Historical Style Generator

BS-SE-21-89

Academic Advisor - Dr. Irina Rabaev

Yahav Bar David
yahavba@sce.ac.il

Leor Arial Rose
leorro@sce.ac.il

Summary

Historical documents can reveal a great deal of information about our past, such as, form of writing, wording, content that did not exist and more. In order to perform computational learning (Machine Learning) a huge amount of classified data (Classified Data) is needed. The process of creating classified data (Annotations) is expensive and tedious work, and therefore in the field of historical documents, the databases that exist for training models are small. These datasets do not allow training deep models to get high results.

In order to create a large database of data, in an easy way that requires less resources, it is necessary to create synthetic data.

In the "Historical Style Generator" project, we researched a method for creating synthetic historical data and developed a system (website) that allows each user to synthesize documents himself.

Our method is a deep learning method based on neural style transfer. In order to improve the results of the method, we used several techniques of computer vision, such as Binarization, Dilation and Image Processing.

The system is accessible to any user via GitHub and was developed using the following technologies: Tensorflow, Keras, FastApi, Python, Angular, Docker.

Keywords: Historical Documents, Classified Data, Synthetic Data, Deep Learning, Neural Style Transfer, Computer Vision, Binarization, Dilation, Image Processing.