Fast(er) Reconstruction of Shredded Text Documents via Self-Supervised Deep Asymmetric Metric Learning

Supplementary Material

Paper ID 10016

March 29, 2020

1 Local Samples Nearest Neighbors

An interesting way to verify how the models are pairing complementary patterns is by fixing 32×32 samples (query samples) from one of the boundaries and recovering samples of the complementary side. As illustrated in Figure 1, one can select $\mathbf{x}_r$ as a query sample and try to recover the top-1 $\mathbf{x}_l$'s, i.e., that the sample of the left boundary which minimizes the distance to the anchor in the embedding space. Figure 9 in our manuscript has shown some queries for both $\mathbf{x}_r$ and $\mathbf{x}_l$ restricted to one shredded document of the test collection. Here, we mixed samples from 3 documents and, similarly, show 28 query samples and their respective top-3 complementary samples (distance increasing from top to bottom).

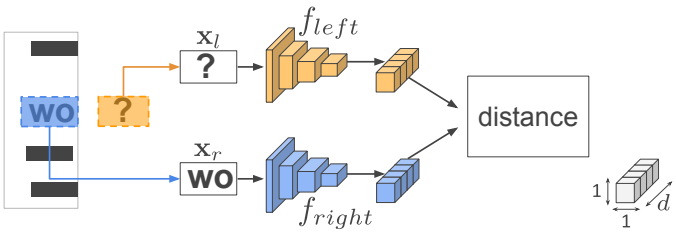

Figure 1: Querying $\mathbf{x}_l$ samples by fixing $\mathbf{x}_r$.

Figure 2: Local samples nearest neighbors. In the top row, the largest square is the "query" sample (before binarization) followed, below, by its binary version and its 3 nearest neighbors side-by-side (with the closest in the top row). The blue and orange samples were projected by f_{right} and f_{left}, respectively. The bottom row shows some examples in which the "query" is projected by the f_{left} instead.

2 Reconstruction of D2

The dataset D2 comprises 20 single-page documents, totaling 505 shreds. Figure 3 shows the reconstruction of the entire D2 dataset, i.e., after mixing all shreds. The shreds were placed side-by-side according to the solution (permutation) computed with the proposed metric learning-based method which achieved the accuracy of 97.22%. The pairwise compatibility evaluation took less than 4 minutes.

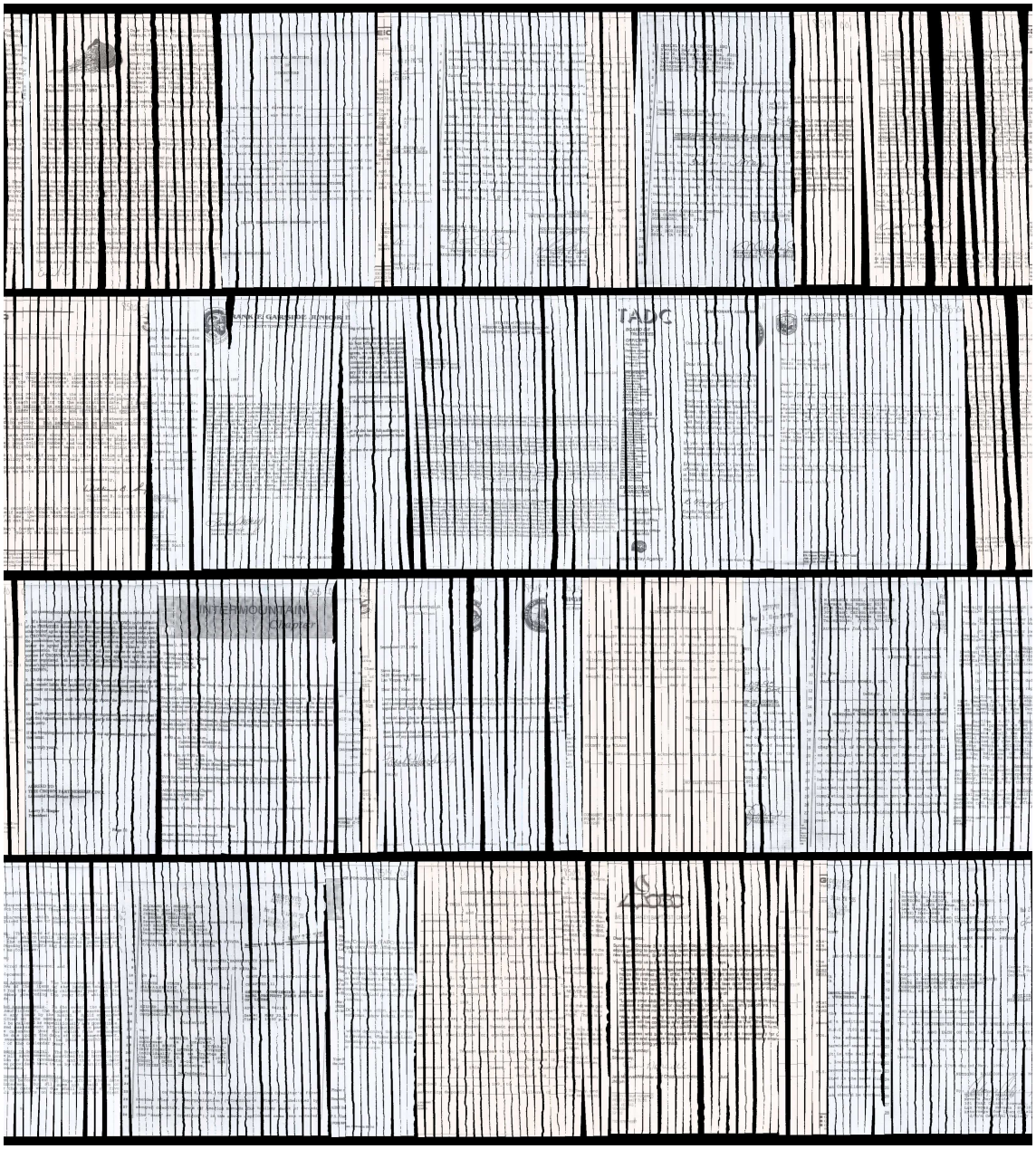

Figure 3: Reconstruction of D2. The generated image was split into 4 parts for better visualization.

3 Embedding Space

Figure 2 of our manuscript illustrates the embedding space onto which the local samples are projected. For a more concrete view of this space, four charts (Figures 4 to 7) were plotted showing local embeddings produced from a real-shredded document (25 shreds).

For each chart, there is a single anchor embedding (in blue), which was produced from an anchor sample $\mathbf{x}_r$ randomly cropped from the right boundary of an arbitrary shred. The other

points (embeddings) in the chart (in orange) corresponds to the samples from the other 24 shreds vertically aligned with the anchor sample, i.e., those which are candidates to match the anchor sample. Notice that the embeddings are numbered according to the shred they belong to, being $0, 1, 2, \ldots, 24$ the ground-truth order of the document. Therefore, the anchor (blue point) indicated by s should match the embedding (orange point) indicated by $s + 1$ (a dashed line linking the respective points was made in each chart).

For 2-D visualization, embeddings in the original space ($\mathbb{R}^{128}$) were projected to the plane by using t-SNE [1, 4]. It is worthy to mention that we analyzed the produced charts to ensure that pairwise distances in $\mathbb{R}^2$ are roughly consistent with those in the original space. Also, no vertical alignment between shreds was performed.

3.1 Case 1

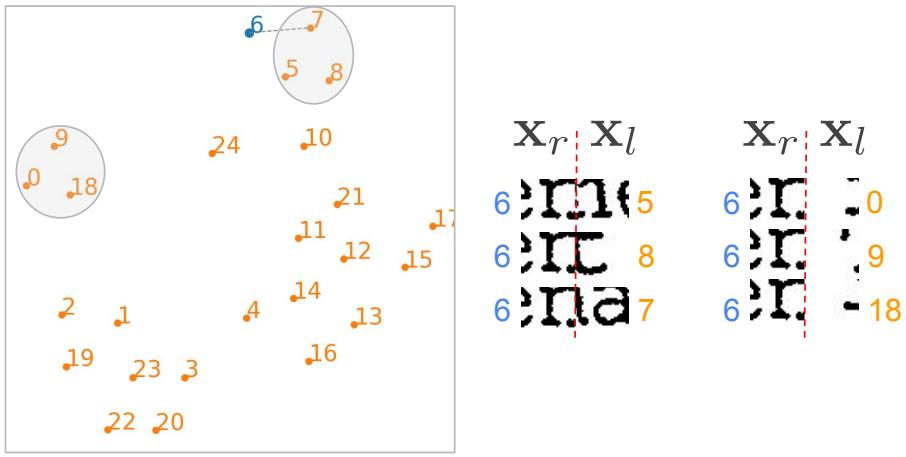

Figure 4: Case 1.

In Figure 4, samples from two clusters ($\{5, 7, 8\}$ and $\{0, 9, 18\}$) were shown at the right side of the 2-D chart. Although the pairing $(6, 8)$ looks incompatible based on the knowledge of the Latin alphabet, we noticed that the vertical alignment and the emerging horizontal were essential for their close positioning. For the cluster $\{0, 9, 18\}$, it is interesting to note that the information (black pixels) in the $\mathbf{x}_l$ samples is concentrated in the last columns.

3.2 Case 2

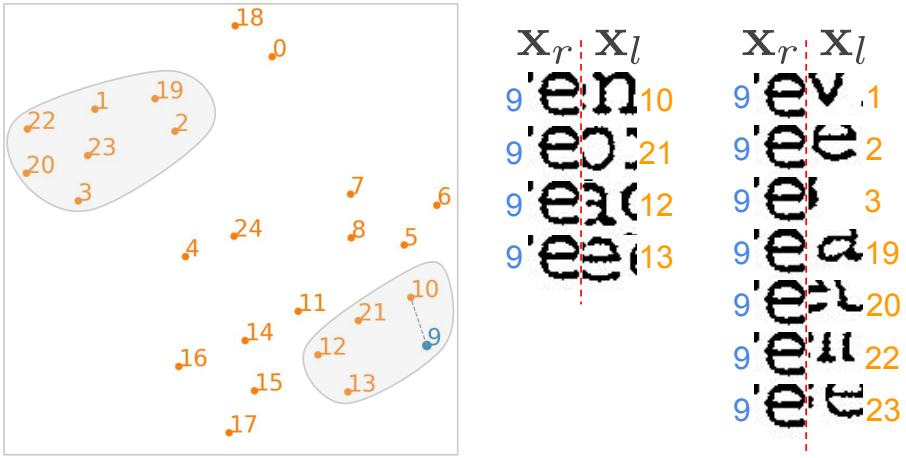

Figure 5: Case 2.

In Figure 5, two clusters were illustrated. As in the previous case, the vertical alignment plays an import role in the positioning of the embeddings. From the cluster $\{1, 2, 3, 19, 20, 22, 23\}$, it can

be observed that the $\mathbf{x}_l$ samples are similarly shifted up compared to the baseline of the anchor. Finally, although the unrealistic pairing $(9, 12)$ yields a distance superior to $(9, 10)$, they are kept close due to the vertical alignment and the emerging connections (three horizontal lines).

3.3 Case 3

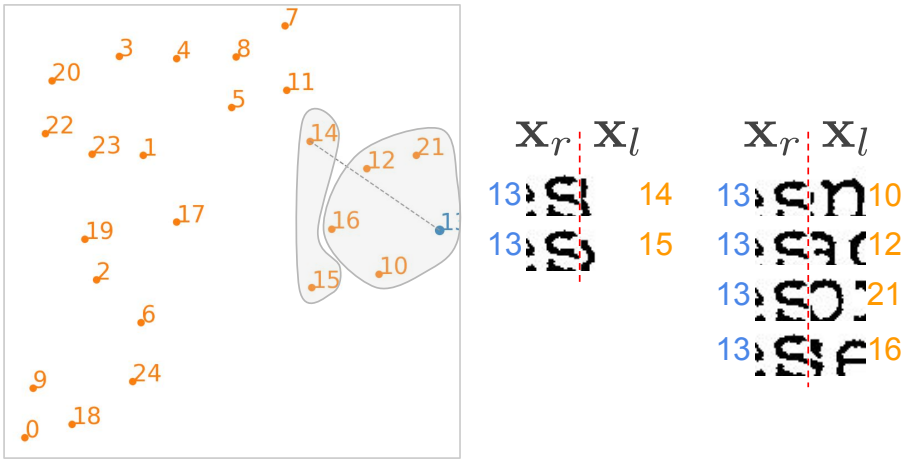

Figure 6: Case 3.

The third case, illustrated in Figure 6, depicts a situation where a couple of matchings are better evaluated than the corrected one, i.e., $(13, 14)$. In addition to the realistic appearance of the competitors (pairings formed with samples in $\{10, 12, 16, 21\}$), we noticed that the low number of blacks in $\mathbf{x}_l$ (and analogously in $\mathbf{x}_r$) leads to some instability in the projection. This issue may occur in very particular situations where the cut happens almost in the blank area following a symbol and either there are no symbols in the sequence or the blank area is large enough so that $\mathbf{x}_l$ is practically blank.

3.4 Case 4

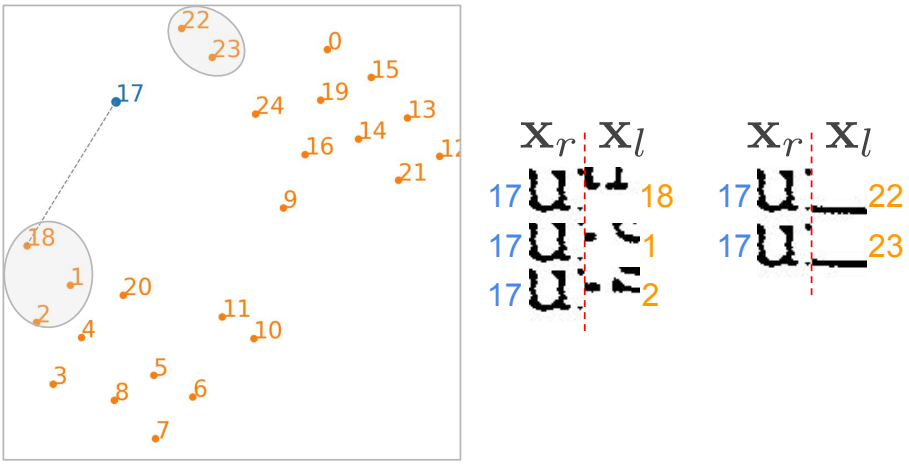

Figure 7: Case 4.

The last selected case (Figure 7) emphasizes the relevance of the vertical alignment stage of our method (Section 3.2 of the manuscript). By observing the correct pairing $(17, 18)$, it is noticeable vertical misalignment between the shreds. The samples 22 and 23 are very similar, and therefore they are mapped closely in the embedding space. Also, these samples are good competitors because of the alignment with the anchor's baseline. Finally, it can be observed (as in Case 2) the clustering induced by the displaced content of $\mathbf{x}_l$.

4 Sensitivity analysis w.r.t. sample size

As [2], we use small samples (32×32) to explore features at text line (local) level based under the assumption of weak feature correlation across text lines. In a previous investigation, we observed that the accuracy of [2] decreases for larger samples. This is also verified in our method when the sample height (s_y) is increased, as seen in Figure 8.

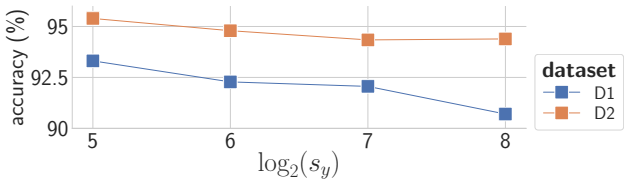

Figure 8: Reconstruction accuracy w.r.t. to the sample height (s_y).

5 Statistical test

Considering a threshold of 5%, the proposed method was statistically equivalent to [2] in D2 and superior to [3] in both D1 and D2. Table 1 shows the p-values of the page-wise paired t-test.

	D1 ∪ D2	D1	D2
Proposed vs. [2]	0.016	0.007	**0.522**
Proposed vs. [3]	0.000	0.000	0.004

Table 1: Page-wise paired t-test.

References

[1] Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-SNE. *Journal of machine learning research*, 9(Nov):2579–2605, 2008.

[2] T. Paixão, R. Berriel, M. Boeres, C. Badue, A. De Souza, and T. Oliveira-Santos. A deep learning-based compatibility score for reconstruction of strip-shredded text documents. In *Conf. on Graphics, Patterns and Images (SIBGRAPI)*, pages 87–94, 2018.

[3] T. Paixão, M. Boeres, C. Freitas, and T. Oliveira-Santos. Exploring Character Shapes for Unsupervised Reconstruction of Strip-shredded Text Documents. *IEEE Transactions on Information Forensics and Security*, 14(7):1744–1754, 2018.

[4] Laurens Van Der Maaten. Accelerating t-SNE using tree-based algorithms. *The Journal of Machine Learning Research*, 15(1):3221–3245, 2014.