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

Thiago M. Paixão*[1,2], Rodrigo F. Berriel[2], Maria C. S. Boeres[2], Alessandro L. Koerich[3], Claudine Badue[2], Alberto F. De Souza[2] and Thiago Oliveira-Santos[2]

[1]Federal Institute of Espírito Santo (IFES), Serra, Brazil
[2]Federal University of Espírito Santo (UFES), Vitória, Brazil
[3]École de Technologie Supérieure (ÉTS), Montreal, Canada

Abstract

The reconstruction of shredded documents consists in arranging the pieces of paper (shreds) in order to reassemble the original aspect of such documents. This task is particularly relevant for supporting forensic investigation as documents may contain criminal evidence. As an alternative to the laborious and time-consuming manual process, several researchers have been investigating ways to perform automatic digital reconstruction. A central problem in automatic reconstruction of shredded documents is the pairwise compatibility evaluation of the shreds, notably for binary text documents. In this context, deep learning has enabled great progress for accurate reconstructions in the domain of mechanically-shredded documents. A sensitive issue, however, is that current deep model solutions require an inference whenever a pair of shreds has to be evaluated. This work proposes a scalable deep learning approach for measuring pairwise compatibility in which the number of inferences scales linearly (rather than quadratically) with the number of shreds. Instead of predicting compatibility directly, deep models are leveraged to asymmetrically project the raw shred content onto a common metric space in which distance is proportional to the compatibility. Experimental results show that our method has accuracy comparable to the state-of-the-art with a speed-up of about 22 *times for a test instance with* 505 *shreds (*20 *mixed shredded-pages from different documents).*

1. Introduction

Paper documents are of great value in forensics because they may contain supporting evidence for criminal investigation (e.g., fingerprints, bloodstains, textual information). Damage on these documents, however, may hamper or even prevent their analysis, particularly in cases of chemical destruction. Nevertheless, recent news [9] shows that documents are still being physically damaged by hand-tearing or using specialized paper shredder machines (mechanically shredding). In this context, a forensic document examiner (FDE) is typically required to reconstruct the original document for further analysis.

To accomplish this task, FDEs usually handle paper fragments (shreds) manually, verifying the compatibility of pieces and grouping them incrementally. Despite its relevance, this manual process is time-consuming, laborious, and potentially damaging to the shreds. For these reasons, research on automatic digital reconstruction has emerged since the last decade [14, 32]. Traditionally, hand-tearing and mechanical-shredding scenarios are addressed differently since shreds' shape tends to be less relevant in the latter. Instead, shreds' compatibility is almost exclusively determined by appearance features, such as color similarity around shreds extremities [17, 29].

As with the mechanical shredding, ad hoc strategies have been also developed for binary text documents to cope with the absence of discriminative color information [6, 11, 16, 30]. More recently, Paixão *et al*. [22] substantially improved the state-of-the-art in terms of accuracy on the reconstruction of strip-shredded text documents, i.e., documents cut in the longitudinal direction only. Nevertheless, time efficiency is a bottleneck because shreds' compatibility demands a costly similarity assessment of character shapes. In a follow-up work [21], the group proposed a deep learning-based compatibility measure, which improved the accuracy even further as well as the time efficiency of the reconstruction. In [21], shreds' compatibility is estimated pairwise by a CNN trained in a self-supervised way, learn-

*Corresponding author: paixao@gmail.com.

ing from intact (non-shredded) documents. Human annotation is not required at any stage of the learning process. A sensitive issue, however, is that model inference is required whenever a pair of shreds has to be evaluated. Although this is not critical for a low number of shreds, scalability is compromised for a more realistic scenario comprising hundreds/thousands of shreds from different sources.

To deal with this issue, we propose a model in which the number of inferences scales linearly with the number of shreds, rather than quadratically. For that, the raw content of each shred is projected onto a space in which the distance metric is proportional to the compatibility. The projection is performed by a deep model trained using a metric learning approach. The goal of metric learning is to learn a distance function for a particular task. It has been used in several domains, ranging from the seminal work of the Siamese networks [5] in signature verification, to an application of the triplet loss [33] in face verification [28], to the lifted structured loss [20], to the recent connection with mutual information maximization [31] and many others. Unlike most of these works, however, the proposed method does not employ the same model to semantically different samples. In our case, right and left shreds are (asymmetrically) projected by two different models onto a common space. After that, the distances between the right and left shreds are measured, the compatibility matrix is built and passed on to the actual reconstruction. To enable fair comparisons, the actual reconstruction was performed by coupling methods for compatibility evaluation to an external optimizer. The experimental results show that our method achieves accuracy comparable to the state-of-the-art (97.22%) while taking only 3.73 minutes to reconstruct 20 mixed pages with 505 shreds compared to 1 hour and 20 minutes of [21], i.e., a speed-up of ≈ 22 times.

In summary, the main contributions of our work are:

1. This work proposes a compatibility evaluation method leveraging metric learning and the asymmetric nature of the problem;

2. The proposed method does not require manual labels (trained in a self-supervised way) neither real data (the model is trained with artificial data);

3. The experimental protocol is extended from a single-page to a more realistic and time demanding scenario with a multi-page multi-document reconstruction;

4. Our proposal scales the inference linearly rather than quadratically as in the current state-of-the-art, achieving a speed-up of ≈ 22 times for 505 shreds, and even more for more shreds.

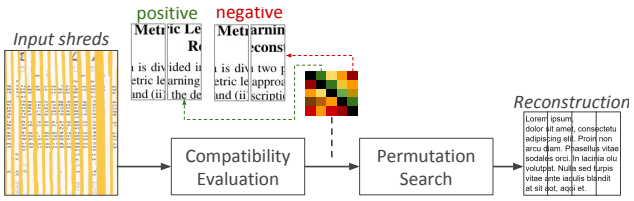

Figure 1. Classical approach for automatic document reconstruction. Shreds' compatibility is evaluated pairwise and then an optimization search process is conducted (based on the compatibility values) in order to find the shreds' permutation that best represents the original document.

2. Problem Definition

For simplicity of explanation, let us first consider the scenario where all shreds belong to the same page: single-page reconstruction of strip-shredded documents. Let $\mathcal{S} = \{s_1, s_2, \ldots, s_n\}$ denote the set of n shreds resulting from longitudinally shredding (strip-cut) a single page. Assume that the indices determine the ground-truth order of the shreds: s_1 is the leftmost shred, s_2 is the right neighbor of s_1, and so on. A pair (s_i, s_j) – meaning s_j placed right after s_i – is said to be "positive" if $j = i + 1$, otherwise it is "negative". A solution of the reconstruction problem can be represented as a permutation $\pi_{\mathcal{S}} = (s_{\pi_1}, s_{\pi_2}, \ldots, s_{\pi_n})$ of $\mathcal{S}$. A perfect reconstruction is that for which $\pi_i = i$, for all $i = 1, 2, \ldots, n$.

Automatic reconstruction is classically formulated as an optimization problem [18, 24] whose objective function derives from pairwise compatibility (Figure 1). Compatibility or cost, depending on the perspective, is given by a function $c : \mathcal{S} \times \mathcal{S} \rightarrow \mathbb{R}$ that quantifies the (un)fitting of two shreds when placed side-by-side (order matters). Assuming a cost interpretation, $c(s_i, s_j)$, $i \neq j$, denotes the cost of placing s_j to the right of s_i. In theory, $c(s_i, s_j)$ should be low when $j = i + 1$ (positive pair), and high for other cases (negative pairs). Typically, $c(s_i, s_j) \neq c(s_j, s_i)$ due to the asymmetric nature of the reconstruction problem.

The cost values are the inputs for a search procedure that aims to find the optimal permutation $\pi^*_{\mathcal{S}}$, i.e., the arrangement of the shreds that best resembles the original document. The objective function C to be minimized is the accumulated pairwise cost computed only for consecutive shreds in the solution:

$$C(\pi_{\mathcal{S}}) = \sum_{i=1}^{n-1} c(s_{\pi_i}, s_{\pi_{i+1}}). \quad (1)$$

The same optimization model can be applied in the reconstruction of several shredded pages from one or more documents (multi-page reconstruction). In a stricter formulation, a perfect solution in this scenario can be represented by a sequence of shreds which respects the ground-truth order in each page, as well as the expected order (if any) of

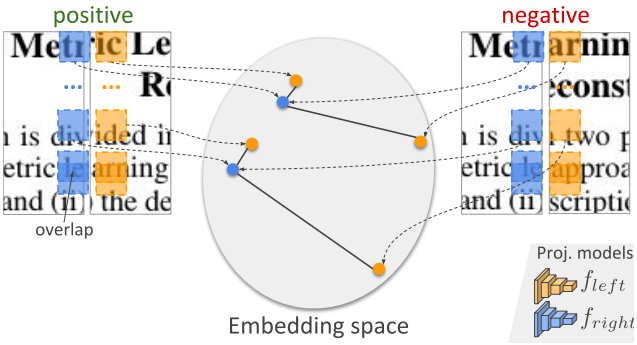

Figure 2. Metric learning approach for shreds' compatibility evaluation. Embeddings generated from compatible regions are expected to be closer in the embedding space, whereas those from non-fitting regions are expected to be mapped far from each other.

the pages themselves. If page order is not relevant (or does not apply), the definition of a positive pair of shreds can be relaxed, such that a pair (s_i, s_j) is also positive if s_i and s_j are, respectively, the last and first shreds of different pages, even for $j \neq i+1$. The optimization problem of minimizing Equation 1 has been extensively investigated in literature, mainly using genetic algorithms [4, 10, 11, 35] and other metaheuristics [2, 25, 27]. The focus of this work is, nevertheless, on the compatibility evaluation between shreds (i.e., the function c), which is critical to lead the search towards accurate reconstructions.

To address text documents, literature started to evolve from the application of pixel-level similarity metrics [3, 11, 17], which are fast but inaccurate, towards stroke continuity analysis [12, 23] and symbol-level matching [22, 34]. Strokes continuity across shreds, however, cannot be ensured since physical shredding damages the shreds' borders. Techniques based on symbol-level features, in turn, tend to be more robust. However they may struggle to segment symbols in complex documents, and to cope efficiently with the wide variability of symbols' shape and size. These issues have been addressed in [21], wherein deep learning has been successfully used for accurate reconstruction of strip-shredded documents. Nonetheless, the large number of network inferences required for compatibility assessment hinders scalability for multi-page reconstruction.

This work addresses precisely the scalability issue. Although our self-supervised approach shares some similarities with their work, the training paradigm is completely distinct since the deep models here do not provide compatibility (or cost) values. Instead, deep models are used to convert pixels into embedding representations, so that a simple distance metric can be applied to measure shreds' compatibility. This is better detailed in the next section.

3. Compatibility Evaluation via Deep Metric Learning

The general intuition behind the proposed approach for compatibility evaluation is illustrated in Figure 2. The underlying assumption is that two side-by-side shreds are globally compatible if they locally fit each other along the touching boundaries. The local approach relies on small samples (denoted by $\mathbf{x}$) cropped from the boundary regions. Instead of comparing pixels directly, the samples are first converted to an intermediary representation (denoted by $\mathbf{e}$) by projecting them onto a common embedding space $\mathbb{R}^d$. Projection is accomplished by two models (CNNs): f_{left} and f_{right}, $f_\bullet : \mathbf{x} \mapsto \mathbf{e}$, specialized on the left and right boundaries, respectively.

Assuming that these models are properly trained, boundary samples (indicated by the orange and blue regions in Figure 2) are then projected, so that embeddings generated from compatible regions (mostly found on positive pairings) are expected to be closer in this metric space, whereas those from non-fitting regions should be farther apart. Therefore, the global compatibility of a pair of shreds is measured in function of the distances between corresponding embeddings. More formally, the cost function in Equation 1 is such that:

$$c(s_i, s_j) \propto \phi(\mathbf{e}_i, \mathbf{e}_j), \tag{2}$$

where $\mathbf{e}_\bullet$ represents the embeddings associated with the shred $s_\bullet$, and ϕ is a distance metric (e.g., Euclidean).

The interesting property of this evaluation process is that the projection step can be decoupled from the distance computation. In other words, the process scales linearly since each shred is processed once by each model, and pairwise evaluation can be performed with the embeddings produced. Before diving into the details of the evaluation, we first describe the self-supervised learning of these models. Then, a more in-depth view of evaluation will be presented, including the formal definition of a cost function that composes the objective function in Equation 1.

3.1. Learning Projection Models

For producing the shreds' embeddings, the models f_{left} and f_{right} are trained simultaneously with small $s \times s$ samples. The two models have the same fully convolutional architecture: a base network for feature extraction appended with a convolutional layer. The added layer is intended to work as a fully connected layer when the base network is fed with $s \times s$ samples. Nonetheless, weight sharing is disabled since models specialize on different sides of the shreds, hence deep asymmetric metric learning. The base network comprises the first three convolutional blocks of SqueezeNet [13] architecture (i.e., until the *fire3* block).

SqueezeNet has been effectively used in distinguishing between valid and invalid symbol patterns in the context

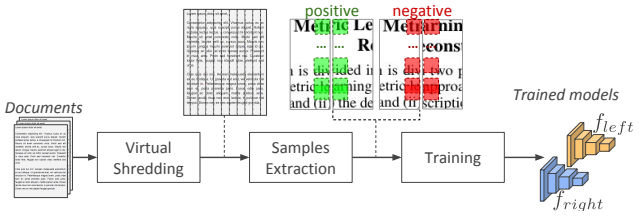

Figure 3. Self-supervised learning of the models with samples extracted from digital documents.

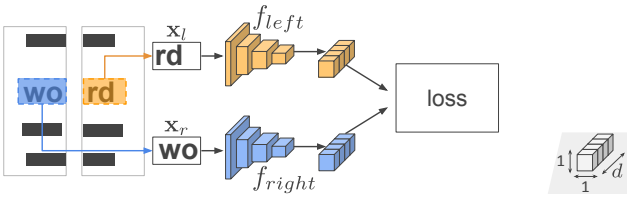

Figure 4. Learning projection models for shreds' compatibility evaluation. The models are jointly trained with sample pairs guided by the contrastive loss function. The input vectors for the loss function are encoded as $1 \times 1 \times d$ tensors.

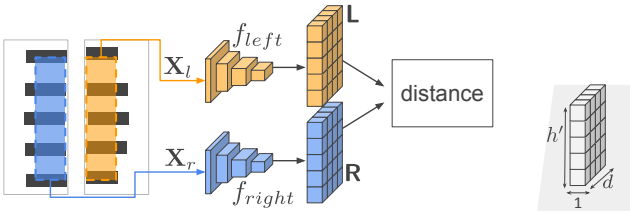

Figure 5. Compatibility evaluation of a pair of shreds. Local embeddings, represented by the $h' \times 1 \times d$ tensors $\mathbf{L}$ and $\mathbf{R}$, are extracted along the boundary regions. Compatibility is a real value given by the squared Euclidean distance between $\mathbf{L}$ and $\mathbf{R}$ (computed over the flattened tensors).

of compatibility evaluation [21]. Nevertheless, preliminary evaluations have shown that the metric learning approach is more effective with shallower models, which explains the use of only the first three blocks. For projection onto $\mathbb{R}^d$ space, a convolutional layer with d filters of dimensions $s/4 \times s/4$ (base network's dimensions when fed with $s \times s$ samples) and sigmoid activation was added to the base network.

Figure 3 outlines the self-supervised learning of the models with samples extracted from digital documents. First, the shredding process is simulated so that the digital documents are cut into equally shaped rectangular "virtual" shreds. Next, shreds of the same page are paired side-by-side and sample pairs are extracted top-down along the touching edge: one sample from the s rightmost pixels of the left shred (r-sample), and the other from the s leftmost pixels of the right shred (l-sample). Since shreds' adjacency relationship is provided for free with virtual shredding, sample pairs can be automatically labeled as "positive" (green boxes) or "negative" (red boxes). Self-supervision comes exactly from the fact that labels are automatically acquired by exploiting intrinsic properties of the data.

Training data comprise tuples $(\mathbf{x}_r, \mathbf{x}_l, y)$, where $\mathbf{x}_r$ and $\mathbf{x}_l$ denote, respectively, the r- and l-samples of a sample pair, and y is the associated ground-truth label: $y = 1$ if the sample pair is positive, and $y = 0$, otherwise. Training is driven by the contrastive loss function [7]:

$$\mathcal{L}(f_{left}, f_{right}, \mathbf{x}_l, \mathbf{x}_r, y) = \frac{1}{2}\{(1-y) \cdot dist^2 + y \cdot [\max(0, m - dist)]^2\}, \quad (3)$$

where $dist = \|f_{left}(\mathbf{x}_l) - f_{right}(\mathbf{x}_r)\|_2$, and m is the margin parameter. For better understanding, an illustration is provided in Figure 4. The models handle a positive sample pair that, together, composes the pattern "word". Since it is positive ($y = 1$), the loss value would be low if the resulting embeddings are close in $\mathbb{R}^d$, otherwise, it would be high. Note that weight-sharing would result in the same loss value for the swapped samples (pattern "rdwo"), which is undesirable for the reconstruction application. Implementation details of the sample extraction and training procedure are described in experimental methodology (Section 4.3).

3.2. Compatibility Evaluation

In compatibility evaluation, shreds' embedding and distance computation are two decoupled steps. Figure 5 presents a joint view of these two steps for better understanding of the model's operation. Strided sliding window is implicitly performed by the fully convolutional models. To accomplish this, two vertically centered $h \times s$ regions of interest are cropped from the shreds' boundaries (s is the sample size): $\mathbf{X}_r$, comprising the s rightmost pixels of the left shred, and $\mathbf{X}_l$, comprising the s leftmost pixels of the right shred. Inference on the models produces $h' \times 1 \times d$ feature volumes represented by the tensors $\mathbf{L} = f_{left}(\mathbf{X}_l)$ (l-embeddings) and $\mathbf{R} = f_{right}(\mathbf{X}_r)$ (r-embeddings). The h' rows from the top to the bottom of the tensors represent exactly the top-down sequence of d-dimensional local embeddings illustrated in Figure 2.

If it is assumed that vertical misalignment among shreds is not significant, compatibility could be obtained by simply computing $\|\mathbf{R} - \mathbf{L}\|_2$. For a more robust definition, shreds can be vertically "shifted" in the image domain to account for misalignment [21]. Alternatively, we propose to shift the tensor $\mathbf{L}$ "up" and "down" δ units (limited to δ_{max}) in order to determine the best-fitting pairing, i.e., that which yields the lowest cost. This formulation helps to save time since it does not require new inferences on the models. Given a tensor $\mathbf{T} = (T_{i,j,k})_{h' \times 1 \times d}$, let $\mathbf{T}_{a:b} = (T_{i,j,k})_{a \le i \le b,\ j=1,\ 1 \le k \le d}$ denote a vertical slice from row a to b. Let $\mathbf{R}^{(i)}$ and $\mathbf{L}^{(j)}$ represent, respectively, the r- and l-embeddings for a pair of shreds (s_i, s_j). When shifts are

restricted to the upward direction, compatibility is defined by the function:

$$c_{\uparrow}(s_i, s_j) = \min_{0 \leq \delta \leq \delta_{max}} \left\| \mathbf{R}^{(i)}_{1:1+n_{rows}} - \mathbf{L}^{(j)}_{1+\delta:1+n_{rows}+\delta} \right\|_2, \tag{4}$$

where $n_{rows} = h' - \delta_{max}$ is the number of rows effectively used for distance computation. Analogously, for the downward direction:

$$c_{\downarrow}(s_i, s_j) = \min_{0 \leq \delta \leq \delta_{max}} \left\| \mathbf{R}^{(i)}_{1+\delta:1+n_{rows}+\delta} - \mathbf{L}^{(j)}_{1:1+n_{rows}} \right\|_2 \tag{5}$$

Finally, the proposed cost function is a straightforward combination of Equations (4) and (5):

$$c(s_i, s_j) = \min(c_{\uparrow}(s_i, s_j), c_{\downarrow}(s_i, s_j)). \tag{6}$$

Note that, if δ_{max} is set to 0 (i.e., shifts are not allowed), then $n_{rows} = h'$, therefore:

$$c(i, j) = c_{\uparrow}(i, j) = c_{\downarrow}(i, j) = \left\| \mathbf{R}^{(i)} - \mathbf{L}^{(j)} \right\|_2. \tag{7}$$

4. Experimental Methodology

The experiments aim to evaluate the accuracy and time performance of the proposed method, as well as to compare with the literature in document reconstruction focusing on the deep learning method proposed by Paixão *et al*. [21] (hereafter referred to as Paixão-b). For this purpose, we followed the basic protocol proposed in [22] in which the methods are coupled to an exact optimizer and tested on two datasets (D1 and D2). Two different scenarios are considered here: single- and multi-page reconstruction.

4.1. Evaluation Datasets

D1. Produced by Marques and Freitas [17], it comprises 60 shredded pages scanned at 300 dpi. Most pages are from academic documents (e.g., books and thesis), part of such pages belonging to the same document. Also, 39 instances have only textual content, whereas the other 21 have some graphic element (e.g., tables, diagrams, photos). Although a real machine (Cadence FRG712) has been used, the shreds present almost uniform dimensions and shapes. Additionally, the text direction is nearly horizontal in most cases.

D2. This dataset was produced by Paixão *et al*. [22] and comprises 20 single-page documents (legal documents and business letters) from the ISRI-Tk OCR collection [19]. The pages were shredded with a Leadership 7348 strip-cut machine and their respective shreds were arranged side-by-side onto a support yellow paper sheet, so that they could be scanned at once and, further, easily segmented from background. In comparison to D1, the shreds of D2 have less uniform shapes and their borders are significantly more damaged due to the machine blades wear. Besides, the handling of the shreds before scanning caused slight rotation and (vertical) misalignment between the shreds. These factors render D2 as a more realistic dataset compared to D1.

4.2. Accuracy Measure

Similar to the neighbor comparison measure [1], the accuracy of a solution is defined here as the fraction of adjacent pairs of shreds which are "positive". For multi-reconstruction, the relaxed definition of "positive" is assumed (as discussed in Section 2), i.e., the order in which the pages appear is irrelevant. More formally, let $\pi_{\mathcal{S}} = (s_{\pi_1}, s_{\pi_2}, \ldots, s_{\pi_n})$ be a solution for the reconstruction problem for a set of shreds $\mathcal{S}$. Then, the accuracy of $\pi_{\mathcal{S}}$ is calculated as

$$\text{accuracy}(\pi_{\mathcal{S}}) = \frac{1}{n-1} \sum_{i=1}^{n-1} \mathbf{1}[(s_{\pi_i}, s_{\pi_{i+1}}) \text{ is positive}], \tag{8}$$

where $\mathbf{1}[\cdot]$ denotes the 0-1 indicator function.

4.3. Implementation Details

Sample Extraction. Training data consist of 32×32 samples extracted from 100 binary documents (forms, emails, memos, etc.) scanned at 300 dpi of the IIT-CDIP Test Collection 1.0 [15]. For sampling, the pages are split longitudinally into 30 virtual shreds (amount estimated for the usual A4 paper shredders). Next, the shreds are individually thresholded with Sauvola's algorithm [26] to cope with small fluctuations in pixel values of the original images. Sample pairs are extracted page-wise, which means that the samples in a pair come from the same document. The extraction process starts with adjacent shreds in order to collect positive sample pairs (limited to 1,000 pairs per document). Negative pairs are collected subsequently, but limited to the number of positive pairs. During extraction, the shreds are scanned from top to bottom, cropping samples every two pixels. Pairs with more than 80% blank pixels are considered ambiguous, and then they are discarded for future training. Finally, the damage caused by mechanical shredding is roughly simulated with the application of salt-and-pepper random noise on the two rightmost pixels of the r-samples, and on the two leftmost pixels of the l-samples.

Training. The training stage leverages the sample pairs extracted from the collection of 100 digital documents. From the entire collection, the sample pairs of 10 randomly picked documents are reserved for validation where the best-epoch model should be selected. By default, the embeddings dimension d is set to 128. The models are trained from scratch (i.e., the weights are randomly initialized) for 100 epochs using the stochastic gradient descent (SGD) with a learning rate of 10^{-1} and mini-batches of size

256. After each epoch, the models' state is stored, and the training data are shuffled for the new epoch (if any). The best-epoch model selection is based on the ability to project positive pairs closer in the embedding space, and negative pairs far. This is quantified via the standardized mean difference (SMD) measure [8] as follows: for a given epoch, the respective f_{left} and f_{right} models are fed with the validation sample pairs and the distances among the corresponding embeddings are measured. Then, the distance values are separated into two sets: $dist^+$, comprising distances calculated for positive pairs, and $dist^-$, for negative ones. Ideally, the difference between the mean values of the two sets should be high, while the standard deviations within the sets should be low. Since these assumptions are addressed in SMD, the best f_{left} and f_{right} are taken as those which maximize $\mathrm{SMD}(dist^+, dist^-)$.

4.4. Experiments

The experiments rely on the trained models f_{left} and f_{right}, as well as on the Paixão-b's deep model. The latter was retrained (following the procedure described in [21]) on the CDIP documents to avoid training and testing with documents of the same collection (ISRI OCR-Tk). In practice, no significant change was observed in the reconstruction accuracy with this procedure.

The shreds of the evaluation datasets were also binarized [26] to keep consistency with training samples. The default parameters of the proposed method includes $d = 128$ and $\delta_{max} = 3$. Non-default assignments are considered in two of the three conducted experiments, as better described in the following subsections.

Single-page Reconstruction. This experiment aims to show whether the proposed method is able to individually reconstruct pages with accuracy similar to Paixão-b, and how the time performance of both methods is affected when the vertical shift functionality is enabled since it increases the number of pairwise evaluations. To this intent, the shredded pages of D1 and D2 were individually reconstructed with the proposed and Paixão-b methods, first using their default configuration, and after disabling the vertical shifts (in our case, it is equivalent to set $\delta_{max} = 0$). Time and accuracy were measured for each run. For a more detailed analysis, time was measured for each reconstruction stage: projection (pro) – applicable only for the proposed method –, pairwise compatibility evaluation (pw), and optimization search (opt).

Multi-page Reconstruction. This experiment focuses on the scalability with respect to time while increasing the number of shreds in multi-page reconstruction. In addition to the time performance, it is essential to confirm whether the accuracy of both methods remains comparable. Rather than individual pages, there are two large reconstruction instances in this experiment: the $1{,}370$ mixed shreds of D1 and the 505 mixed shreds of D2. Each instance was reconstructed with the proposed and Paixão-b methods, but now only with their default configuration (i.e., vertical shifts enabled). Accuracy and time (segmented by stage) were measured. Additionally, time processing was estimated for different instance sizes based on the average elapsed time observed for D2.

Method	D1 ∪ D2	D1	D2
Proposed	**93.71 ± 11.60**	**93.14 ± 12.93**	**95.39 ± 6.02**
Paixão-b [21]	96.28 ± 5.15	96.78 ± 4.44	94.78 ± 6.78
Paixão *et al*. [22]	74.85 ± 22.50	71.85 ± 23.14	83.83 ± 18.12
Marques and Freitas [17]	23.90 ± 17.95	29.18 ± 17.43	8.05 ± 6.60

Table 1. Single-page reconstruction performance: average accuracy ± standard deviation (%).

Sensitivity Analysis. The last experiment assesses how the proposed method is affected (time and accuracy) by testing with different embedding dimensions (d): $2, 4, 8, \ldots, 512$. Note that this demands the retraining of f_{left} and f_{right} for each d. After training, the D1 and D2 instances were individually reconstructed, and then accuracy and time processing were measured.

4.5. Experimental Platform

The experiments were carried out in an Intel Core i7-4770 CPU @ 3.40GHz with 16GB of RAM running Linux Ubuntu 16.04, and equipped with a TITAN X (Pascal) GPU with 12GB of memory. Implementation[1] was written in Python 3.5 using Tensorflow for training and inference, and OpenCV for basic image manipulation.

5. Results and Discussion

5.1. Single-page Reconstruction

A comparison with the literature for single-page reconstruction of strip-shredded documents is summarized in the Table 1. Given the clear improvement in the performance, the following discussions will focus on the comparison with [21]. The box-plots in Figure 6 show the accuracy distribution obtained with both the proposed method and Paixão-b for single-page reconstruction. Likewise [21], we also observe that vertical shifts affect only D2 since the D1's shreds are practically aligned (vertical direction). The methods did not present significant difference in accuracy for the dataset D2. For D1, however, Paixão-b slightly outperformed ours: the proposed method with default configuration (vertical shift "on") yielded accuracy of $93.14 \pm 12.88\%$ (arithmetic mean ± standard deviation), while Paixão-b achieved $96.78 \pm 4.44\%$. The higher variability in our approach is

[1] https://github.com/thiagopx/deeprec-cvpr20.

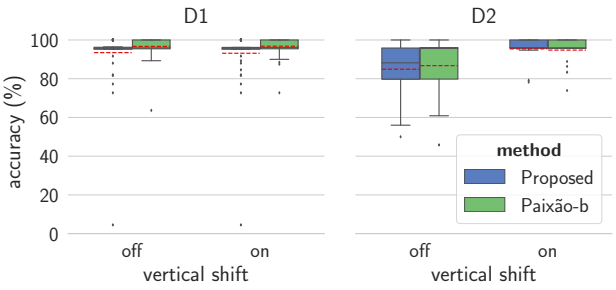

Figure 6. Accuracy distribution for single-page reconstruction with the proposed and Paixão-b methods. Accuracies are calculated document-wise and the average values are represented by the red dashed lines.

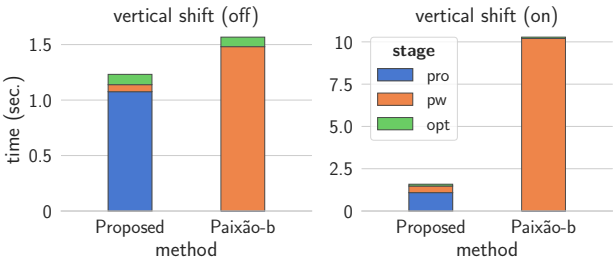

Figure 7. Time performance for single-page reconstruction. The stacked bars represent the average elapsed time for each reconstruction stage: projection (pro), pairwise compatibility evaluation (pw), and optimization search (opt).

mainly explained by the presence of documents with large areas covered by filled graphic elements, such as photos and colorful diagrams (which were not present in the training). By disregarding these cases (12 in a total of 60 samples), the accuracy of our method increases to 95.88%, and the standard deviation drops to 3.84%.

Time performance is shown in Figure 7. The stacked bars represent the average elapsed time in seconds (s) for each reconstruction stage: projection (pro), pairwise compatibility evaluation (pw), and optimization search (opt). With vertical shift disabled (left chart), the proposed method spent much more time producing the embeddings (1.075s) than in pairwise evaluation (0.063s) and optimization (0.092s). Although Paixão-b does not have the cost of embedding projection, pairwise evaluation took 1.481s, about 23 times the time elapsed in the same stage in our method. This difference becomes more significant (in absolute values) when the number of pairwise evaluations increases, as it can be seen with the enabling of vertical shifts (right chart). In this scenario, pairwise evaluation took 0.389s in our method, against the 10.197s spent in Paixão-b (approx. 26 times slower). Including the execution time of the projection stage, our approach yielded a speed-up of almost 7 times for compatibility evaluation. Note that, without vertical shifts, the accuracy of Paixão-b would drop

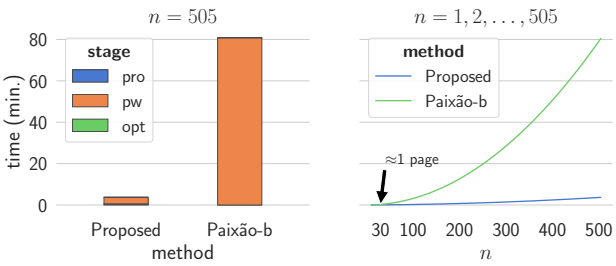

Figure 8. Time performance for multi-page reconstruction. Left: the time demanded in each stage to reconstruct D2 entirely ($n = 505$ shreds). Right: predicted processing time in function of the number of shreds.

from 94.77% to 86.74% in D2.

Finally, we provide an insight into what the embedding space might look like by showing a local sample and its three nearest neighbors. As shown in Figure 9, the models tend to form pairs that resemble something realistic. It is worth noting that the samples are very well aligned vertically, even in cases where the sample is shifted slightly to the top or bottom and the letters are appearing only in half (see more samples in the Supplementary Material).

5.2. Multi-page Reconstruction

For multi-page reconstruction, the proposed method achieved 94.81% and 97.22% of accuracy for D1 and D2, respectively, whereas Paixão-b achieved 97.08% and 95.24%. Overall, both methods yielded high-quality reconstructions with low difference in accuracy (approx. ± 2 p.p.), which is an indication that their accuracy is not affected by the increase of instances.

Concerning time efficiency, however, the methods behave notably different, as evidenced in Figure 8. The left chart shows the average elapsed time of each stage to process the 505 shreds of D2. In this context, with a larger number of shreds, the optimization cost became negligible when compared to the time required for pairwise evaluation. Remarkably, Paixão-b demanded more than 80 minutes to complete evaluation, whereas our method took less than 4 minutes (speed-up of approx. 22 times). Based on the average time for the projection and the pairwise evaluation, estimation curves were plotted (right chart) indicating the predicted processing time in function of the number of shreds (n). Viewed comparatively, the growing of the proposed method's curve (in blue) seems to be linear, although pairwise evaluation time (not the number inferences) grows quadratically with n. In summary, the greater the number of shreds, the higher the speed-up ratio.

5.3. Sensitivity Analysis

Figure 10 shows, for single-page reconstruction, how accuracy and time processing (mean values over pages) are

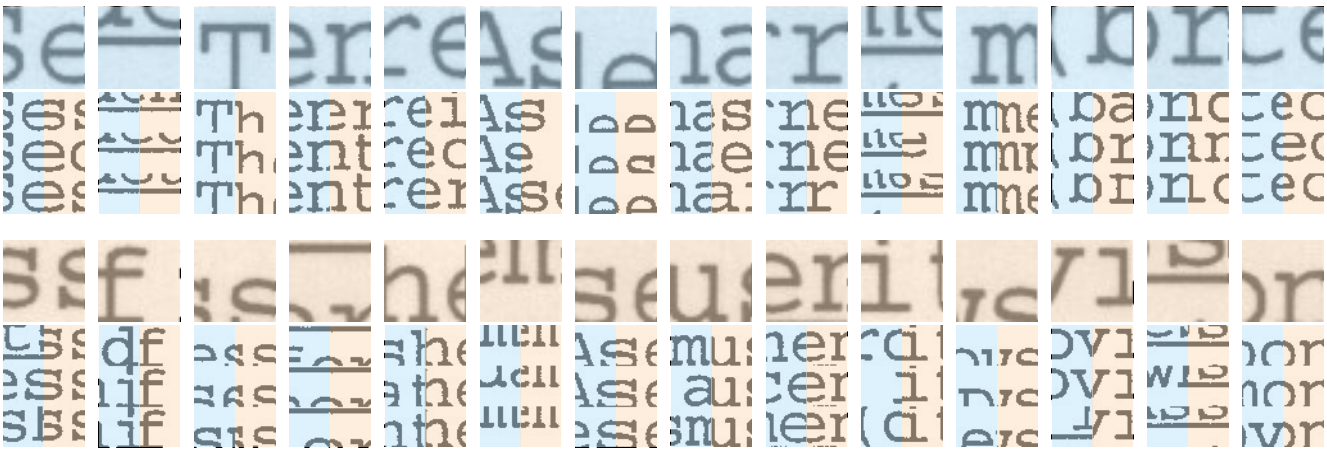

Figure 9. Local samples nearest neighbors. In the top row, the largest square is the "query" sample (before binarization) followed, below, by its binary version and its three nearest neighbors side-by-side (with the closest in the top row). The blue and orange samples were projected by f_{right} and f_{left}, respectively. The bottom row shows some examples in which the "query" is projected by the f_{left} instead.

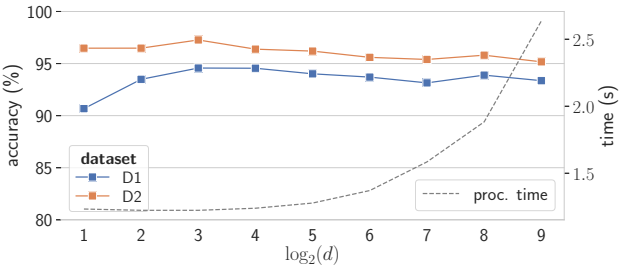

Figure 10. Sensitivity analysis with respect to embeddings dimension (d). The best accuracy was observed for $d = 8$: 94.57% and 97.27% for D1 and D2, respectively. This reduced embedded size yielded a reduction of 23% on processing time.

affected by the embedding dimension (d). Remarkably, projecting onto 2-D space ($d = 2$) is sufficient to achieve average accuracy superior to 90%. The highest accuracies were observed for $d = 8$: 94.57% and 97.27% for D1 and D2, respectively. Also, the average reconstruction time for $d = 8$ was 1.224s, which represents a reduction of nearly 23% when compared to the default value (128). For higher dimensions, accuracy tends to decay slowly (except for $d = 256$). Overall, the results suggest that there is space for improvement on accuracy and processing time by focusing on small values of d, which will be better investigated in future work.

6. Conclusion

This work addressed the problem of reconstructing mechanically-shredded text documents, more specifically the critical part of evaluating compatibility between shreds. Focus was given to the time performance of the evaluation. To improve it, we proposed a deep metric learning-based method as a compatibility function in which the number of inferences scales linearly rather than quadratically [21] with the number of shreds of the reconstruction instance. In addition, the proposed method is trained with artificially generated data (i.e., does not require real-world data) in a self-supervised way (i.e., does not require annotation).

Comparative experiments for single-page reconstruction showed that the proposed method can achieve accuracy comparable to the state-of-the-art with a speed-up of ≈ 7 times on compatibility evaluation. Moreover, the experimentation protocol was extended to a more realistic scenario in this work: multi-page multi-document reconstruction. In this scenario, the benefit of the proposed approach is even greater: our evaluation compatibility method takes less than 4 minutes for a set of 20 pages, compared to the approximate time of 1 hour and 20 minutes (80 minutes) of the current state-of-the-art (i.e., a speed-up of ≈ 22 times), while preserving a high accuracy (97.22%). Additionally, we show that the embedding dimension is not critical to the performance of our method, although a more careful tuning can lead to better accuracy and time performance.

Future work should include the generalization of the proposed method to other types of cut (e.g., cross-cut and hand-torn), as well as to other problems related to jigsaw puzzle solving [1].

Acknowledgements

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001. We thank NVIDIA for providing the GPU used in this research. We also acknowledge the scholarships of Productivity on Research (grants 311120/2016-4 and 311504/2017-5) supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, Brazil).

References

[1] F. Andaló, G. Taubin, and S. Goldenstein. PSQP: Puzzle solving by quadratic programming. *IEEE Trans. Patt. Anal. Mach. Intell.*, 39(2):385–396, 2017.

[2] H. Badawy, E. Emary, M. Yassien, and M. Fathi. Discrete grey wolf optimization for shredded document reconstruction. In *Int. Conf. Adv. Intell. Sys. Inf.*, pages 284–293, 2018.

[3] J. Balme. Reconstruction of shredded documents in the absence of shape information. Technical report, Working paper, Dept. of Comp. Sci., Yale University, USA, 2007.

[4] B. Biesinger, C. Schauer, B. Hu, and G. Raidl. Enhancing a Genetic Algorithm with a Solution Archive to Reconstruct Cross Cut Shredded Text Documents. In *Int. Conf. on Comput. Aided Sys. Theory*, pages 380–387, 2013.

[5] J. Bromley, I. Guyon, Y. LeCun, E. Säckinger, and R. Shah. Signature Verification using a "Siamese" Time Delay Neural Network. In *Adv. in Neural Inf. Process. Syst.*, 1994.

[6] G. Chen, J. Wu, C. Jia, and Y. Zhang. A pipeline for reconstructing cross-shredded English document. In *Int. Conf. on Image, Vision and Comput.*, pages 1034–1039, 2017.

[7] S. Chopra, R. Hadsell, and Y. LeCun. Learning a similarity metric discriminatively, with application to face verification. In *Conf. on Comput. Vision and Pattern Recognit. (CVPR)*, pages 539–546, 2005.

[8] J. Cohen. Statistical Power Analysis for the Behavioral Sciences. *Technometrics*, 31(4):499–500, 1988.

[9] J. Delk. FBI is reconstructing shredded documents obtained during Cohen raid. *The Hill*, May 2018.

[10] Y. Ge, Y. Gong, W. Yu, X. Hu, and J. Zhang. Reconstructing Cross-Cut Shredded Text Documents: A Genetic Algorithm with Splicing-Driven Reproduction. In *Conf. on Genetic and Evolutionary Comput. (GECCO)*, pages 847–853, 2015.

[11] Y. Gong, Y. Ge, J. Li, J. Zhang, and W. Ip. A splicing-driven memetic algorithm for reconstructing cross-cut shredded text documents. *Applied Soft Comput.*, 45:163–172, 2016.

[12] S. Guo, S. Lao, J. Guo, and H. Xiang. A Semi-automatic Solution Archive for Cross-Cut Shredded Text Documents Reconstruction. In *Int. Conf. on Image and Graphics*, pages 447–461, 2015.

[13] F. Iandola, S. Han, M. Moskewicz, K. Ashraf, W. Dally, and K. Keutzer. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size. *arXiv preprint arXiv:1602.07360*, 2016.

[14] E. Justino, L. Oliveira, and C. Freitas. Reconstructing shredded documents through feature matching. *Forensic Sci. Int.*, 160(2):140–147, 2006.

[15] D. Lewis, G. Agam, S. Argamon, O. Frieder, D. Grossman, and J. Heard. Building a Test Collection for Complex Document Information Processing. In *Conf. on Res. and Develop. in Inf. Retrieval*, pages 665–666, 2006.

[16] H. Lin and W. Fan-Chiang. Reconstruction of shredded document based on image feature matching. *Expert Syst. with Appl.*, 39(3):3324–3332, 2012.

[17] M. Marques and C. Freitas. Document Decipherment-restoration: Strip-shredded Document Reconstruction based on Color. *IEEE Lat. Am. Trans.*, 11(6):1359–1365, 2013.

[18] W. Morandell. Evaluation and reconstruction of strip-shredded text documents. Master's thesis, Inst. of Comp. Graphics and Algorithms, Vienna Univ. of Technol., 2008.

[19] T. Nartker, S. Rice, and S. Lumos. Software tools and test data for research and testing of page-reading OCR systems. In *Electron. Imag.*, 2005.

[20] H. Oh Song, Y. Xiang, S. Jegelka, and S. Savarese. Deep metric learning via lifted structured feature embedding. In *Conf. on Comput. Vision and Pattern Recognit. (CVPR)*, pages 4004–4012, 2016.

[21] 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*, pages 87–94, 2018.

[22] 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.

[23] T. Phienthrakul, T. Santitewagun, and N. Hnoohom. A Linear Scoring Algorithm for Shredded Paper Reconstruction. In *Int. Conf. on Signal-Image Technol. & Internet-Based Syst. (SITIS)*, pages 623–627, 2015.

[24] M. Prandtstetter and G. Raidl. Combining forces to reconstruct strip shredded text documents. In *Int. Workshop on Hybrid Metaheuristics*, 2008.

[25] Matthias Prandtstetter and G. Raidl. Meta-heuristics for reconstructing cross cut shredded text documents. In *Conf. on Genetic and Evolutionary Comput.*, pages 349–356, 2009.

[26] J. Sauvola and M. Pietikäinen. Adaptive document image binarization. *Pattern Recognit.*, 33(2):225–236, 2000.

[27] C. Schauer, M. Prandtstetter, and G. Raidl. A memetic algorithm for reconstructing cross-cut shredded text documents. In *Int. Workshop on Hybrid Metaheuristics*, 2010.

[28] F. Schroff, D. Kalenichenko, and J. Philbin. FaceNet: A Unified Embedding for Face Recognition and Clustering. In *Conf. on Comput. Vision and Pattern Recognit. (CVPR)*, pages 815–823, 2015.

[29] A. Skeoch. An investigation into automated shredded document reconstruction using heuristic search algorithms. *Unpublished Ph. D. Thesis in the University of Bath, UK*, page 107, 2006.

[30] A. Sleit, Y. Massad, and M. Musaddaq. An alternative clustering approach for reconstructing cross cut shredded text documents. *Telecommun. Syst.*, 52(3):1491–1501, 2013.

[31] M. Tschannen, J. Djolonga, P. Rubenstein, S. Gelly, and M. Lucic. On mutual information maximization for representation learning. *arXiv preprint arXiv:1907.13625*, 2019.

[32] A. Ukovich, G. Ramponi, H. Doulaverakis, Y. Kompatsiaris, and M. Strintzis. Shredded document reconstruction using MPEG-7 standard descriptors. In *Symp. on Signal Process. and Inf. Technol.*, pages 334–337, 2004.

[33] K. Weinberger and L. Saul. Distance metric learning for large margin nearest neighbor classification. *Journal of Mach. Learning Res.*, 10(Feb):207–244, 2009.

[34] Nan Xing, Siqi Shi, and Yuhua Xing. Shreds Assembly Based on Character Stroke Feature. *Procedia Comput. Sci.*, 116:151–157, 2017.

[35] H. Xu, J. Zheng, Z. Zhuang, and S. Fan. A solution to reconstruct cross-cut shredded text documents based on character recognition and genetic algorithm. *Abstract and Applied Anal.*, 2014.