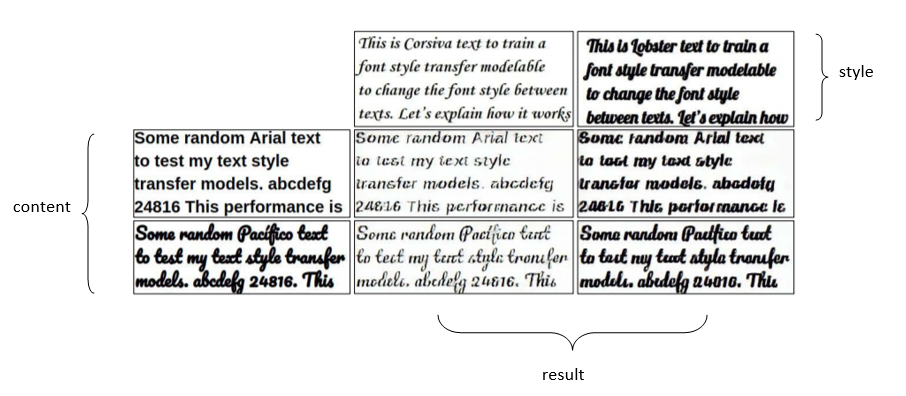


Figure 5 – Taken from “Selective Style Transfer for Text”. The Figure shows the result of their model to transfer content with given style.

Someresearchers combined the original model architecture with VGG19 with other techniques, like in the paper of Guan et.al. [7]. The work proposes a style conditioned generative adversarial network (SC-GAN). This network is used to transfer the styles of real handwriting images to skeleton images extracted from handwriting samples to generate photo-realistic text line images. In their paper the researchers used a VGG19 model to extract content and style from given images. First, given a content image they create a skeleton image. Then the skeleton and style image are given to a VGG19 model and combined after that by a generator that uses AdaIN and a discriminator GAN model. In this paper the researchers showed that they can create synthetic images, whose styles look quite similar with the corresponding style images. Some of their results are presented in Figure 6.

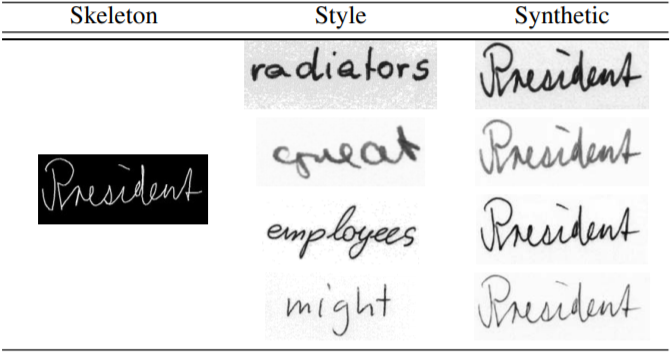


Figure 6 – Taken from "Improving Handwritten OCR with Augmented Text Line Images Synthesized from Online Handwriting Samples by Style-Conditioned GAN". The Figure shows the results of transferring style to skeleton image.

Anotherexample is the paper of Kang et.al. [8]. The researchers suggest a generator aimed at transferring writing style features from one sample to another in an image-to-image translation approach.   
Given a pair of handwritten word images (, ), the proposed module extracts content and style using the VGG19 model, and then combined by another model to generate a new handwritten word image. To combine the content and style, instead of using Gram matrix, they used a generator model that consists of two residual blocks with AdaIN normalization layers followed by four convolutional modules with nearest neighbor up-sampling, and a final tanh activation layer. Some of their results are illustrated in Figure 7, and as we can see from it, the researchers managed to combine style and content to a new image.

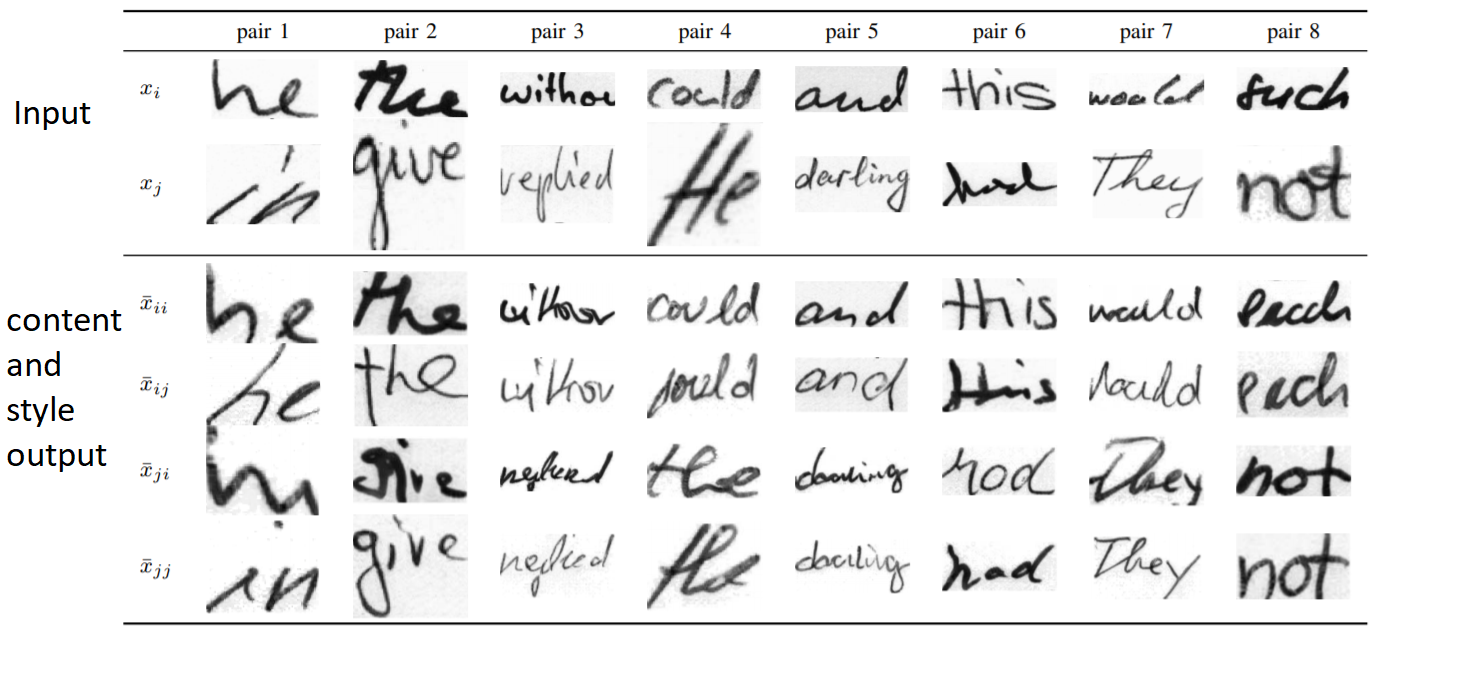


Figure 7 – Taken from “Distilling Content from Style for Handwritten Word Recognition”. The Figure shows the results of transferring style to content given style and content image.

Similar to Guan et.al. [7] and Kang et.al. [8], in the paper of Liu et.al. [9] the researchers propose a word image generating method called Synth-Text Transfer Network (STN).   
They use an Encoder-AdaIN-Decoder architecture, where the encoder is a pre-trained VGG-19 with first few layers fixed in order to encode both content and style images. The decoder is learned to invert the AdaIN output to the image spaces. The AdaIN layer is used to conduct style transfer in the feature space. Then, they compute the content loss and the style loss by using the same VGG encoder. The researchers found that AdaIN is the most proper module for arbitrary style transfer. However, they mentioned that their method has limitation in the case where the style image has complicated background texture or uncommon font. Some of their results are shown in Figure 8.

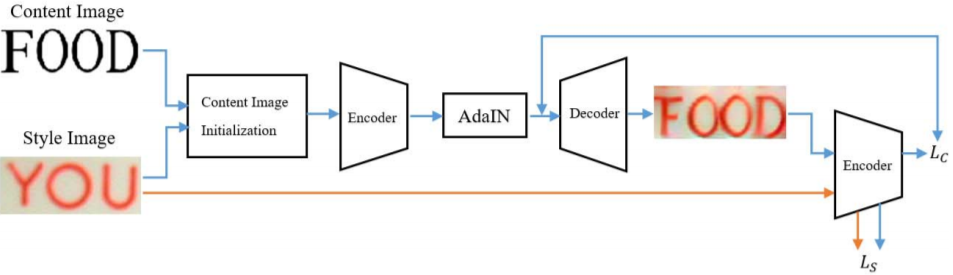


Figure 8 – Taken from “Synthesizing Scene Text Images for Recognition”. The Figure shows the Encoder-AdaIN-Decoder architecture of the proposed method.

In addition, there are different attempts to transfer image style in writing with different models and techniques.

Forexample, in the paper of Mayr et.al. [10] the researchers present a method for online handwriting synthesis, given a handwritten sample they produce a new style-adapted realistic-looking text.  
In this paper the researchers tested their model on words and sentences. In order to extract the content they use Graves’ algorithm (LSTM), and in order to extract the style and combine it with the content they used the pix2pix generator network. To conclude, the researchers managed to create a fully automatic method to imitate handwriting using spatial-temporal style transfer. Some of their results are shown in Figure 9.

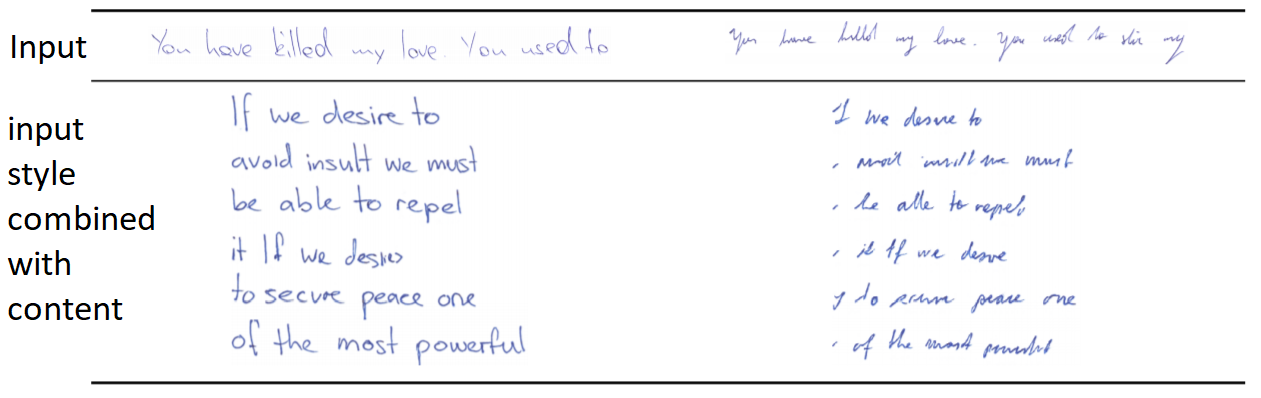


Figure 9 – Taken from “Spatio-Temporal Handwriting Imitation”. The Figure shows the results of transferring input handwriting style to different texts.

Anotherexample is the paper of Davis et.al. [11]. The researchers present a GAN for generating images of handwritten lines conditioned on arbitrary text and style vectors.  
Given three inputs, content, style and noise, the model generates handwriting.  
The researchers used a big and complex model made of six models in order to get the final result: (1) A generator network G to produce images from spaced text, a style vector and noise-based on StyleGAN; (2) A style extractor network S, that produces a style vector from an image and the recognition predictions; (3) A spacing network C, which predicts the horizontal text spacing based on the style vector; (4) A patch-based convolutional discriminator D; (5) A pretrained handwriting recognition network R to encourage image legibility and correct content; (6) A pretrained encoder E, to compute a perceptual loss.  
This method has presented a system capable of directly generating the pixels of a handwriting image of arbitrary length. Their model can extract a style from example images and their method does well at capturing the variations of global style in handwriting, such as slant and size. Some of their results are illustrated in Figure 10.

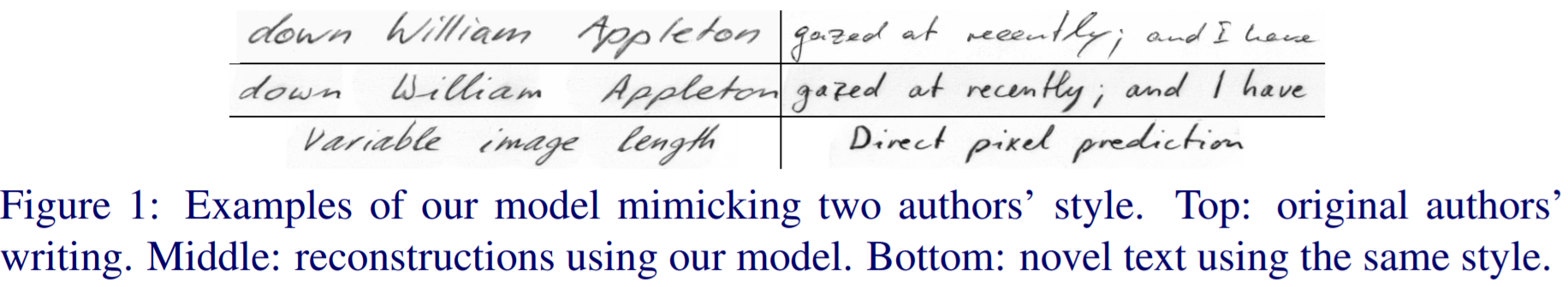


Figure 10 – Taken from “Text and Style Conditioned GAN for Generation of Offline Handwriting Lines”. The Figure shows the results of mimicking two authors style, top sentence is the original author writing bottom is generated sentence.

All the aforementioned papers focus more about the font transfer of the writing and less for the style of the document and the other aspects of document. In addition, none of the papers try to implement the transfer on a full document image. In the paper of Pondenkandath et.al. [12] the researchers apply document style transfer on historical documents. They use two models and compare results between them. The first model is GAN (cycleGAN) and the second model is VGG-19 CNN. The main motivation of this research is to generate massive synthetic datasets of "historical" documents which can be used for the training of document analysis systems. The researchers found that the synthetic images generated by the cycleGAN appear significantly better than those generated with NST. The researchers notice many similarities between the target domain samples and the synthetic samples. The overall style content of the target domain is well expressed. However, in a structural content point of view the initials are not well detected and expressed. The two column-mode is not at all expressed. When considering the synthetic documents produced with the NST, the structural content is better preserved. However, the style is mixed and standardized over the entire synthetic document, leading to the presence of a lot of colored artifacts. Also, the font does not change as compared to the synthetic images generated by the GAN. Some of their results are shown in Figure 11.



Figure 11 – Taken from “Historical Document Synthesis with Generative”. The Figure shows the source transform given target with cycleGAN and NST.

As we can see there are various researches dealing with synthesizing writing/fonts from style image into content image. These studies are a great base for our research, but are missing a few aspects of what we want to investigate. Most of the discussed above work/deal with words/sentences/short paragraphs but not with *full* document images. In addition, all these papers do not consider the background/paper style to be part of the wanted style to transfer and they work with grey/binary images. The only paper that tries to create a full document is the paper of Pondenkandath et.al. [12], but this work and tested only on Latin languages and focuses more about the font and less about the style of the document. In our project, the goal is to transfer historical style (color, degradation of document and more features). In addition, we want to explore this transformation with the original simple model using VGG19 and not using GAN, transformers, and complex models.

To conclude, from the researches available today it seems our research is possible. We will want to answer a few questions in our research. Our main research question is (1) “Is it possible to transfer style from a full historical document image to a modern document image by using the simple image transfer style model?”. Additional questions we would like to explore (2) “What is the ratio/trade-off between content and style?”, (3) “Can we perform style transfer on documents that are not written in the same language?”.

# **References:**

|  |  |
| --- | --- |
| [1] | L. A. Gatys, A. S. Ecker and M. Bethge, "A Neural Algorithm of Artistic Style," *CoRR,* 2015. |
| [2] | L. A. Gatys, A. S. Ecker and M. Bethge, "Image Style Transfer Using Convolutional Neural Networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 2414-2423. |
| [3] | Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. B. Girshick, S. Guadarrama and T. J. Darrell, "Caffe: Convolutional Architecture for Fast Feature Embedding," in *Proceedings of the 22nd ACM International Conference on Multimedia*, New York, NY, USA, Association for Computing Machinery, 2014, p. 675–678. |
| [4] | A. Ter-Sarkisov, "Network of Steel: Neural Font Style Transfer from Heavy Metal to Corporate Logos," in *Proceedings of the 9th International Conference on Pattern Recognition Applications and Methods (ICPRAM)*, Valletta, Malta, 2020, pp. 621-629. |
| [5] | G. Atarsaikhan, B. K. Iwana, A. Narusawa, K. Yanai and S. Uchida, "Neural Font Style Transfer," in *14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, Kyoto, Japan, 2017, pp. 51-56. |
| [6] | R. Gomez, A. F. Biten, L. Gomez, J. Gibert, D. Karatzas and M. Rusiñol, "Selective Style Transfer for Text," in *2019 International Conference on Document Analysis and Recognition (ICDAR)*, Sydney, Australia, 2019, pp. 805-812. |
| [7] | M. Guan, H. Ding, K. Chen and Q. Huo, "Improving Handwritten OCR with Augmented Text Line Images Synthesized from Online Handwriting Samples by Style-Conditioned GAN," in *17th International Conference on Frontiers in Handwriting Recognition (ICFHR)*, Dortmund, Germany, 2020. |
| [8] | L. Kang, P. Riba, M. Rusiñol, A. Fornés and M. Villegas, "Distilling Content from Style for Handwritten," in *17th International Conference on Frontiers in Handwriting Recognition (ICFHR).*, Dortmund, Germany, 2020. |
| [9] | H. Liu and A. Zhu, "Synthesizing Scene Text Images for Recognition with Style Transfer," in *2019 International Conference on Document Analysis and Recognition Workshops (ICDARW)*, Sydney, Australia, 2019, pp. 8-13. |
| [10] | M. Mayr, M. Stumpf, A. Nikolaou, M. Seuret, A. Maier and V. Christlein, "Spatio-Temporal Handwriting Imitation," 2020. |
| [11] | B. Davis, C. Tensmeyer, B. Price, C. Wigington, B. Morse and R. Jain, "Text and Style Conditioned GAN for Generation of Offline Handwriting Lines," 2020. |
| [12] | V. Pondenkandath, M. Alberti, M. Diatta, R. Ingold and M. Liwicki, "Historical Document Synthesis with Generative Adversarial Networks," in *2019 International Conference on Document Analysis and Recognition Workshops (ICDARW)*, Sydney, Australia, 2019, pp. 146-151. |