SelfDoc: Self-Supervised Document Representation Learning
Supplementary Material

Peizhao Li[1], Jiuxiang Gu[2], Jason Kuen[2], Vlad I. Morariu[2], Handong Zhao[2],
Rajiv Jain[2], Varun Manjunatha[2], Hongfu Liu[1]
[1]Brandeis University, [2]Adobe Research
{peizhaoli,hongfuliu}@brandeis.edu
{jigu,kuen,morariu,hazhao,rajijain,vmanjuna}@adobe.com

A. Visualization of Modality-Adaptive Attention

We visualize the attention mechanism in Fig. 1. The provided samples show that it gives attention scores for each modality adaptively on different types of documents. Some documents with heavy-word or fewer clues in vision have a larger value in w_{lang}, while some forms with multiple font styles or unrecognizable hand-written enjoy a larger value in w_{visn}.

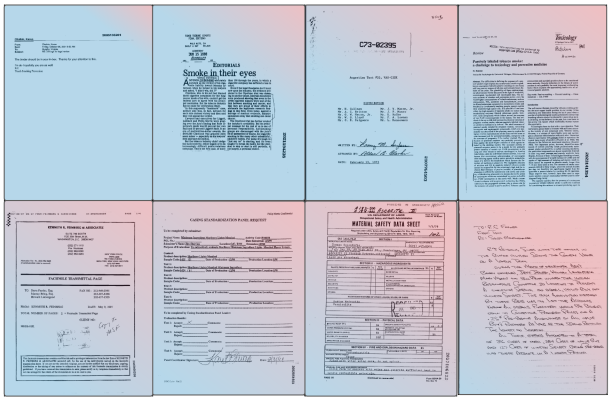

Figure 1. Visualization of Modality-Adaptive Attention on document classification. The size of covering area in blue and red represents the value in w_{lang} and w_{visn}, respectively.

B. Experiments in Document Clustering

B.1. Evaluative metrics

$$\text{Acc.} = \frac{\sum_{i=1}^{n} \mathbb{1}_{y_i=\text{map}(\hat{y}_i)}}{n},$$

$$\text{NMI} = \frac{\sum_{i,j} n_{ij} \log \frac{n \cdot n_{ij}}{n_{i+} \cdot n_{+j}}}{\sqrt{(\sum_i n_{i+} \log \frac{n_{i+}}{n})(\sum_j n_{j+} \log \frac{n_{+j}}{n})}},$$

where $\mathbb{1}$ is the indicator function, and $\text{map}(\hat{y}_i)$ is permutation mapping function that maps each cluster label $\hat{y}_i$ to the ground truth label y_i using linear sum assignment, n_{ij}, n_{i+} and n_{+j} represent the co-occurrence number and cluster size of i-th and j-th clusters in the obtained partition and ground truth, respectively, and n is the total data instance number.

B.2. Label sets

Cluster = 4: {email, form, handwritten, letter}
Cluster = 6: {email, form, handwritten, letter, news article, resume}
Cluster = 8: {email, form, handwritten, letter, news article, questionnaire, resume, scientific publication}
Cluster = 10: {email, file folder, form, handwritten, letter, news article, questionnaire, resume, scientific publication, specification}
Cluster = 12: {email, file folder, form, handwritten, letter, memo, news article, questionnaire, resume, scientific publication, scientific report, specification}