**Abstract**

Optical Character Recognition (OCR) is a process in which text from scanned images is converted to computer enabled characters to be used in further computational analysis. This process works best on documents that contain highly visible and clean images to reliably extract information. However, not all documents are ready made to be scanned into OCR and require additional pre-processing steps to enhance the image sufficiently. These ‘noisy’ documents contain such elements as stains, marks, crumples, and other items. The objective of this project is to build an autoencoder capable of taking noisy images and removing these elements to improve OCR recognition. With a dataset from RVL-CDIP at Carnegie Melon University, our autoencoder is trained using a Convolutional Neural Network (CNN). This neural network learns through the process of convolution and pooling how to remove pixel information that is least relevant to reconstructing the final image. These weights from the trained model are used in the autoencoder to predict an enhanced version of the input image which improves the ability for characters to be properly recognized via OCR.