

Introduction

In this exercise, we developed a machine learning pipeline for keyword spotting in historical document images. The goal was to find the most similar words based on a query image in The George Washington Database. We used HOG features extracted from preprocessed word images and computed Dynamic Time Warping (DTW) distances between pairs.

Preprocessing

1. Cropping: We extracted word regions using SVG polygon coordinates.
2. Binarization: Applied Sauvola thresholding to binarize word images.
3. Normalization: Resized images to 100x100 pixels and normalized intensity to [0.1].

The preprocessed data will be used as a consistent input for feature extraction and similarity computation.

Feature Extraction

We used Histogram of Gradients (HOG) to encode word shapes:

- Resized images to 64x64 for HOG descriptor.
- Block size: 16x16
- stride: 8x8
- cell size: 8x8
- 9 orientation bins
- Features were L2-normalized

The feature vector from HOG is then used in DTW to measure similarity between query and target words.

Similarity Computation

For each query word, the distance to all other words in the same file were computed using Dynamic Time Warping (DTW):

$$DTW(x, y) = \frac{1}{n+m} \sum_{i,j} |x_i - y_j|$$

The smaller the distance the more similar word pairs are.

Example of top 5 most similar word pairs:

Query ID	Target ID	Distance
275-01-01	275-26-06	0.00275
275-01-01	275-05-03	0.002839
301-03-01	301-03-02	0.002893
273-01-01	273-20-05	0.002928
300-02-01	300-20-03	0.002933

Example of top 5 least similar word pairs:

Query ID	Target ID	Distance
272-02-01	272-10-06	0.0040
304-01-01	304-04-08	0.0040
272-02-01	272-23-05	0.0040
272-02-01	272-17-03	0.0040
277-02-01	277-09-12	0.0040

Conclusion

The proposed method successfully retrieves words based on visual similarity, independent of text transcription. DTW with HOG features allows flexible matching despite variations in handwriting or word length.