**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Historical documents can reveal a lot of information, such as form of writing, wording, content that did not exist and more.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 these documents that are usually written in diaries, pages or letters that have a final lifespan, there are certain periods that lack of historical documents.  
In this study we want to test whether it is possible to synthesize historical documents using an historical document and a source document (plain text) that we will synthesize into a historical document.

]אנא תוסיפו מוטיבציה, כגון, אלגוריתמי למידת מוכנה דורשים כמות ענקית של נתונים שלא נמצאים לרשותינו במקרה של מסכים היסטוריים. בנוסף, supervised learning דורש annotation- תהליך ארוך ומייגע[

Transferring the style from one image into another can be considered a problem of texture transfer. In texture transfer the goal is to synthesise 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. [2] and [1], 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 classifying 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.  
The researchers were able in this study 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.

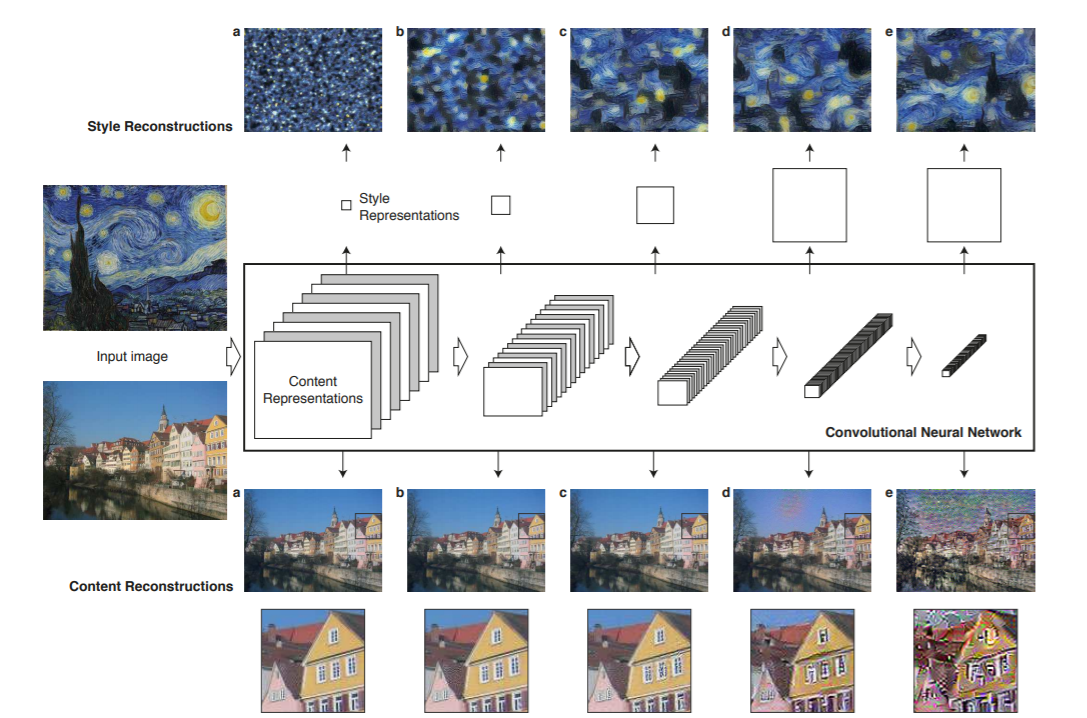


Figure 1 – Taken from “Image Style Transfer Using Convolutional Neural Networks”.  
The figure shows the content and style in the VGG network.

Basedon Gatys et.al papers [1][2] a few papers were released about image style transfer writing. For example, in the paper of Ter-Sarkisov [8] the researcher tried transferring 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 [1] 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)  
The researcher found there is a big tradeoff between content and style in order to maintain the logo readable. Because the researcher only wanted to transfer the font style, he suggests that this model is not suitable for this mission and that maybe he needs to approach 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.

Another example of this technique being used is in Atarsaikhan et.al. paper [9]. The researchers generate fonts by using neural style transfer. In this paper the researchers used a VGG16 model to extract and determine the content representation of a source image and feature maps of a style image to synthesize them together.  
In their paper the researchers 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.  
In this paper the researches showed 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 visible in Figure 3.

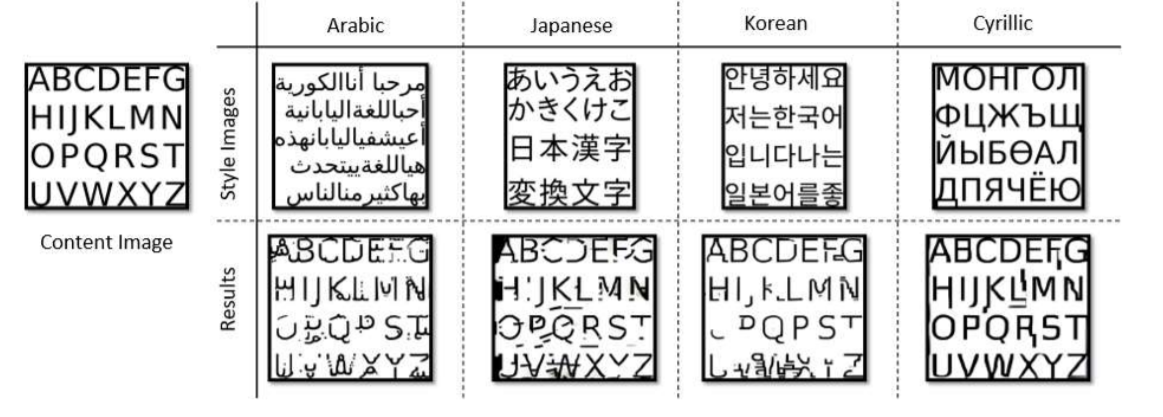


Figure 3 – Taken from “Neural Font Style Transfer”. The Figure shows the results of transferring style to content image from English content and foreign languages.

Someresearchers combined the original model architecture with VGG19 with other techniques, like in the paper of Guan et.al. [5]. The researchers propose 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 this 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 visible in Figure 4.

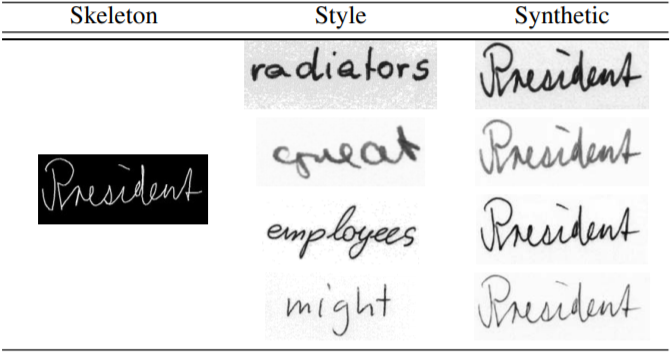


Figure 4 – taken from "Improving Handwritten OCR with Augmented Text Line Images Synthesized from Online Handwriting Samples by Style-Conditioned GAN" the Figure shows the results of transferring style to skeleton image.

Anotherexample is the paper of Kang et.al. [4]. 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, with a final tanh activation layer. Some of their results are visible in Figure 5,   
and as we can see from it the researchers managed to combine style and content to a new image.

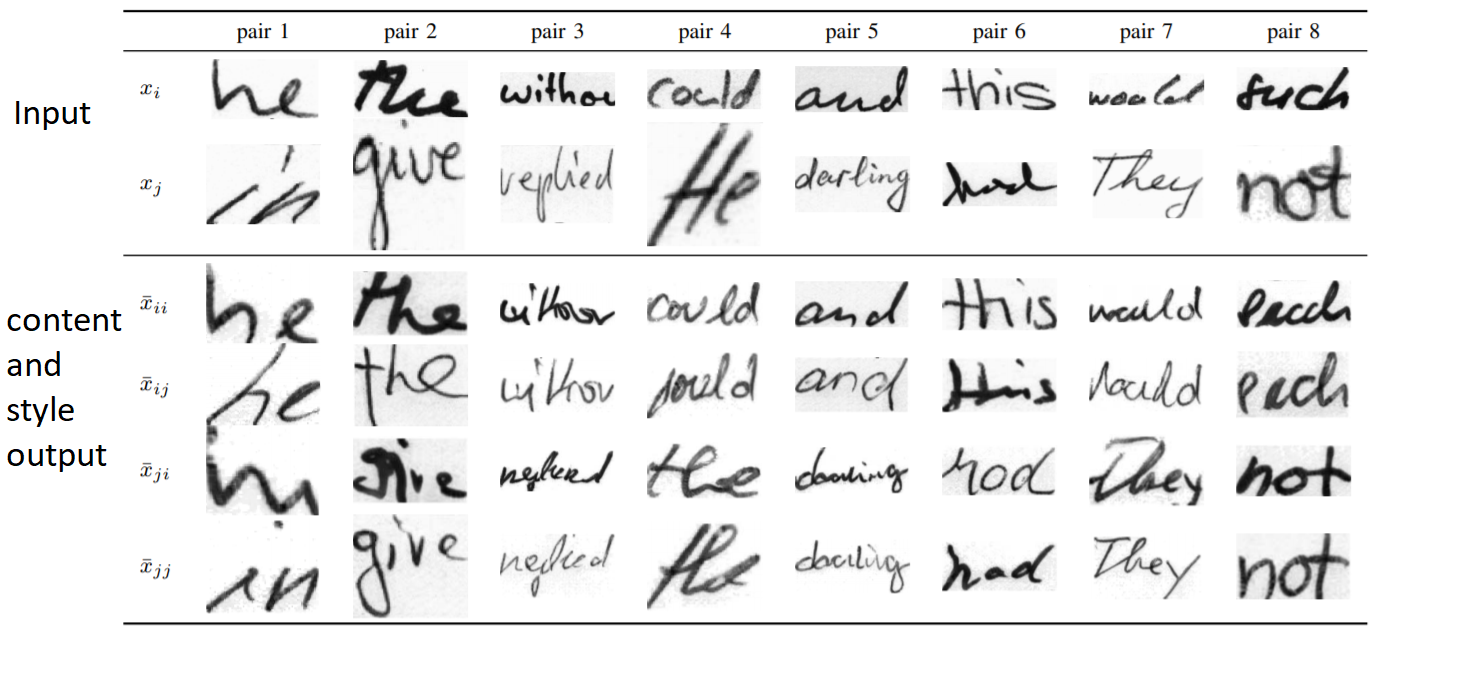


Figure 5 – taken “Distilling Content from Style for Handwritten Word Recognition” the Figure shows the results of transferring style to content given style and content image.

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

Forexample, In the paper of Mayr et.al. [6] 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 visible in Figure 6.

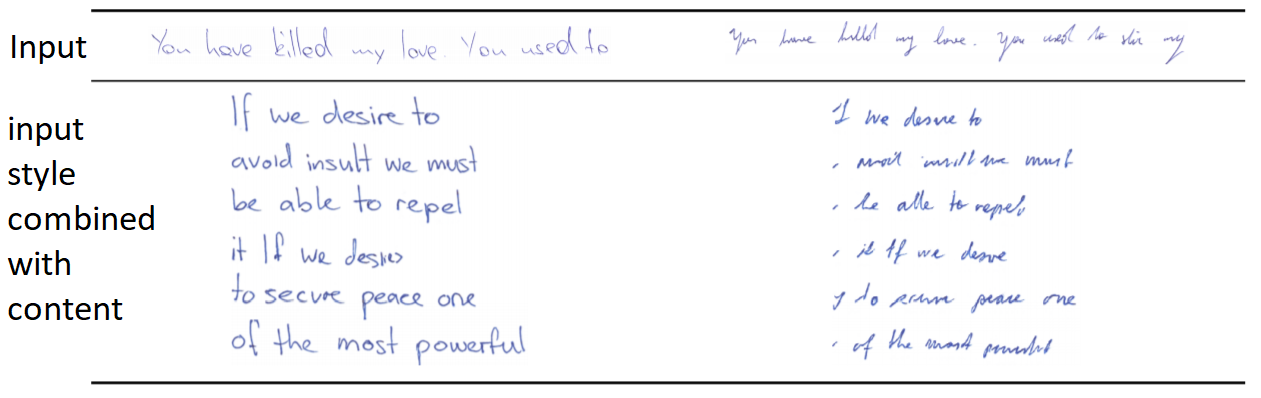


Figure 6 – taken “Spatio-Temporal Handwriting Imitation” the Figure shows the results of transferring input handwriting style to different texts.

Anotherexample is the paper of Davis et.al. [7]. The researchers present a GAN for generating images of handwritten lines conditioned on arbitrary text and style vectors.  
Given 3 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 visible in Figure 7.

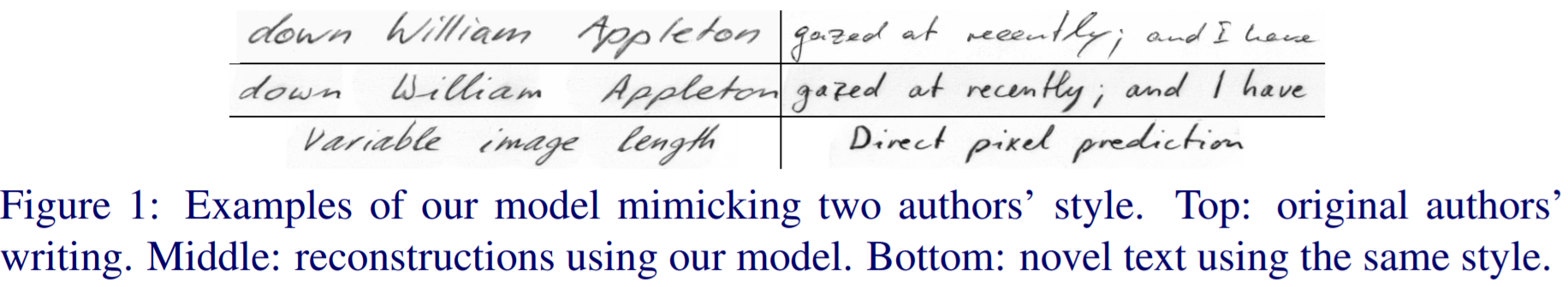


Figure 7 – taken “Text and Style Conditioned GAN for Generation of Offline Handwriting Lines” the Figure shows the results of mimicking two authors style, top sentence is the original author writing bottom is generated sentence.

As we can see there are various of researches in 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. All the discussed works\deal with words\sentences\short paragraphs but not with *full* document images. Also, 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. In this project, our goal is to transfer historical style this is a very important aspect to transfer. 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 to a modern document by using the simple image transfer style model?”. Additional questions we want to explore (2) “What is the ratio\trade-off between content and style?”, (3) “Can we transfer style from documents that are not written in same language?”

**References:**

[1] L. A. Gatys, A. S. Ecker and M. Bethge. "Image Style Transfer Using Convolutional Neural Networks". In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2414-2423. (2016).

[2] L. A. Gatys, A. S. Ecker, M. Bethge. "A Neural Algorithm of Artistic Style". In *Journal of Vision* 2016;16(12):326. (2016).

[3] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, T. Darrell. "Caffe: Convolutional Architecture for Fast Feature Embedding". In *Proceedings of the 22nd ACM international conference on Multimedia* (MM '14), pp. 675–678. (2014).

[4] L. Kang, P. Riba, M. Rusiñol, A. Fornés, M. Villegas. "Distilling Content from Style for Handwritten Word Recognition". In *17th International Conference on Frontiers in Handwriting Recognition (ICFHR)*. (2020).

[5] M. Guan, H. Ding, K. Chen, 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)*. (2020).

[6] M. Mayr, M. Stumpf, A. Nikolaou, M. Seuret, A. Maier, V. Christlein. "Spatio-Temporal Handwriting Imitation". (2020).

[7] B. Davis, C. Tensmeyer, B. Price, C. Wigington, B. Morse, R. Jain. "Text and Style Conditioned GAN for Generation of Offline Handwriting Lines". (2020).

[8] 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 - Volume 1*: ICPRAM, ISBN 978-989-758-397-1, pp. 621-629. (2020).

[9] G. Atarsaikhan, B. K. Iwana, A. Narusawa, K. Yanai, S. Uchida. "Neural Font Style Transfer". In *14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, pp. 51-56. (2017).