THIAGO MEIRELES PAIXÃO

DEEP LEARNING-BASED RECONSTRUCTION OF SHREDDED DOCUMENTS

VITÓRIA, ES

APRIL 2022

THIAGO MEIRELES PAIXÃO

DEEP LEARNING-BASED RECONSTRUCTION OF SHREDDED DOCUMENTS

A thesis submitted in partial fulfillment of of the requirements for the degree of Doctor of Philosophy in Computer Science.

Postgraduate Program in Computer Science – PPGI

Federal University of Espírito Santo – UFES

Supervisor: Prof. Dr. Thiago Oliveira dos Santos

Co-supervisor: Prof. Dr. Maria Claudia Silva Boeres

VITÓRIA, ES

APRIL 2022

Ficha catalográfica disponibilizada pelo Sistema Integrado de Bibliotecas - SIBI/UFES e elaborada pelo autor

---

Paixão, Thiago Meireles, 2918-
P142d Deep learning-based reconstruction of shredded documents / Thiago Meireles Paixão. - 2022.
74 f. : il.

Orientador: Thiago Oliveira dos Santos.
Coorientadora: Maira Claudia Silva Boeres.
Tese (Doutorado em Informática) - Universidade Federal do Espírito Santo, Centro Tecnológico.

1. Inteligência artificial. 2. Redes neurais (Computação). 3. Processamento de imagens. 4. Reconstrução de imagens. I. dos Santos, Thiago Oliveira. II. Boeres, Maira Claudia Silva. III. Universidade Federal do Espírito Santo. Centro Tecnológico. IV. Título.

CDU: 004

Deep Learning-based Reconstruction of Shredded Documents

Thiago Meireles Paixão

Tese submetida ao Programa de Pós-Graduação em Informática da Universidade Federal do Espírito Santo como requisito parcial para a obtenção do grau de Doutor em Ciência da Computação.

Aprovada em 10 de maio de 2022.

Prof. Dr. THIAGO OLIVEIRA DOS SANTOS
Orientador, participação remota

Profª. Drª. MARIA CLAUDIA SILVA BOERES
Coorientadora, participação remota

, Profª. Drª. LUCIA CATABRIGA
Membro Interno, participação remota

Prof. Dr. FRANCISCO DE ASSIS BOLDT
Membro Externo, participação remota

Prof. Dr. ALCEU DE SOUZA BRITTO JR.
Membro Externo, participação remota

Prof. Dr. ROBERTO HIRATA JR.
Membro Externo, participação remota

UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO
Vitória/ES, 10 de maio de 2022

There are far, far better things ahead than any we leave behind.

— C. S. Lewis

Affectionately dedicated to my loves Fernanda and Liz, and
to the loving memory of Marta (Martinha, mãe, *mainha*).

1959 – 2015

Break a vase, and the love that reassembles the fragments is stronger than that love which took its symmetry for granted when it was whole. The glue that fits the pieces is the sealing of its original shape.

— Derek Walcott

ACKNOWLEDGMENTS

Soli Deo gloria. What can I say? I own Him everything I am, and everything I have: friends, family, and an amazing work team of LCAD[1]. Especially, I thank God for my wife Fernanda and my daughter Liz, and for all the "PX team": my father Marcos and Elda, and my brothers Daniel and Lucas.

I am very thankful to my church (Igreja Evangélica Vida) represented by Pr. Valteci Moreira: "A friend loves at all times, and a brother is born for a time of adversity." (Proverbs 17:17).

Many thanks to my supervisor Prof. Dr. Thiago Oliveira dos Santos and co-supervisor Prof. Dr. Maria Claudia Silva Boeres for the confidence in my work. I also thank my colleagues from IFES[2] whose support enabled me to be exclusively focused on my research.

A huge thanks to Prof. Dr. Alessandro Koerich (ÉTS[3]) for having me in Montreal (Canada) for an internship period. I am grateful to all friends I made there: Jonathan de Matos, Steve Ataky, Sajjad Abdoli, and to all LIVIA[4] members. In particular, I was very happy in knowing Wellington and his beautiful family who assisted me and my family during all my stay in Canada.

Finally, I would like to thank Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001 for the scholarship, NVIDIA for providing me a GPU to be used in this research, and finally to AWS Cloud Credits for Research program that provided me cloud computing resources.

1 Laboratório de Computação de Alto Desempenho.
2 Instituto Federal do Espírito Santo.
3 École de Technologie Supériéure.
4 The Laboratory of Imaging, Vision and Artificial Intelligence.

Therefore, all we have are fragments,
gathered in several bundles,
whose final form we can only imagine.

— Os Guiness, on *Pensées* of Blaise Pascal

RESUMO

A reconstrução de documentos fragmentados é uma tarefa importante em diversas situações, tais como na investigação forense e na reconstrução de fatos históricos. Como alternativa ao processo manual, pesquisadores têm desenvolvido métodos para reconstruir documentos (semi-)automaticamente no domínio digital. Apesar dos diversos trabalhos na área, tratar adequadamente dados reais obtidos com uso de máquinas fragmentadoras é um problema crítico na literatura. Neste contexto, duas direções de pesquisa foram abordadas nesta tese: a avaliação robusta de compatibilidade entre os fragmentos, que é o foco do nosso trabalho, e a interação homem-máquina no processo de reconstrução.

Com respeito à avaliação de compatibilidade, verificou-se que as técnicas tradicionais baseadas em análise de *pixel* não são robustas à fragmentação real, enquanto técnicas mais sofisticadas comprometem significativamente a eficiência (tempo de processamento). Esta tese propõe duas abordagens baseadas em *deep learning* para cenários mais complexos/realísticos envolvendo, além da fragmentação mecânica, a mistura de fragmentos provenientes de diversas páginas de documentos (*multi-page reconstruction* ou *multi-reconstruction*). A primeira abordagem modela a avaliação de compatibilidade como um problema de reconhecimento de padrões envolvendo duas classes (válida e inválida). A segunda abordagem, baseada no paradigma *deep metric learning*, propõe separar as etapas de extração de características e avaliação de compatibilidade para melhor eficiência na reconstrução de maiores instâncias de reconstrução.

A interação humana é explorada num segundo momento para se obter maior acurácia comparada aos métodos automáticos. Em relação a este tema, um fator crítico é que os métodos propostos na literatura não escalam eficientemente com o aumento do número de fragmentos (cenário mais realístico). Isso se deve ao fato do usuário ser totalmente responsável pela organização dos fragmentos, e/ou porque ele precisa visualizar todo o documento reconstruído para designar fragmentos a serem analisados. Diante deste desafio, propusemos um *framework* que explora a interação homem-máquina e que automaticamente seleciona potenciais erros na solução (pareamentos incorretos) para serem analisados pelo usuário.

Palavras-chave: Deep Learning; Metric Learning; Reconstrução de Documentos Fragmentados; Avaliação de Compatibilidade; Problema de Quebra Cabeça; Otimização Combinatorial.

ABSTRACT

The reconstruction of shredded documents is a relevant task in various domains, such as forensic investigation and history reconstruction. As an alternative for the manual reconstruction, researchers have been investigating ways to perform (semi-)automatic digital reconstruction. Despite the several works on this topic, dealing with real-shredded data is a very sensitive issue in the current literature. Two research directions are addressed in this thesis to face this scenario: properly evaluating the fitting of shreds (the bulk of this work) and integrating the human into the reconstruction process.

Regarding the fitting (compatibility) evaluation, it was verified that traditional pixel-based approaches are not robust to real shredding, while more sophisticated techniques compromise significantly time performance. This thesis presents two deep learning self-supervised approaches that have achieved state-of-the-art accuracy in more realistic/complex scenarios involving several real-shredded documents where the shreds are mixed (multi-page reconstruction or multi-reconstruction). The first approach models the compatibility evaluation as a two-class (valid or invalid) pattern recognition problem. The second approach, based on deep metric learning, proposes decoupling feature extraction from compatibility evaluation to improve scalability (time performance) for large reconstruction instances.

Human interaction is explored to improve the accuracy of automatic methods. A critical issue regarding this topic is that the proposed methods do not scale well for large instances (real scenario), either because the user has the entire responsibility of arranging the shreds, or because he/she has to visualize the reconstruction and designate the shreds to be analyzed. In face of this challenge, we propose a human-in-the-loop framework that automatically selects potential mistakes (wrong pairings) in the solution for user analysis.

Keywords: Deep Learning; Metric Learning; Reconstruction of Shredded Documents; Compatibility Evaluation; Jigsaw Puzzle Solving; Combinatorial Optimization.

CONTENTS

1 INTRODUCTION 1
1.1 Motivation and Scope . . . 2
1.2 Contributions and Publications . . . 5
1.2.1 Other Publications . . . 6
1.3 Outline . . . 6
2 THEORETICAL BACKGROUND 7
2.1 Document Reconstruction: A Jigsaw Puzzle Problem . . . 7
2.2 Reconstruction of Strip-shredded Text Documents . . . 9
2.2.1 Problem Definition . . . 10
2.2.2 Compatibility Evaluation . . . 12
2.2.3 An Optimal Reconstruction Formulation . . . 14
2.2.4 Semi-automatic Reconstruction . . . 15
2.3 Concluding Remarks . . . 16
3 DEEP RECONSTRUCTION: A CLASSIFICATION-BASED APPROACH 17
3.1 The Reconstruction Method . . . 17
3.1.1 Learning from Simulated-shredded Documents . . . 18
3.1.2 Reconstruction of Mixed Shredded Documents . . . 21
3.2 Experimental Assessment . . . 23
3.2.1 Training Datasets . . . 23
3.2.2 Evaluation Datasets . . . 24
3.2.3 Experiments . . . 25
3.2.4 Experimental Platform . . . 27
3.3 Results and Discussion . . . 27
3.3.1 Experiment 1: Default Configuration . . . 27
3.3.2 Experiment 2: Ablation Study . . . 30
3.3.3 Experiment 3: Comparative Evaluation . . . 31
3.4 Concluding Remarks . . . 33
4 DEEP RECONSTRUCTION: AN ASYMMETRIC METRIC-LEARNING APPROACH 35
4.1 The Reconstruction Method . . . 35
4.1.1 Learning Projection Models . . . 37
4.1.2 Pairwise Compatibility Evaluation . . . 38
4.2 Experimental Assessment . . . 39
4.2.1 Implementation Details . . . 40
4.2.2 Experiments . . . 41
4.2.3 Experimental Platform . . . 42
4.3 Results and Discussion . . . 42
4.3.1 Experiment 1: Single-page Reconstruction . . . 42

4.3.2 Experiment 2: Multi-page Reconstruction . . . 44
4.3.3 Experiment 3: Sensitivity Analysis . . . 45
4.4 Concluding Remarks . . . 46
5 A HUMAN-IN-THE-LOOP RECONSTRUCTION FRAMEWORK 47
5.1 HIL Framework . . . 47
5.1.1 Optimality-based Strategies . . . 49
5.1.2 Uncertainty-based Strategies . . . 49
5.2 Experimental Assessment . . . 50
5.2.1 Experiments . . . 51
5.2.2 Implementation Details. . . . 52
5.2.3 Experimental Platform . . . 52
5.3 Results and Discussion . . . 52
5.3.1 Experiment 1: Workload Experiment . . . 53
5.3.2 Experiment 2: Multi-iteration Experiment . . . 54
5.4 Concluding Remarks . . . 55
6 CONCLUDING REMARKS AND FUTURE WORK 56

I APPENDIX 58
A APPENDIX: AN ASSYMETRIC METRIC-LEARNING APPROACH 59
A.1 Local Samples Nearest Neighbors . . . 59
A.2 Reconstruction of S-ISRI-OCR . . . 60
A.3 Embedding Space . . . 61
A.3.1 Case 1 . . . 61
A.3.2 Case 2 . . . 62
A.3.3 Case 3 . . . 62
A.3.4 Case 4 . . . 63
A.4 Sensitivity analysis w.r.t. sample size . . . 63
A.5 Statistical test . . . 64
B APPENDIX: OTHER PUBLICATIONS 65

BIBLIOGRAPHY 67

LIST OF FIGURES

Figure 1 Manual reconstruction (1979 Iran Hostage Crisis). 1
Figure 2 Classical approach for automatic document reconstruction. 2
Figure 3 Shredding type. 3
Figure 4 Document sample of our local dataset. 3
Figure 5 Jigsaw solving-related problems. 8
Figure 6 Tiles panel reconstruction. 8
Figure 7 Document shreds spliced onto a yellow background. 9
Figure 8 DARPA Shredder Challenge instances. 10
Figure 9 Overview of the proposed system. 18
Figure 10 Simulated shredding. 19
Figure 11 Association image extraction. 21
Figure 12 Compatibility computation as a sliding window operation. 22
Figure 13 Samples of the shredded datasets. 24
Figure 14 Multi-reconstruction accuracy across datasets. 27
Figure 15 Reconstruction of a test instance. 28
Figure 16 Reconstruction accuracy for S-Cdip across categories. 28
Figure 17 Challenging reconstruction instances. 29
Figure 18 Accuracy sensitivity w.r.t. the parameter $\rho_{\texttt{black}}$. 30
Figure 19 Accuracy sensitivity w.r.t. the size of training samples. 30
Figure 20 Visual ambiguity between “wo” and “vo”. 31
Figure 21 Accuracy sensitivity w.r.t. the parameter $\nu_{\texttt{shift}}$. 31
Figure 22 Comparative accuracy performance. 32
Figure 23 Comparative time performance. 33
Figure 24 Metric learning approach for shreds’ compatibility evaluation. . . . 36
Figure 25 Self-supervised learning of the models. 37
Figure 26 Learning projection models for shreds’ compatibility evaluation. . 38
Figure 27 Compatibility evaluation of a pair of shreds. 38
Figure 28 Accuracy distribution for single-page reconstruction. 43
Figure 29 Time performance for single-page reconstruction. 43
Figure 30 Local samples nearest neighbors. 44
Figure 31 Time performance for multi-page reconstruction. 45
Figure 32 Sensitivity analysis w.r.t. embeddings dimension. 46
Figure 33 Overview of the proposed HIL reconstruction framework. 48
Figure 34 Reconstruction accuracy w.r.t. workload (Deeprec-CL). 53
Figure 35 Reconstruction accuracy w.r.t. workload (Deeprec-ML). 53

Figure 36 Reconstruction accuracy w.r.t. the number of iterations (DEEPREC-ML, OPT-R). 54
Figure 37 Querying $\mathbf{x}_l$ samples by fixing $\mathbf{x}_r$. 59
Figure 38 Local samples nearest neighbors (appendix). 59
Figure 39 Reconstruction of S-ISRI-OCR. The generated image was split into 4 parts for better visualization. 60
Figure 40 Case 1. 61
Figure 41 Case 2. 62
Figure 42 Case 3. 62
Figure 43 Case 4. 63
Figure 44 Reconstruction accuracy w.r.t. to the sample height (s_y). 64

LIST OF TABLES

Table 1 Single-page reconstruction performance 42
Table 2 Multi-page reconstruction performance. 52
Table 3 Accuracy improvement w.r.t. the query strategies (DEEPREC-ML, $\alpha_{load} = 0.25$). 53
Table 4 Accuracy improvement w.r.t. the number of iterations (DEEPREC-ML, $\alpha_{load} = 0.25$, OPT-R). 54
Table 5 Page-wise paired t-test. 64

ACRONYMS

AL Active Learning

ATSP Asymmetric Traveling Salesman Problem

CNN Convolutional Neural Network

DARPA Defense Advanced Research Projects Agency

DL Deep Learning

FDE Forensic Document Examiner

FCNN Fully Convolutional Neural Network

HIL human-in-the-loop

LCAD High Performance Computing Laboratory

MCHPP Minimum-Cost Hamiltonian Path Problem

OCR Optical Character Recognition

RSSTD Reconstruction of Strip-shredded Text Documents

RCCTD Reconstruction of Cross-cut Text Documents

SGD Stochastic Gradient Descent

SMD Standardized Mean Difference

TSP Travelling Salesman Problem

1 INTRODUCTION

The paper shredder machine was invented in the early 1900s [51] to physically destroy waste paper making its content unintelligible. This kind of device is still commonly seen today in different organizational environments, where huge amounts of documents must be constantly disposed preventing the disclosure of sensitive information. Besides, paper shredders have become popular for personal use due to the decrease in prices, which allows people to manage personal files more safely.

Historically, shredding is also associated with the destruction of espionage content, as in the Iran hostage crisis [25][1] or in the case of the documents left behind by the official state security service of former East Germany (*Stasi*) after the fall of the Berlin Wall [50]. Additionally, shredding may be illicitly motivated when the objective is to destroy evidence of fraud and other sorts of crimes. In this context, revealing the original content of shredded papers is of great relevance for forensic investigation, which can be achieved by first joining coherently the paper shreds as in a jigsaw puzzle. For instance, the manual reconstruction of shredded documents (*e. g.*, bills, letters, messages) (Figure 1) conducted by free-lance reporters helped to reveal an influence-peddling scheme involving South Korean figures and U.S. congressmen, which became known as the *Koreagate* scandal [8, 56]. Manual reconstruction is also portrayed in the Netflix production *The Mechanism*[2], where criminal evidence was exposed after a police chief has manually spliced dozens of paper shreds.

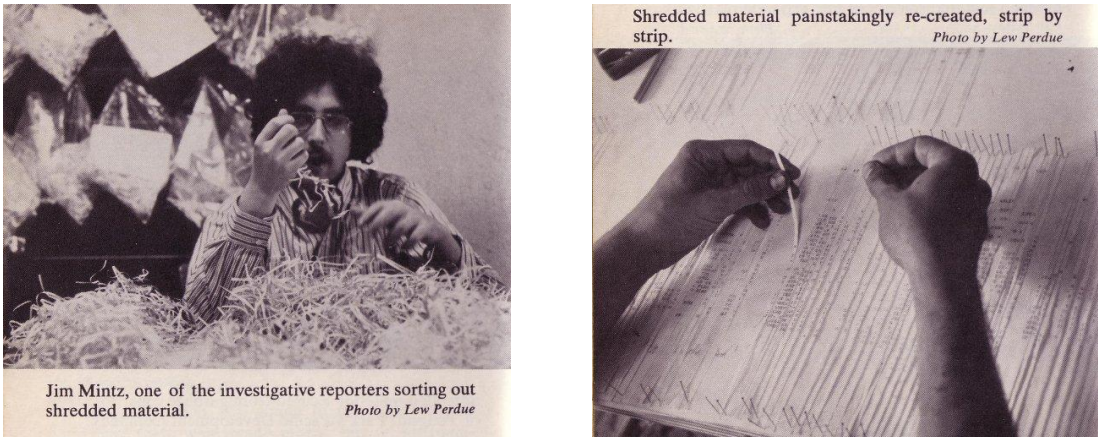


Figure 1: Manual reconstruction of the shredded material during the 1979 Iran hostage crisis. **Source:** http://lewisperdue.com/archives/4052.

1 The Iran hostage crisis – including the reassembling of the shredded documents – is portrayed in the movie "Argo".

2 A fictional series inspired by the *Carwash* operation, the largest anti-corruption operation ongoing in Brazil. Official trailer (accessed on 2020-04-07): https://www.youtube.com/watch?v=130tvUxOcUU.

Regardless of its importance, the manual reconstruction is potentially damaging to the paper due to the continuous direct contact with the documents' shreds, besides being a slow and tedious process for humans. Those facts motivated the development of the digital and automatic reconstruction process [96]. There is commercial software, such as *Unshredder*[3], that could assist people and corporations to recover destroyed (whether intentionally or not) paper documents, or that could be leveraged by Forensic Document Examiners (FDEs) in criminal investigations. On the other hand, such technology enables the use of disposed or robbed documents (*e. g.*, industrial espionage) for invasion of privacy and illicit use of sensitive data. Therefore, as with hacker intrusion, the reconstruction technology can be used to assess the safety level of shredding and disposal services provided by specialized companies, such as ShredIt[4].

This thesis aims to advance the state-of-the-art on reconstruction of shredded documents in two main directions. First, it proposes two approaches for automatic reconstruction that leverages the Deep Learning (DL) revolution for high accuracy performance in realistic scenarios. Second, it proposes a reconstruction framework that benefits from human interaction to overcome the accuracy limitation of fully automatic methods.

1.1 MOTIVATION AND SCOPE

Digital reconstruction of shredded documents has emerged in the past decade mainly motivated by historical and forensics needs [15, 37, 96]. The laborious manual effort is alleviated by algorithms capable of assessing the fitting (compatibility) of the shreds and grouping them by optimizing the overall compatibility. Therefore, fragments are manipulated only during the preliminary acquisition procedure, and the human participation is restricted to specific interventions (semi-automatic reconstruction [15, 72, 103]), or even not required at all (automatic reconstruction).

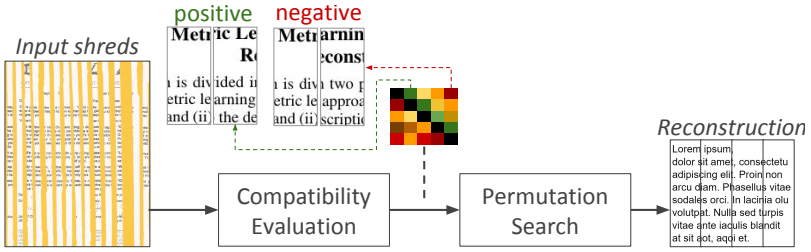


Figure 2: Classical approach for automatic document reconstruction. Shreds' compatibility is evaluated pairwise and then an optimization search process is conducted (based on the compatibility values) in order to find the shreds' permutation that best represents the original document [65].

Typically, compatibility evaluation and optimization are treated separately, as depicted in Figure 2. Evaluating compatibility is an image-based problem that aims to quantify

3 https://www.unshredder.com.
4 https://www.shredit.com.

how fitting two shreds are, *i.e.*, the probability of two shreds being adjacent (in the correct order). The optimization process is modeled as a combinatorial problem whose goal is to find a coherent arrangement of the shreds guided by the compatibility values. In the example, the document was cut only in the longitudinal direction (strip-shredding), therefore the reconstruction is a permutation of shreds. A harder problem, however, is the reconstruction of cross-cut documents, *i.e.*, documents shredded both in the longitudinal and transverse directions (Figure 3).

(a) (b)

Figure 3: Shredding type. (a) Strip-shredding; (b) Cross-cut shredding.

The literature on reconstruction of shredded documents has mostly focused on improving the optimization search process, relegating the image-based problem of verifying shreds' compatibility to the background. Most works are restricted to the reconstruction of simulated-shredded documents [30, 45, 57, 70, 76, 84, 91], and, in this situation, the criticalness of the compatibility evaluation step in the reconstruction pipeline is masked, giving rise to potentially misleading conclusions. For those addressing real-shredded data (*e.g.*, [47, 48]), the main issue is the assessment with few test instances ($\leqslant$ 3 documents) that also yields biased conclusions. Literature also lacks studies on the impact of mixing shreds from different documents on reconstruction accuracy.

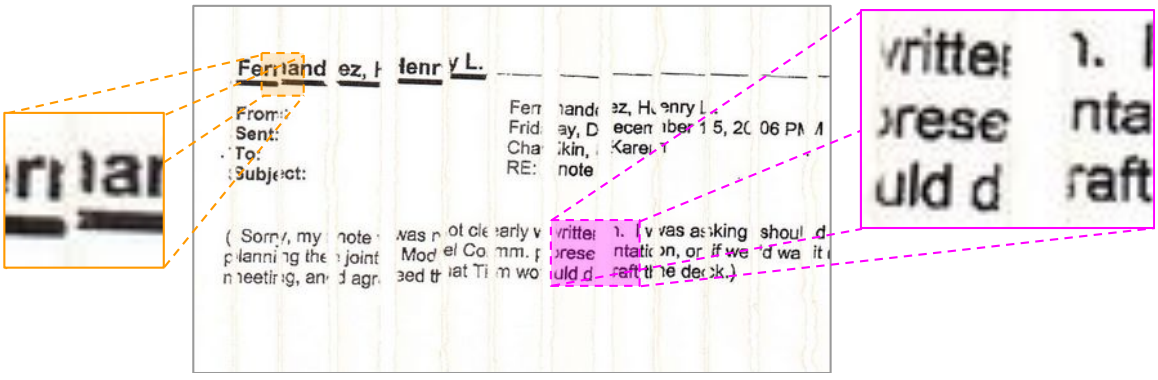


Figure 4: Document sample of our local dataset. Note that shreds can be curved, and how the borders are trimmed.

The aforementioned limitations have motivated this work, mainly the problem of evaluation shreds' compatibility, which we consider the major research gap in the reconstruction literature. To provide a robust solution for real-shredded data (see Figure 4), we leveraged DL, more specifically with the use of Convolutional Neural Networks (CNNs)

[28, 42, 43], an architecture type that have been enabling state-of-the-art performance for several image-based problems [4, 31, 40, 49, 78]. The experimental assessment considered several realistic scenarios, including multi-page reconstruction.

As an additional contribution, this thesis also investigated the benefit of user inputs to overcome the accuracy limitation of fully automatic methods. Unlike most of the hybrid methods [15, 22, 32, 87], where the user is inherently part of the reconstruction process, we propose a framework where the user interaction is optional and provides feedback on the obtained solutions. This is closely related to the work of Prandtstetter and Raidl [76], where the user role is to confirm whether pairs of adjacent shreds in the solution are also adjacent in the original document. In comparison to [76], our approach relieves the user from the burden of identifying relevant pairs for annotation, being this task relegated to a *recommender module* designed to identify potential mistakes (wrong pairs). This makes user participation more feasible for reconstruction instances with several shreds.

The scope of our work assumes:

- Strip-shredded documents: the cuts are restricted to the vertical direction;
- Correctly-oriented shreds: the shreds are set upwards (possibly slightly skewed);
- Single-sided shreds: the documents' content is only in one of the paper faces;
- Black-and-white appearance: the target of this thesis is text documents, which, in general, are low-color images (*e.g.*, forms, legal documents, and business letters.
- Shreds with nearly the same dimensions: the popular paper shreds produce nearly the same number of shreds, therefore it is safe to assume that the shreds have nearly the same height and width.

Although there are works addressing more complex scenarios from the optimization perspective (*e.g.*, cross-cut documents), the current solutions are not robust for real shredded data, which is a more complex scenario from the image analysis perspective. Therefore, we believe that the first step for solving cross-cutting is a robust approach for strip-shredding (*i.e.*, vertical cuts). Our focus is precisely on the compatibility evaluation between shreds, an essential step to solve the reconstruction problem for which little progress has been observed in the literature.

From the application perspective, it is important that the reconstruction solutions not only manage real-shredded data, but that they can generalize for different shredder machines, which implies dealing with shreds in different damage levels. Despite the outstanding results with our DL approaches, this investigation was not feasible due to the lack of public datasets and equipment limitation (only one shredded was available to assemble our collection).

1.2 CONTRIBUTIONS AND PUBLICATIONS

The development of this thesis has contributed to the literature in significant ways. In terms of methods for compatibility evaluation, there are a few consolidated contributions reported in four articles (enumerated according to the chronological development):

1. T. M. Paixão, M. C. S. Boeres, C. O. A. Freitas, and T. Oliveira-Santos. "Exploring Character Shapes for Unsupervised Reconstruction of Strip-shredded Text Documents." In: *IEEE Trans. Inf. Forensics Secur.* 14.7 (2019), pp. 1744–1754. ISSN: 1556-6013 — An unsupervised approach based on character shapes published as a journal article [61] (qualis A2);

2. T. M. Paixão, R. F. Berriel, M. C. S. Boeres, C. Badue, A. F. De Souza, and T. Oliveira-Santos. "A deep learning-based compatibility score for reconstruction of strip-shredded text documents." In: *Conf. on Graph., Patterns and Images*. 2018, pp. 87–94 — A preliminary investigation on reconstruction based on DL published as a conference article [62] (qualis B1)

3. T. M. Paixão, R. F. Berriel, M. C. S. Boeres, A. L. Koerich, C. Badue, A. F. De Souza, and T. Oliveira-Santos. "Self-supervised deep reconstruction of mixed strip-shredded text documents." In: *Pattern Recognit.* 107 (2020), p. 107535 — Extension of [62] addressing the reconstruction of several mixed shredded documents published as a journal article [63] (qualis A1);

4. T. M. Paixão, R. F. Berriel, M. C. S. Boeres, A. L. Koerich, C. Badue, A. F. D. Souza, and T. Oliveira-Santos. "Fast(er) Reconstruction of Shredded Text Documents via Self-Supervised Deep Asymmetric Metric Learning." In: *IEEE/CVF Conf. on Comp. Vision and Pattern Recognit.* 2020, pp. 14343–14351 — A deep metric learning approach published as an international conference article [65] (qualis A1).

Although this thesis focuses on the DL approaches (items 2 and 3), the first work [61], which relies on traditional computer vision techniques, brought relevant findings and methodology that enabled further progress. First, it showed, through extensive experimentation, that the state-of-the-art methods fail when dealing with real mechanically-shredded documents. Also, it settled the ground for the experimental methodology used in the more recent works, which includes metrics, a new dataset (with 20 shredded documents), and the exact formulation of the optimization problem for comparative purposes.

Concerning the DL methods, Chapter 3 covers in detail the classification-based approach, including other relevant contributions, such as a comprehensive investigation on the reconstruction of mixed shredded documents (*i. e.*, multi-page reconstruction) and the release of a new dataset with 100 real strip-shredded documents (totaling 2,292 shreds). In this context, multi-page reconstruction denotes the reassembly of individual pages, which means that the reconstruction is not concerned in grouping the pages of a same

document. Together with the 20 instances released in [61], we have made available the largest collection of shredded documents, totaling 120 instances with ground-order annotation. Chapter 4 presents the deep metric-learning approach, an alternative way to use neural networks that significantly improves the time scalability when processing larger instances.

The contributions related to the interactive (human-in-the-loop) reconstruction framework were compiled into an article currently under review for a journal.

1. T. M. Paixão, R. F. Berriel, M. C. S. Boeres, A. L. Koerich, C. Badue, A. F. De Souza, and T. Oliveira-Santos. "A human-in-the-loop recommendation-based framework for reconstruction of mechanically shredded documents." In: *Pattern Recognit. Letters* (under review) (qualis A1).

Chapter 5 introduces the reconstruction framework and four different strategies the recommender module employs to query the user for annotation. The chapter also covers a novel methodology to assess the human impact on the quality improvement of the solutions.

1.2.1 *Other Publications*

The development of this thesis has resulted in several publications (listed in Appendix B) in the field of machine learning with members of the High Performance Computing Laboratory (LCAD). These collaborations contributed in many ways to develop skills in the machine learning theory, as well as in scientific methodology (*i. e.*, experimental design, results analysis, critical literature review, etc.). These skills were of great value for the construction of this thesis.

1.3 OUTLINE

The remainder to the text is organized as follows:

- Chapter 2 describes the theoretical background and related work;
- Chapter 3 introduces the classification-based approach and discusses relevant results on multi-page reconstruction;
- Chapter 4 describes the deep metric-learning approach and discusses, through experiments, how it improves the time scalability of deep models in the reconstruction application;
- Chapter 5 discusses the interactive reconstruction framework, and how the user feedback can improve the reconstruction accuracy;
- Finally, Chapter 6 presents the concluding remarks and future work.

2

THEORETICAL BACKGROUND

This chapter presents the theoretical aspects concerning the problem of reconstructing shredded documents. The purpose is to situate the reader with relevant literature on the topic, and formally introduce the optimization model for the reconstruction problem.

The text is divided into three main sections. The first section introduces document reconstruction as a jigsaw puzzle problem. Then, the discussion moves on to the automatic reconstruction of strip-shredded documents, addressing the reconstruction problem definition, the compatibility evaluation sub-problem, and an optimal reconstruction formulation (solver). In our approach, the compatibility evaluation is decoupled from the solver, thus different strategies to compute compatibilities can be compared under a common ground. Finally, we review the literature on semi-automatic reconstruction, paving the way to discuss the human in the reconstruction loop (Chapter 5).

2.1 DOCUMENT RECONSTRUCTION: A JIGSAW PUZZLE PROBLEM

An interesting analogy for the document reconstruction problem is the jigsaw puzzle solving. In both, the goal is to fit the pieces to compose the full image following the clues provided by the appearance and shape of the pieces. In fact, the literature relates the computational problem of reconstructing documents to the classical problem of solving jigsaw puzzles automatically [15, 37, 76].

Originally, "apictorial" jigsaw puzzles were addressed, thus the solution was based solely on the uniqueness of the pieces' shape fitness [27]. Three decades later, the pictorial information started to be explored under the assumption that neighbor pieces tend to be similar in appearance (color and texture), more specifically in the areas around the contact edges [39]. The combined use of shape and appearance enabled the solution of several real problems, such as the reconstruction of ancient artifacts (Figure 5a) and hand-torn documents (Figure 5b).

In a more recent version of the problem, Nielsen, Drewsen, and Hansen [59] addressed the assembling of jigsaw puzzles with equal-size rectangular pieces. They argue that this puzzle can be solved by leveraging only appearance features (*e. g.*, color and texture). Such features can be also useful even when the pieces are not exactly regular, but their shapes present some ambiguity. In this domain, there is the interesting application of tiles panels reconstruction (Figure 6), which is of great historical and artistic value.

Despite sharing some similarity with this more recent version of the jigsaw puzzle problem, the reconstruction of documents start to demand more customized approaches to deal with the edge of the shreds, damaged by the machine during the shredding pro-

(a) (b)

Figure 5: Jigsaw solving-related problems. (a) Reconstruction of pottery from the Apollonia-Arsuf archeological site in Israel [99]; (b) Reconstruction of a receipt document [37].

Figure 6: Reconstruction of an ancient Portuguese tiles panel: potential application for the jigsaw puzzle solving with equal-size pieces [2].

cess (Figure 7), and to deal with the shape of the shreds, since the modern paper shredders – despite the rectangular cut pattern – usually produce irregular (*i. e.*, wrinkled, torn, curved) fragments. In this regard, it is worthy to mention the DARPA Shredder Challenge [23], a competition promoted in 2011 by the DARPA, a research agency of the U.S. Department of Defense. The challenge included five problems, each one involving the reconstruction of a cross-cut document. The fragments have edges remarkably irregular and their content is very diverse, comprising, for instance, manuscript text, hand-made drawings, color, and colorful pictures Figure 8).

Historically, the challenge was a milestone that fostered research on the particular application of shredded documents, including the fact the organizers released a public dataset. Butler, Chakraborty, and Ramakrishan [15], for instance, used the *Puzzle 1* to validate their interactive document reconstruction system (*Deshredder*). The proposed system explores

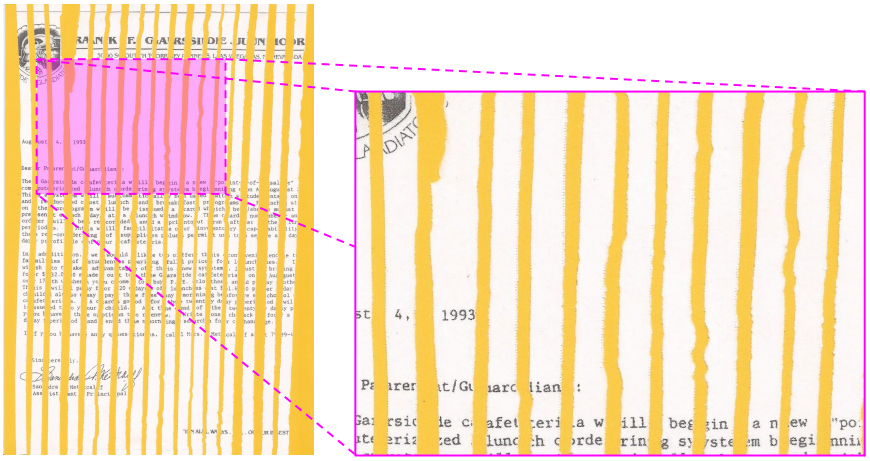


Figure 7: Document shreds spliced onto a yellow background.

color and the very irregular fragments' shapes in feature representation. The same puzzle was used by Zhang, Lai, and Bächer [103] in a similar context of assisted reconstruction. Their tool (*Hallucination*) requires the user to sketch scribbles that serve as a clue to determine neighbor fragments. According to the authors, evaluating the compatibility between the fragments is the core challenge in the reconstruction problem.

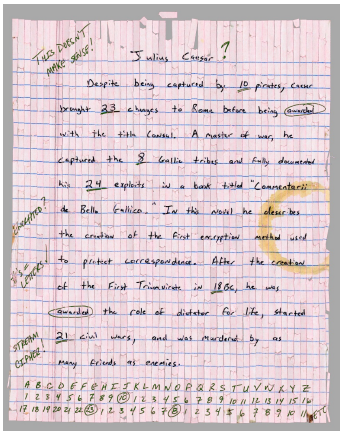
The dataset assembled for the challenge resembles that produced by Saboia and Goldenstein [81]. However, text documents (the target of our research) usually depict low color and texture information, being the primary content typewritten text (usually black) onto a paper substrate (usually white). Additionally, the use of modern shredders (as assumed in this project) yields shreds with edges significantly less irregular. This context coupled with the privacy requirements of documents used in real investigation forces the researchers to produce their own collections by using paper shredders (real shredding) or by simulating cuts in documents (artificial/virtual shredding).

2.2 RECONSTRUCTION OF STRIP-SHREDDED TEXT DOCUMENTS

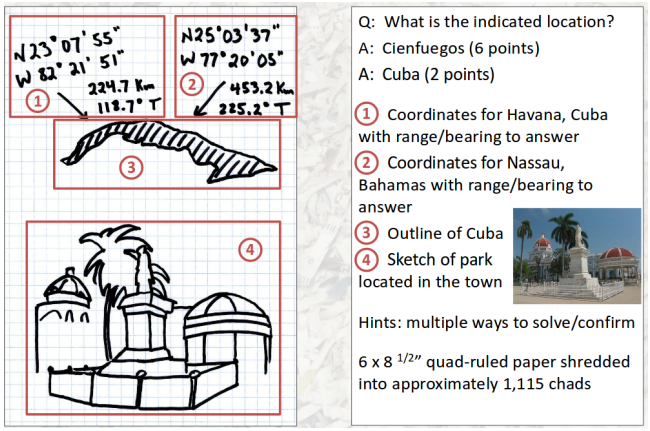
In the scope of mechanically-shredded text documents, the literature poses two main variations of the reconstruction problem [76]: (i) Reconstruction of Strip-shredded Text Documents (RSSTD) and (ii) Reconstruction of Cross-cut Text Documents (RCCTD). The RSSTD can be viewed as a particular case of RCCTD where the puzzle has one only row of shreds. The RCCTD is, therefore, a more complex problem mainly from the optimization perspective. Since our focus is on the compatibility evaluation of the shreds rather than on the optimization process, the scope of the discussion here is on the strip-shredding version of the problem. Nonetheless, it is important to notice that part of the related works also addresses cross-cut documents and that the findings of our thesis can help to improve the reconstruction of cross-cut documents.

The discussion in this section starts with a reconstruction optimization model that is based on a pairwise cost (or, alternatively, compatibility) function that measures the fit-

(a) Puzzle 1.



(b) Puzzle 2.



(c) Puzzle 3.

Figure 8: Defense Advanced Research Projects Agency (DARPA) Shredder Challenge instances [23].

ting between shreds. Then, a thorough review of compatibility evaluation strategies is presented, which includes, among other things, the design/choice of features, algorithms, and (dis)similarity measures. Subsequently, we comment on the challenges to establish the state-of-the-art for this problem. The section ends with a presentation of the optimal reconstruction formulation adopted in our research that has enabled us to better compare the different compatibility evaluation methods.

2.2.1 *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 $\mathcal{S} = \{s_i\}_{i=1}^{n}$ 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 $\pi_S = (s_{\pi_i})_{i=1}^{n}$. 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 [57, 76] whose objective function derives from pairwise compatibilities (Figure 2). Compatibility – or cost, depending on the perspective – is given by a function $\phi : S^2 \to \mathbb{R}^+$ that quantifies the (un)fitting of two shreds when placed side-by-side (order matters). Assuming a cost interpretation, $\phi(s_i, s_j)$, $i \neq j$, denotes the cost of placing s_j to the right of s_i. In theory, $\phi(s_i, s_j)$ should be low when $j = i + 1$ (positive pair), and high for other cases (negative pairs). Typically, $\phi(s_i, s_j) \neq \phi(s_j, s_i)$ due to the asymmetric nature of the reconstruction problem. For computational purposes, it is useful to represent such graph as an adjacency cost matrix $\mathbf{\Phi} = (\Phi_{i,j})_{n \times n}$, where $\Phi_{i,j} = \phi(s_i, s_j)$.

The cost values are the inputs for a search procedure that aims to find the optimal permutation π_S^*, which is the arrangement of the shreds that best resembles the original document. The objective function Obj_ϕ to be minimized is the accumulated pairwise cost computed only for consecutive shreds in the solution:

$$\mathrm{Obj}_\phi(\pi_S) = \sum_{i=1}^{n-1} \phi(s_{\pi_i}, s_{\pi_{i+1}}). \tag{1}$$

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 that respects the ground-truth order on each page, as well as the expected order (if any) of 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 $j \neq i + 1$. Based on this definition, Equation (2) quantifies the quality (accuracy) of a solution π_S as the proportion of positive pairs of shreds in a solution [3, 61, 63, 65]:

$$\mathrm{acc} = \frac{1}{n-1} \sum_{i=1}^{n-1} [(s_{\pi_i}, s_{\pi_{i+1}}) \text{ is positive}], \tag{2}$$

where $[\cdot]$ is the Iverson Bracket notation, *i.e.*, $[P] = 1$ if P is True, and $[P] = 0$, otherwise. Note that accuracy ranges in the interval $[0, 1]$, where 0 implies a fully disordered reconstruction, and 1 is achieved only by a perfect reconstruction.

The optimization problem of minimizing Equation (1) has been extensively investigated in literature, mainly using genetic algorithms [12, 29, 30, 102] and other metaheuristics [9, 75, 84]. The focus of this thesis is, nevertheless, on the compatibility evaluation between shreds (*i.e.*, the function ϕ), which is critical to lead the search towards accurate reconstructions.

2.2.2 *Compatibility Evaluation*

The literature on reconstruction of shredded documents has mostly focused on improving the optimization search process, relegating the image-based problem of verifying shreds' compatibility to the background. Most works are restricted to the reconstruction of simulated-shredded documents [30, 57, 70, 76, 91], and, in this situation, the criticalness of the compatibility evaluation step in the reconstruction pipeline is masked, giving rise to potentially misleading conclusions on the accuracy of the methods for real-shredded documents. Therefore, this review focuses on different approaches to assess compatibility between shreds, which is the main contribution of this work.

Inspired by the jigsaw-puzzle solving problem, most literature on reconstruction explores low-level features for compatibility evaluation. The most naive approach in this context is to apply distance metrics (*e. g.*, Hamming, Euclidean, Canberra, Manhattan) on the raw pixels of opposite boundaries of two shreds [16, 18, 24, 53, 90]. Some of these methods rely on the very edge [53, 90], being more sensitive to the corruption of the shreds' extremities caused by the mechanical cut. To alleviate this, Marques and Freitas [53] suggest removing some border pixels, which, in practice, results in limited improvement. Additionally, different color spaces have been investigated (RGB [24], HSV [53, 90], gray-scale [16]) without great success for text documents due to their poor chromatic information.

More sophisticated compatibility measures were designed to solve image puzzles with rectangular tiles, such as prediction-based [73] measure. In the context of document shreds, Andaló, Taubin, and Goldenstein [3] proposed a modified version of the measure proposed by Pomeranz, Shemesh, and Ben-Shahar [73], reaching near 100% of accuracy for simulated shredding with documents from ISRI-Tk OCR dataset [58], a collection of images commonly used to assess Optical Character Recognition (OCR) software. Nevertheless, our previous investigation [61] demonstrated that the accuracy of their method decreases dramatically when dealing with real-shredded documents of the same image collection.

Some authors designed compatibility measures focusing on text documents [11, 57, 77]. Balme [11] and Morandell [57] addressed the problem of vertical misalignment between pixels around the cutting section, *i. e.*, the area near the touching edges of two adjacent shreds. Both of them adopt binary image representation given the black-and-white appearance of the text documents. Balme's measure is used in several works [17, 30, 76], and consists of a weighted pixel correlation intended to mitigate the misalignment issue, whereas Morandell [57] quantifies the degree of misalignment between corresponding text lines ("black" pixels) as a measure of compatibility. In an unsupervised approach, Ranca [77] proposed learning the expected arrangement of pixels around the cutting section using information from the pixels inside the shreds. The best results were achieved with a simple probabilistic model, although they have also evaluated, unsuccessfully, feed-forward neural networks. Their experiments were also limited, given that they were carried out only on simulated-shredded data. Text-line detection was exploited in [48, 72,

91], however, these methods struggle with ambiguities typically found in common text documents.

At a higher level of abstraction, compatibility can be assessed by exploring the matching degree of fragmented characters around the cutting section [61, 69, 70, 100, 101]. The continuity of character strokes was used as a matching criterion in [70, 100]. In such an approach, the reconstruction accuracy strongly relies on the vertical alignment of the shreds and the image quality around the shreds' boundaries.

Alternatively, learning-based matching has also been proposed in the literature [61, 69, 101]. Matching in [69] leverages OCR-based on keypoint features. OCR tends to work well for general text recognition, but its application on corrupted characters is quite unstable, which turns it into a drawback of this formulation. Instead of identifying symbols, Xing and Zhang [101] proposed a learning model to identify valid combinations of symbols (restricted to the Chinese language) based on structural features. The work of our group Paixão et al. [61] (preliminary to our current deep-learning approaches) analyzes the types of symbol combinations based on their shapes. Both [61, 101] depend on segmenting text information from shreds, which is a challenging task given the condition of the shredded documents.

In a recent and relevant work, Liang and Li [47] proposed the *word path metric*, which combines pixel- and character-level information (low-level metrics) with word-level information (high-level metric). A central procedure in their method is sampling candidate sequences and applying OCR for word recognition to improve pair compatibility estimation. Despite reporting accuracy comparable to our deep learning approach (preliminary published in [62]), their validation relies on solely three real-shredded instances. For two of them, those which are up to 39 shreds, their method achieved accuracy above 70%, while the third instance, with 56 shreds, yielded 41.8% of accuracy. As mentioned by the authors, scalability for larger instances (*i. e.*, with more shreds) is still an issue firstly due to the OCR working overhead and its accuracy degradation. Additionally, for better accuracy, the number and size of candidate sequences have to be increased, which compromises the run-time performance (it performed ≈ 6 times slower than [62] for a 60-shreds instance).

In the last years, deep learning started to be used in the context of jigsaw puzzle solving with simulated-cut tiles. Le and Li [41] applied Convolutional Neural Networks (CNNs) to verify potential matching pieces in order to reduce the search space for the posterior optimization process. Paumard, Picard, and Tabia [67] solved small (3×3) 2D-tile puzzles following the seminal ideas introduced in [26, 60], in which CNNs are trained in a self-supervised way to predict the relative positions of patches cropped from a reference image. More related to our work, Sholomon, David, and Netanyahu used a fully connected network to measure pairwise compatibility between 2D-tile. Boundary pixels of two tiles are fed to the network and the network's output, *i. e.*, the predicted adjacency probability, is assigned as the pair compatibility. Although these works are promissory to solve jigsaw-related problems, they only considered a non-realistic scenario with simulated-cut pieces.

To the best of our knowledge, by developing this project, we were the first to explore deep learning in a realistic scenario that includes multi-page real-shredded documents. The use of deep models aims to improve robustness in real shredding context, where the damage to the shreds' borders prevents the use of similarity evaluation at pixel-level or of stroke continuity analysis. Additionally, our approach is able to cope with more heterogeneous content because the fitting of patterns is learned in a self-supervised fashion from large-scale data without segmenting symbols, as will be discussed in the next chapters.

ESTABLISHING THE STATE-OF-THE-ART. The investigation of the literature revealed a challenge in properly establishing the state-of-the-art for the reconstruction of strip-shredded text documents. This is due to the lack of public codes and datasets (with real-shredded documents) to reproduce the performed experiments. In this direction, our first contribution in this research field [61] was relevant not only for the achieved results in comparison with other tested methods, but because it provided the audience with the performance of the main methods in literature in a comparative way, besides the source code and datasets to enable reproducibility of our work[1].

With the publication of [61], the interested audience may have a picture of the main methods in literature (those which were possible to test) and their performance. From that point, we started to improve our reconstruction technique migrating from traditional computer vision to deep learning techniques [62, 63, 65]. We started to be cited in literature with the work of Liang and Li [47], which used the preliminary model developed in [62] for comparative analysis with their method. This enforces our compromise not only in developing and evaluating software but helping to push further the research in this area contributing in significant ways to the literature.

2.2.3 *An Optimal Reconstruction Formulation*

Since the main contribution of our reconstruction technique lies in the compatibility evaluation, it is necessary to define an optimization formulation that will serve as a ground for literature comparison. The adopted formulation was first addressed in our previous work [61]. On that occasion, an optimal reconstruction formulation was presented by reducing the original reconstruction problem to the Travelling Salesman Problem (TSP), whose solution can be obtained by 3-rd party software (*e. g.*, Concorde TSP Solver [5]).

Optimality here should be understood as the ability of the optimization strategy in providing an optimal solution given the previously computed costs, *i. e.*, a minimum-cost solution for the Equation (1). Notice, however, that this does not guarantee that the correct solution/permutation of shreds will be found since the computed costs may not reflect reality. As consequence, the expected permutation (ground-truth solution) may not have the optimal (minimum) cost – from the optimization perspective –, or there may

1 https://github.com/thiagopx/docrec-tifs18.

be multiple optimal solutions and the expected solution is one of those. Despite that, an optimal formulation is still desired because low-accuracy solutions can be directly regarded as a failure in evaluating compatibilities.

Following a similar approach in [76], the original reconstruction problem is first reduced to the problem of finding the Minimum-Cost Hamiltonian Path Problem (MCHPP) in a directed weighted graph $G = (\mathcal{V}, \mathcal{A}, w)$. Here, the vertices in $\mathcal{V}$ are uniquely associated to shreds, and $w(a)$, given an arc $a = (v_i, v_j) \in \mathcal{A}$, carries the cost $\phi(s_i, s_j)$ as weight. If $v_{\pi_1} v_{\pi_2} \dots v_{\pi_n}$ is an optimal solution for the MCHPP on G, then the permutation $\pi_S^* = (s_{\pi_i})_{i=1}^n$ is an optimal solution for the reconstruction problem given the Equation (1). As in [76], MCHPP is reduced to the Asymmetric Traveling Salesman Problem (ATSP) by adding a dummy vertex v' into the original graph, and also adding a set of zero-weight arcs, $\mathcal{A}_0$, connecting v' to the previously existing vertices. More formally, an ATSP instance $G' = (\mathcal{V}', A', w')$ arising from $G = (\mathcal{V}, A, w)$ is such that $\mathcal{V}' = \mathcal{V} \cup \{v'\}$, $\mathcal{A}' = \mathcal{A} \cup \mathcal{A}_0$, where $\mathcal{A}_0 = \bigcup_{v \in \mathcal{V}} \{(v', v), (v, v')\}$, and, for all $a' \in \mathcal{A}'$,

$$w'(a') = \begin{cases} w(a'), & \text{if } a' \notin \mathcal{A}_0 \\ 0, & \text{otherwise.} \end{cases} \tag{3}$$

Let the cycle $v_{\pi_1} v_{\pi_2} v_{\pi_3} \dots v_{\pi_n} v_{\pi_1}$ be a solution for ATSP, and assume v_{π_1} is the aforementioned dummy vertex. A solution for MCHPP is obtained by removing v_{π_1} and its incident arcs, which results in the simple path $v_{\pi_2} v_{\pi_3} \dots v_{\pi_n}$. ATSP is indirectly solved by reformulating it as a TSP following the transformation proposed by Jonker and Volgenant [36], and then invoking Concorde. To enable optimal solutions, Concorde can be configured with the QSOpt Linear Programming Solver[2].

2.2.4 *Semi-automatic Reconstruction*

The few works addressing semi-automatic reconstruction of shredded documents can be categorized into active and passive paradigms. The active paradigm – which comprises most of the literature [15, 22, 32, 87] – assumes that the user is an inherent part of the reconstruction process. In the passive paradigm (ours), user intervention is optional since a preliminary solution can be automatically obtained. In this case, the user inputs are used to improve an initial/intermediate solution.

Following the active paradigm, Guo et al. [32] proposed a human mediation module in the context of cross-cut documents where the user is occasionally asked to decide whether shreds belong to the same row. This kind of decision is mostly based on text-line alignment, a condition that is barely present in real-shredded data. In *Deshredder* [15], the reconstruction process is predominantly manual, being the user responsible for moving shreds, correcting their orientation, and deciding whether the shreds fit each other with

2 http://www.dii.uchile.cl/~daespino/QSoptExact_doc/main.html.

the assistance of GUI tools (*e. g.*, zoom-in/out, drag-and-drop, and sliders for threshold definition). The automated part of the system is restricted to show the most relevant matching candidates for a query shred, yet based on thresholds defined by the user. Similarly, Shang et al. [87] built visual interfaces to identify correct matches and to arrange the shreds onto a canvas manually. Although visual tools [15, 87] may be intuitive and helpful, the effort and time required from the user to reconstruct large instances (several pages/documents to be reconstructed) are quite prohibitive. In [22], the correct matches are automatically determined by an algorithm (*oracle*, as they call). User interaction is restricted to a few operations: add missing shreds, adjust position, and correct rotation. Although the oracle introduces a new level of automation, the user still has an active role in the reconstruction process, which hinders scalability for larger instances.

In a different direction, Prandtstetter and Raidl [76] formulate the reconstruction of mechanically strip-shredded documents as an optimization problem where solutions can be obtained fully automatically. Optionally, the user can annotate pairs of adjacent shreds in the solution (as many as the user wants) as correct/positive if its shreds are also adjacent in the original document or as wrong/negative, otherwise. However, an issue in their proposal is that the user has to visualize the reconstruction and select – without any system recommendation – the pairs to be annotated. This can be tiresome and time-consuming considering large instances with thousands of shreds (one of the datasets used in this work has $2,292$ shreds). Moreover, positive and negative pairs are arbitrarily chosen, which can slightly improve the solutions.

As will be discussed in Chapter 5, the proposed framework alleviates the user effort since he/she can focus only on the pairs previously selected by the recommender module of the reconstruction framework. In addition to this, the strategies for pairs selection (query strategies) enable better solutions than arbitrary selection.

2.3 CONCLUDING REMARKS

This chapter covered the main aspects of the literature on reconstruction of shredded documents, a particular case of the jigsaw puzzle solving problem. As discussed in Section 2.2.2, little progress has been achieved in evaluating compatibility between shreds for real-shredded documents. These methods explore different levels of features to accomplish this task: pixel-level similarity (or stroke continuity), shape-level matching, and learning-based matching (*e. g.*, Optical Character Recognition). To the best of our knowledge, our work was the first to explore deep neural networks for robust compatibility evaluation. The reconstruction combining the deep learning approaches with the optimization formulation (Section 2.2.3) is discussed in the next two chapters. Finally, we discussed in the scarcity of literature on interactive reconstruction. To address this problem, it was proposed a human-in-the-loop which is presented in Chapter 5.

3

DEEP RECONSTRUCTION: A CLASSIFICATION-BASED APPROACH

This chapter addresses the first deep learning approach to solve the reconstruction problem. The underlying ideas of this approach were first published as a conference paper [62], and further extended and consolidated in a scientific journal [63]. The latter work is a milestone of our research that, besides the reconstruction technique itself, brought the following contributions to the literature:

- An investigation on the reconstruction of mixed shredded documents (*i.e.*, multi-page reconstruction): results have shown that our method is capable of reconstructing 100 mixed shredded documents (2,292 shreds) with accuracy superior to 90%, which brings the state-of-the-art of the document reconstruction problem to another level;
- The release of a new public dataset[1] with 100 real strip-shredded documents (totaling 2,292 shreds). This addresses the lack of publicly available collections representing real scenarios.

The text is organized in four main sections addressing the (i) reconstruction method, (ii) the experiments to assess the proposed method, (iii) the discussion of the obtained results, and (iv) the concluding remarks of the conducted investigation.

3.1 THE RECONSTRUCTION METHOD

The proposed classification-based approach (illustrated in Figure 9) is essentially divided into training (off-line) and reconstruction (on-line) pipelines. The training (top flow) aims to produce a model capable of quantifying the compatibility between shreds based solely on the content around the cutting sections of digitally-cut documents. Small samples (given the whole document) extracted from these documents are the patterns to be learned. This local approach follows the intuition behind the manual reconstruction, where humans analyze the fitting of shreds based on local matching of fragmented patterns (mainly at text line level). These samples should be categorized as positive if they are likely to appear on real documents, or negative otherwise. In practice, positive samples are cropped from pairs of adjacent shreds, and negative from non-adjacent pairs. The learning process is said to be self-supervised because the adjacency relationship is automatically inferred

1 Available at https://github.com/thiagopx/deeprec-pr20.

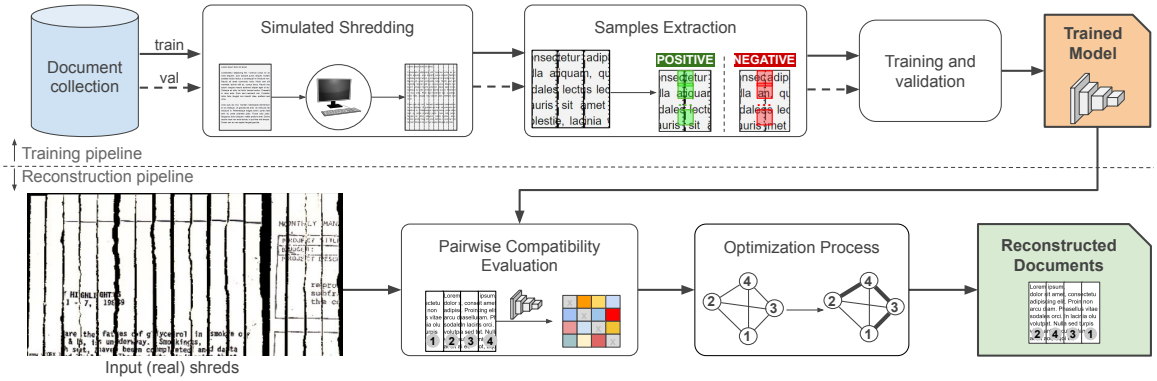


Figure 9: Overview of the proposed system. The training pipeline (top flow) comprises the generation of training data from simulated-shredded documents followed by the training itself, where the best-epoch model is chosen through validation. The reconstruction pipeline (bottom flow) represents the system in operation. The input is a set of real shreds (from one or more documents) and the output is a permutation of shreds (reconstructed documents). The trained model is used for shred's pairwise compatibility evaluation, and the resulting values are the inputs for a graph-based optimization procedure that searches for the best reconstruction.

in simulated shredding. After sampling, the deep model – a Fully Convolutional Neural Network (FCNN) – is trained as a classification model to distinguish between positive and negative samples, being the best model parameters determined through a validation process also using simulated data.

For reconstruction (bottom flow), there is a mild assumption: the shreds of the documents to be reconstructed are already individualized in digital format, *i. e.*, the documents were previously shredded, scanned, and their shreds were segmented. The best model obtained in the training stage is used for pairwise compatibility evaluation of the shreds. The resulting values, arranged as a square matrix, are the input for the graph-based optimization procedure (Section 2.2.3) that estimates a shreds' permutation representing the final reconstruction. The training and reconstruction pipelines are presented in the following subsections alongside a more in-depth description of this reconstruction approach.

3.1.1 *Learning from Simulated-shredded Documents*

The training pipeline aims to produce a model capable of quantifying pairwise compatibility, which means the probability of two shreds being adjacent in a certain order (order matters given the nature of the reconstruction problem). The input can be any collection of digital documents from which all the training data is extracted through simulated shredding. This is particularly beneficial since there is a lack of publicly available datasets containing real-world textual shredded documents, and the generation of such a kind of dataset is tedious, error-prone, and highly demanding because it requires printing, submitting the documents to a paper shredder, manually organizing and scanning the shreds,

and, finally, post-processing them. The generation of training data and the training process itself are discussed in the next sections.

3.1.1.1 *Simulated Shredding*

This process consists of longitudinally slicing digital documents into 30 rectangular regions with the same dimensions, wherein the height of the regions is equal to that of the input image, and the number of regions is an approximation of the number of shreds produced by regular shredders for A4 paper sheets. Text documents in most available public collections are binary or almost binary (*e. g.*, the two collections in Section 3.2.2). This motivated us to adopt a binary representation of the input documents – resulting from applying Sauvola's method [83] – before the simulated shredding.

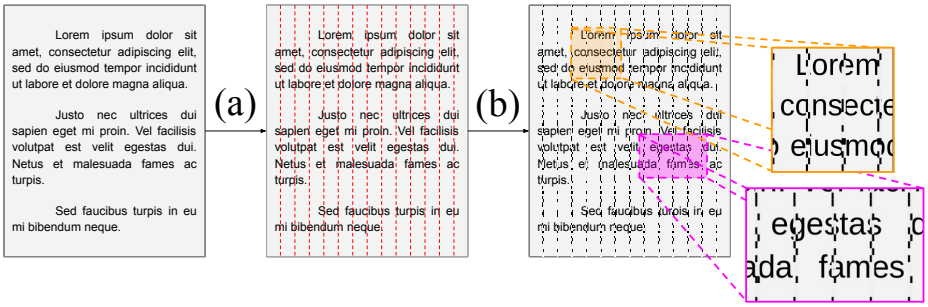


Figure 10: Simulated shredding. The document is digitally shredded into (a) longitudinal cuts (*i. e.*, strip-cut shredding) and random noise is added (b) to the shreds' borders to roughly simulate the damage caused by paper shredders.

The simulated shreds, however, present clean edges, which is very unlikely in real-world mechanical shredding. To cope with that, the original content of the two rightmost and leftmost pixel columns of the shreds is replaced by a black-and-white pattern drawn from the uniform binary distribution $\mathrm{U}(0, 1)$. An overview of the process is depicted in Figure 10.

3.1.1.2 *Sample Extraction*

The input of this step is a document-wise set of digital shreds, and the output is a set of samples to be used in the training of the deep learning model. Given an input document, samples are extracted by pairing shreds and cropping small regions around the touching borders: positive samples come from adjacent shreds (respecting the shreds' ground-truth order) and negatives from non-adjacent shreds (or adjacent shreds in swapped order). As the shreds are automatically obtained, the samples can be self-annotated since the correct sequence of shreds in the original document is known.

Positive and negative training samples were extracted following the same procedure: given a pair of shreds, a sample is a rectangular region of 32×32 pixels (32 rows of the 16 rightmost pixels of the left shred and 32 rows of the 16 leftmost pixels of the right shred).

Such dimensions correspond to the minimum even-valued input size that the adopted network architecture (described in the next section) can handle.

The shreds were sampled every two pixels along the vertical axis, and a limit of 1,000 positive samples per document was fixed. To produce balanced sets, the number of negative samples is limited to the number of positive samples collected in the same document. It is important to mention that the document datasets are available in binary format, as further discussed in Section 3.2.1. In this context, we define the level of information of a sample as the percentage of its text (assumed as black) pixels. For effective training, samples with an information level less than a threshold ρ_{black} are discarded due to the class ambiguity of such cases. This threshold was empirically set to 0.2 based on visual inspection of a few samples: lower than this value, samples usually look like scanning noise.

Before extraction, the pairs of shreds are firstly shuffled to ensure sampling in different regions of the document since the number of samples per document is limited. Note that the extraction procedure is applied to each fragmented document obtained with simulated shredding, one document at a time, resulting in balanced sets of positive and negative samples.

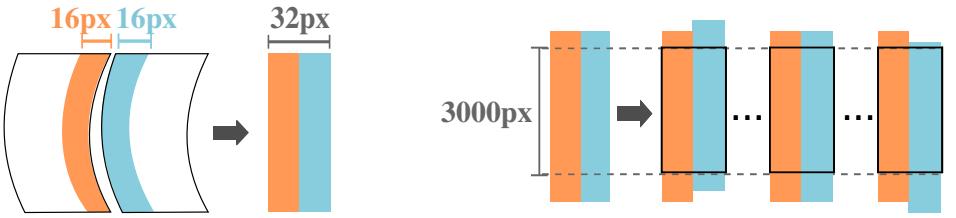
3.1.1.3 *Model Training*

At this point, two balanced sets (positive and negative) of shreds with 32×32 pixels are available for training the deep learning model. The SqueezeNet [35] architecture pre-trained on $227 \times 227 \times 3$ (RGB) images[2] of ImageNet was chosen because it has been shown to be efficient for the classification task, *i. e.*, it can achieve good performance with considerable few parameters, and due to its fully convolutional structure, which is particularly interesting during the inference time, as further discussed in Section 3.1.2. More specifically, the vanilla (*i. e.*, no bypass connections) SqueezeNet v1.1 implementation was adopted, which is a modification of the original SqueezeNet with similar accuracy in the classification task, however, with 2.4 times less computation effort[3].

Since SqueezeNet is fully convolutional, it can be fed with images whose dimensions are different from the original input size. Therefore, training with 32×32 samples does not require any further architectural modifications. To leverage the ImageNet's pre-training, the binary samples were replicated to the three channels of the network instead of reducing the network's input to a single channel. The number of filters in the last convolutional layer was reduced from 1,000 (ImageNet's number of classes) to two filters in order to match the positive and negative classes, and the weights for this layer were initialized under a zero-mean Gaussian distribution with a standard deviation of 0.01, as done in the original SqueezeNet implementation.

2 The $224 \times 224 \times 3$ size reported in [35] seems a typo since $227 \times 227 \times 3$ is the size used in the official implementation (https://github.com/forresti/SqueezeNet).

3 https://github.com/forresti/SqueezeNet/tree/master/SqueezeNet_v1.1.



(a) Rectified interest regions.

(b) Network visual field definition by cropping and sliding operations.

Figure 11: Association image extraction.

From the entire database, 90% of the documents (random selection) were designated for training, and 10% for the validation of the model. Therefore, samples of the same document are used exclusively either to train or validate the model. With the architecture properly adjusted for the problem and the weights initialized, the training can begin. The model was trained during 10 epochs in mini-batches of 256 images using the Adam optimizer with default settings [38] and the categorical cross-entropy loss. The classification accuracy was measured on the validation set at the end of each epoch, and the epoch that yielded the highest accuracy was chosen to determine the "best" model, *i.e.*, the model deployed for compatibility evaluation.

3.1.2 *Reconstruction of Mixed Shredded Documents*

The reconstruction scheme in Figure 9 (bottom pipeline) assumes that the documents were previously fragmented by a paper shredder and that the resulting shreds were scanned and separated into image files at a disk (the semi-automatic segmentation process adopted by our group is detailed in [61]). After loading data, the shreds are also binarized with Sauvola's algorithm [83] since the model was trained with binary samples. Subsequently, as recommended in [76], the blank shreds (*i.e.*, those without black pixels) are discarded since they increase processing overhead without providing relevant information for the forensic examiners. Then, the trained neural model is applied for pairwise compatibility evaluation of the remaining (non-blank) shreds. These compatibility values (arranged as a matrix) are the inputs for the optimization process that determines the reconstruction problem solution: the permutation of shreds that (ideally) reassembles the original documents. As we have commented, the optimization search follows exactly that introduced in Section 2.2.3. The pairwise compatibility evaluation, in its turn, is discussed in the next section.

3.1.2.1 *Pairwise Compatibility Evaluation*

The goal of this stage is to estimate a compatibility value for every pair $(s_i, s_j) \in \mathcal{S}^2$, $i \neq j$, where $\mathcal{S} = \{s_i\}_{i=1}^{n}$ is the set of non-blank shreds resulting from mixing the shredded documents to be reconstructed. The compatibility values are arranged in a square matrix

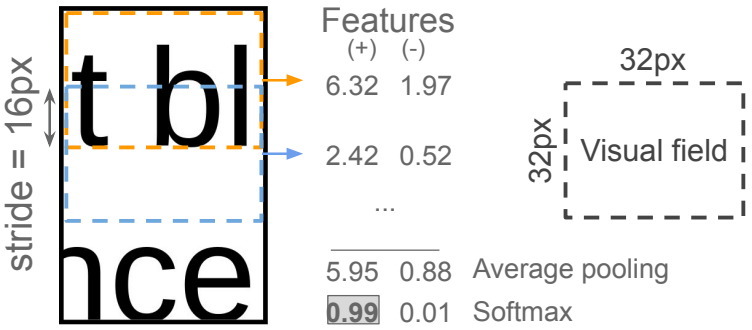


Figure 12: Compatibility computation as a sliding window operation. Since the model was originally trained on 32×32 images, applying it to 3000×32 images is equivalent to sliding vertically the 32×32-size model with implicit stride of 16. In the example, the compatibility candidate value is 0.99, the positive *softmax* probability.

$\Gamma = [\Gamma_{i,j}]_{n \times n}$ where each entry $\Gamma_{i,j}$ matches the compatibility for (s_i, s_j). In other words, $\Gamma_{i,j}$ quantifies how likely s_j is the right neighbor of s_i in the original document. The estimation of $\Gamma_{i,j}$ is focused on regions around the edges of s_i and s_j (see Figure 11a). The 16 rightmost pixels of each row of s_i are joined (at left) with the 16 leftmost pixels of s_j, giving rise to a $H \times 32$ rectified image, where H is the minimum height of both shreds. The rectified image carries the information to be evaluated by the trained model. To account for vertical misalignment, different images are derived from the rectified image by vertically shifting its right part (blue area) s units in the range $[-v_{shift}, v_{shift}]$. Let each of these images to be denoted by I_s, the subscript s indicating the vertical shift. By default, v_{shift} is set to 10, thus 21 different images (*i.e.*, $2v_{shift} + 1$) should be evaluated. Only the 3,000 center rows are considered in computation, as illustrated in Figure 11b.

For faster inference, the derived images are bundled in a batch of size 21 and processed by the deployed neural network. Since the SqueezeNet architecture is fully convolutional and was trained with images of 32×32 pixels, the inference on a 3000×32 image is equivalent to sliding vertically the 32×32-size trained network across the input image with an implicit stride of 16, as illustrated in Figure 12. Note that inference on 32×32 pixels produces a pair of feature values (positive and negative). When applied to a 3000×32 image, an inference produces a 187×2 feature map ($187 = \lfloor \frac{3000}{16} \rfloor$). After global average pooling, the map is reduced to a pair of positive/negative logits from which probabilities are obtained via *softmax*. The compatibility is then set to the highest positive probability in a total of 21 values. More formally,

$$\Gamma_{i,j} = \max_{s \in [-v_{shift}, v_{shift}]} \sigma^+(\mathbf{y}(I_s)), \tag{4}$$

where $\mathbf{y}(I)$ represents the network's logits output given the image I, and $\sigma^+(\mathbf{y})$ the positive probability computed by the softmax function on $\mathbf{y}$.

3.1.2.2 *Optimization Search*

As stated in Section 2.2.1, the objective of the optimization search is to obtain a permutation of shreds π_S^* that minimizes the objective function in Equation (1). Given that the MCHPP is a minimization problem, a cost matrix $\mathbf{\Phi} = [\Phi_{i,j}]_{n \times n}$ is first derived from the compatibility matrix by setting $\mathbf{\Phi} = \max(\mathbf{\Gamma}) - \mathbf{\Gamma}$, where $\max(\mathbf{\Gamma})$ is the maximum value (excluding the diagonal) of the compatibility matrix $\mathbf{\Gamma}$. The cost matrix can be viewed as a complete directed weighted graph $G = (\mathcal{V}, \mathcal{A}, w)$ – instance for the MCHPP –, where a vertex $v_i \in \mathcal{V}$ maps to a shred $s_i \in \mathcal{S}$, $\mathcal{A}$ is the set of arcs, and $w : \mathcal{A} \to \mathbb{R}$ is the weight function defined such that $w((v_i, v_j)) = \phi(s_i, s_j) = \Phi_{i,j}$. With the graph G, a permutation of shreds can be obtained by following the steps described in Section 2.2.3.

3.2 EXPERIMENTAL ASSESSMENT

The general purpose of the experiments is to evaluate the reconstruction accuracy by mixing different quantities of single-page shredded documents (hereafter referred to as documents for simplicity) following an incremental strategy. Besides the two evaluation datasets used in [61], a new collection (referred to as S-CDIP) with 100 documents was assembled specifically for this investigation.

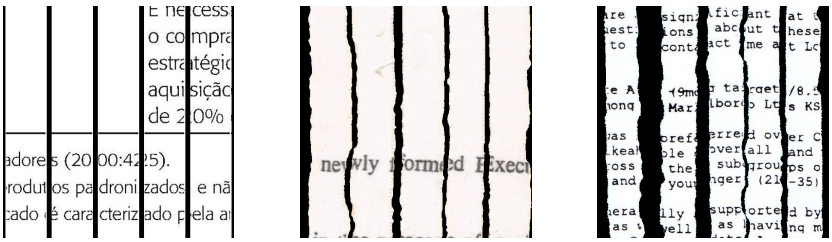
The experiments were divided into three main parts. First, the proposed method was evaluated in its default configuration. Then, an ablation study was conducted to assess the sensibility of our method concerning three key parameters. The last part is a comparative evaluation with state-of-the-art methods available in the literature. The following sections describe, respectively, the datasets used for quality assessment, the conducted experiments, and the computational platform on which the experiments were carried out.

3.2.1 *Training Datasets*

As stated in Section 3.1.1.3, the training of the model for compatibility evaluation should take, as input, any document collection. In fact, different training datasets should be provided to enable cross-database evaluation. Here, two collections of scanned documents were used (one at a time) to extract training samples: ISRI-OCR and CDIP.

ISRI-OCR. This collection comprises a subset of the ISRI-Tk OCR collection [58] which includes 800 binary documents (originally scanned at 300 dpi) labeled as *reports*, *business letters*, or *legal documents*. The structure of these documents has a high degree of similarity, generally focusing on running text at the expense of graphical elements (*i. e.*, pictures, tables, graphs).

CDIP. This dataset comprises 100 documents from the RVL-CDIP collection [34], of which there are 10 documents from each of the following classes: *form*, *email*, *handwritten*,



(a) S-MARQUES. (b) S-ISRI-OCR. (c) S-CDIP.

Figure 13: Samples (cropped view) of the shredded (denoted by the "S-" prefix) datasets.

news article, *budget*, *invoice*, *questionnaire*, *resume*, and *memo*. In summary, this dataset has a more diverse collection of documents. The documents were chosen arbitrarily, except for the restriction that they should present textual content at some level. Since the RVL-CDIP is a subset of the IIT-CDIP Test Collection 1.0 [44] but in lower resolution, we decided to use the corresponding 300 dpi images from the original IIT-CDIP dataset (the resolution matches that of the evaluation datasets). It is worthy to mention that the scanning resolution is up to the user once the paper shreds are available.

3.2.2 *Evaluation Datasets*

Three datasets were used to evaluate the methods: S-MARQUES, S-ISRI-OCR, and S-CDIP (the last is a contribution of this work). The "S-" prefix stands for mechanically "shredded" and was used to differentiate from the training datasets, which comprise the original (unshredded) documents.

S-MARQUES. This collection refers to the 60 text documents in Portuguese of the strip-shredded dataset produced by Marques and Freitas [53]. To create this dataset, the authors collected 60 paper documents (a digital backup is also available along with the dataset) and shredded them using a Cadence FRG712 strip-cut machine. The resulting shreds were scanned at 300 dpi and then separated into JPEG files (one for each shred). Compared to the other datasets, as shown in Figure 13, S-MARQUES' shreds have a more uniform shape, and are less damaged by the shredder's blades, *i. e.*, they are less curved and their borders are less corrupted (smooth serrated effect).

S-ISRI-OCR. This dataset was originally produced and used in previous work of our research group [61] in the context of this project. It was assembled from a set of 20 business letters and legal reports of the ISRI-Tk OCR collection, the same set used in [3] to assess the reconstruction of simulated-shredded documents. The digital documents were printed onto A4 paper and subsequently submitted to a Leadership 7348 strip-cut paper shredder. To expedite the acquisition process, the shreds were spliced onto a high-contrast paper,

and, after scanning (at 300 dpi), they were segmented and stored individually in JPEG files. This process is more detailed in [61].

S-CDIP. The S-CDIP dataset, which is a particular contribution of this work, is the shredded version of the 100 digital documents in CDIP. The same methodology to create S-ISRI-OCR was also adopted for this dataset. As illustrated in Figure 13, the shreds of S-ISRI-OCR and S-CDIP depict a higher degree of vertical misalignment in view of S-MARQUES, as well as more damage at their extremities.

3.2.3 *Experiments*

In the preliminary work [62], the experiments were not cross-database since the documents of S-ISRI-OCR were reconstructed with a model trained on documents of the ISRI-OCR Tk collection. In practice, such experiments assume the availability of training data that share significant appearance and structural similarities with test data. For a more realistic scenario, the experiments conducted here followed a cross-database protocol in which testing on S-CDIP (the dataset produced in this work) leverages the model trained on ISRI-OCR, and testing on S-ISRI-OCR and S-MARQUES uses the model trained on CDIP.

The evaluation is performed incrementally so that new shredded documents are gradually introduced to the reconstruction instance. For the sake of notation, let k denote the number of mixed documents of a particular instance. The main purpose of the incremental approach is to evaluate whether the reconstruction accuracy degrades with the increase of k. Due to the processing burden of this type of experiment, the ablation study evaluates incrementally $k = 1, 2, \ldots, 5$ documents at a time, while the other two experiments also include $k = 10, 15, 20, \ldots, n_{docs}$, where n_{docs} denotes the size of the current evaluation dataset (60, 20, or 100, as described in Section 3.2.2). For each k value, a set of k-size instances (*i. e.*, k mixed documents) should be sampled. Note that the size of the sample space varies significantly over k. For example, there are $\binom{100}{2} = 4{,}959$ possible ways of combining 2 S-CDIP's documents, whereas for $k = 3$, this number rises to 161,700. Instead of independently sampling combinations, the test instances are assembled in such a way that the k-size instances are obtained by adding a single document to each instance of size $k-1$. We assume the documents are arbitrarily ordered and that k-size instances (*i. e.*, groups of k documents) are assembled by grouping consecutive documents.

Formally, let $\{\mathcal{S}_i\}_{i=1}^{n_{docs}}$ be the collection of shredded documents (the order is arbitrary) to be reconstructed, and $\mathcal{S}_{a:b} = \{\mathcal{S}_i\}_{i=a}^{b}$ a subset of this collection. Then, the test instances for a particular k include the sets $\mathcal{S}_{1:k}$, $\mathcal{S}_{2:(k+1)}$, ..., $\mathcal{S}_{(n_{docs}-k+1):n_{docs}}$. Note that this yields overlapping of test instances for the same k, as well as across k values.

EXPERIMENT 1: DEFAULT CONFIGURATION. In the first experiment, the incremental procedure was used to assess the robustness of the proposed method in its default param-

eters' configuration: $\rho_{black} = 0.2$ and samples size of 32×32 (defined in Section 3.1.1.2), and $\nu_{shift} = 10$ (defined in Section 3.1.2.1).

EXPERIMENT 2: ABLATION STUDY. For the second experiment, an ablation study was carried out to evaluate the sensitivity of the system with respect to the aforementioned parameters, one at a time. The parameters' domain for ρ_{black}, sample size, and ν_{shift} were set to $\{0.1, 0.2, 0.3\}$, $\{32 \times 32, 32 \times 64, 64 \times 32, 64 \times 64\}$, and $\{0, 5, 10, 20\}$, respectively. The investigation of ρ_{black} aims to verify the system's robustness with respect to the amount of information contained in the samples, which can vary for different font types and sizes. The desirable behavior is that the average accuracy holds for the widest possible range of ρ_{black}. The motivation behind the analysis of the sample sizes is to confirm whether the locality assumption holds or not. Notice that training with 64-width samples requires adjusting the input window to 3000×64 in the pairwise compatibility evaluation stage (Section 3.1.2.1). The analysis of ν_{shift} evaluates the need for (vertically) aligning the shreds at test time since no image processing was previously applied to this intent.

EXPERIMENT 3: COMPARATIVE EVALUATION. The final experiment aims at comparing our method – referred to as DEEPREC-CL (**Deep rec**onstruction based on **cl**assificaton) – against three relevant methods of literature, here designated with the name of the first author. The first is referred to as Paixão, our preliminary method based on shape matching of characters' fragments. [61]. The original implementation, intended for single-document reconstruction, uses caching of shape dissimilarities to improve time efficiency, which, on the other hand, compromises memory scalability for multi-page reconstruction. Therefore, reconstruction with this method was limited to $k = 5$ documents. The second method is the one proposed by Liang and Li [47], which is referred to as Liang. Due to time restrictions of the provided implementation[4], the multi-reconstruction experiment was run only for the datasets S-MARQUES and S-ISRI-OCR limited to $k = 3$ documents. We adopted the parameters for the real-shredded instances 1 and 2 (a total of 3) of the original work. For the matter of consistency, we configured the OCR software on which the Liang method relies to the Portuguese language when testing on S-MARQUES. The last method, referred to as Marques [53], relies on edge pixel dissimilarity for compatibility evaluation and was chosen due to its superior performance compared to other methods of literature, as can be seen in [61, 62]. While Paixão and DEEPREC-CL share the same optimization formulation, Marques uses a simple greedy nearest-neighbor approach. Thus, for a fairer comparison, our system was also evaluated with Marques' optimization model to emphasize the role of compatibility evaluation in producing accurate reconstructions. The modified method is referred to as DEEPREC-CL-NN.

4 https://github.com/xmlyqing00/DocReassembly.

3.2.4 *Experimental Platform*

The experiments were carried out on two machines: (M1) an Amazon AWS instance with 8 vCPUs (2.3GHz), 60GB RAM, and a GPU NVIDIA Tesla V100 (16GB); (M2) an Intel Xeon E7-4850 v4 (2.10GHz) with 128 vCPUs, 252GB RAM. The ablation study was fully performed on M1. The methods Liang, Paixão, and Marques, which do not require GPU processing, were conducted on M2. As Liang leverages OpenMP[5] directives to improve efficiency, we used 240 threads (120 vCPUs) in the experiments. For **Deeprec-CL**/Deeprec-CL-NN, the compatibility evaluation in experiments 1 and 3 was carried out on M1, while the optimization process was performed in M2 due to the large memory resources. The proposed system was implemented in Python with TensorFlow for training and inference, and with OpenCV for image processing. The code, pre-trained models, and datasets are publicly available at https://github.com/thiagopx/deeprec-pr20.

3.3 RESULTS AND DISCUSSION

3.3.1 *Experiment 1: Default Configuration*

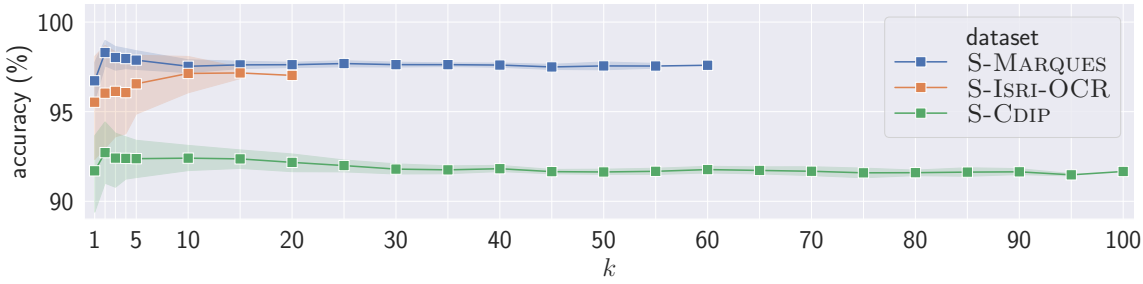


Figure 14: Multi-reconstruction accuracy across datasets (k is the number of mixed documents). The square markers represent the mean values w.r.t. the documents instances, and the shadowed areas represent the 95% confidence interval.

Figure 14 shows the multi-reconstruction accuracy (mean and 95% confidence interval) obtained with the proposed method (default parameters) for three evaluation datasets. Overall, the proposed method performed above 90% for the three datasets, and, comparatively, S-Cdip was verified, as expected, the most challenging test collection (an example of reconstruction is shown in Figure 15). The confidence interval tends to be wider as fewer documents are available, which is the case of S-Isri-OCR. Furthermore, the accuracy tends to stabilize for large k, which means that the insertion of new documents into the reconstruction instance does not degrade accuracy, even though it increases considerably the complexity of the problem.

Breaking down the performance on the S-Cdip dataset, Figure 16 shows accuracy boxplots for single-reconstruction ($k = 1$) across the dataset categories. From the 100 docu-

5 www.openmp.org.

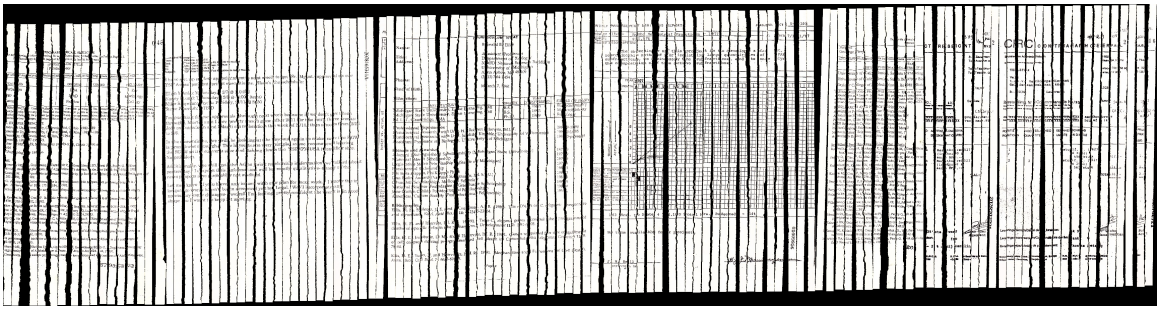
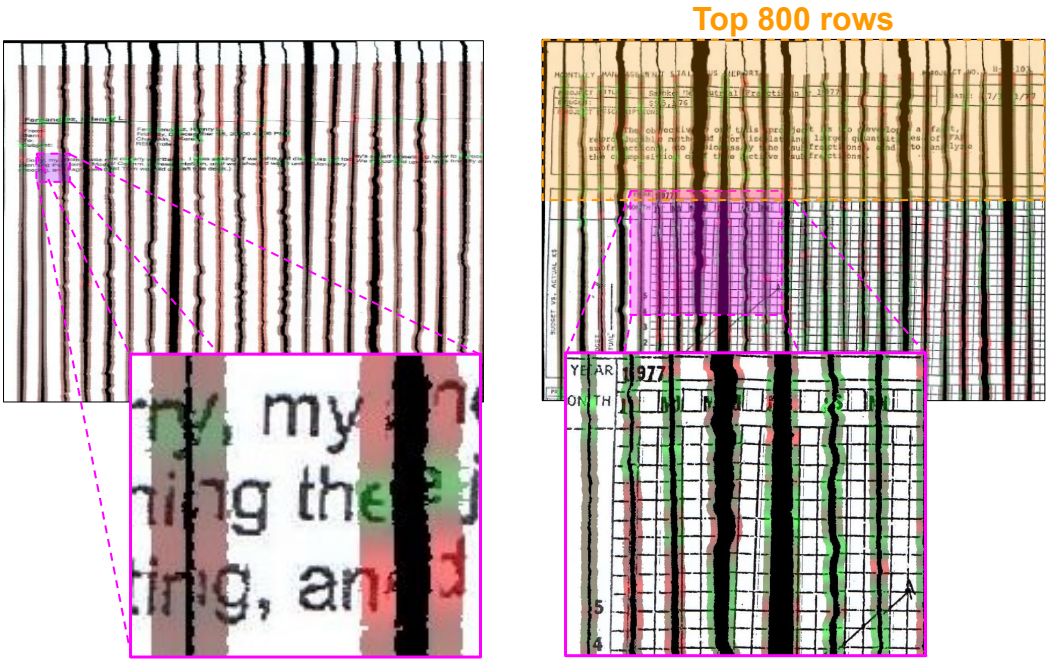


Figure 15: Reconstruction of a test instance comprising shreds from $k = 5$ documents of S-Cdip with accuracy of 82.93%. The shreds were placed side-by-side without any rotation correction. Each new inserted shred was vertically shifted according to the optimal s value in Equation (4). The full reconstruction ($k = 100$) can be viewed at https://htmlpreview.github.io/?https://github.com/thiagopx/docs/blob/master/results_s-cdip.html.

Figure 16: Reconstruction accuracy for S-Cdip across categories ($k = 1$).

ments, 32 were perfectly reconstructed, and only 5 had accuracy lower than 70%. Remarkably, the reconstruction of handwritten documents achieved high accuracy (8 in 10 were superior to 90%) although no handwritten document was used to train the compatibility evaluation model. For documents of the type form, all reconstructions achieved accuracy higher than 90%, being 5 of them perfect. The results for the *handwritten* and *form* categories show that learning is not restricted to the symbolic level, and that lower-level features (*e. g.*, strokes and horizontal lines) can also be learned by the model.

As seen in Figure 16, the *letter* category has a larger variability in comparison to the others. In this category, there are three documents with very small fonts and whose shreds have degraded borders beyond the regular corruption found in most shreds. Although the accuracy for these three documents is low ($< 75\%$), such values are not low enough to be considered outliers, which explains the elongated aspect of the letter's boxplot. The poor outlier performance, more evident in the *budget* and *email* categories, is mainly caused by three factors that may occur in combination or separately: (i) low quality of text symbols (*i. e.*, low resolution, corrupted data), large flat areas (*i. e.*, low amount of information), or (iii) large areas covered by patterns not learned by trained model.



(a) (b)

Figure 17: Challenging reconstruction instances (shreds are placed on the ground-truth order). Positive (green) and negative (red) activation maps for the adjacent shreds were superimposed onto the shreds. (a) Email with large blank areas and corrupted characters for which the reconstruction accuracy was 52%; (b) Budget document with a large grid pattern. The vertical lines of the grid induce negative activation (red) when they get close to the cut section (center of the network visual field). By cropping the top 800 rows of the shreds (roughly indicated by the orange area), the accuracy rises from 45.83 to 75%.

These challenging factors are illustrated in Figure 17. The shreds were placed side-by-side in the ground-truth order and the activation maps from SqueezeNet's last convolutional layer were adjusted and superimposed on the shreds' boundary zones. Green areas represent a high degree of compatibility, while red ones represent the opposite. Neutral zones are usually gray, indicating a balance between the positive and negative classes. Nonetheless, it can be noticed in Figure 17a reddish areas for neutral zones due to bias, or caused by noise (small black regions) in the highlighted areas close to the borders.

In the first case (Figure 17a), an email document with large blank areas and corrupted characters was reconstructed with 52% of accuracy. Due to the low amount of information, the compatibility evaluation and, as a consequence, the reconstruction accuracy is more sensitive to corrupted data. The second document (Figure 17b) is a budget with a large area covered by a grid pattern, and for which the obtained accuracy was 45.83%. Unlike the horizontal lines, which are captured by the model, the vertical lines lead to erroneous evaluations by the model. This is justified by the scarcity of such patterns in the training set, which comprises images from ISRI-OCR. By restricting the shreds to their first 800 rows (orange highlighted region in Figure 17b), the reconstruction accuracy increases to 75%. Although the aforementioned cases yielded low-accuracy reconstructions, it does

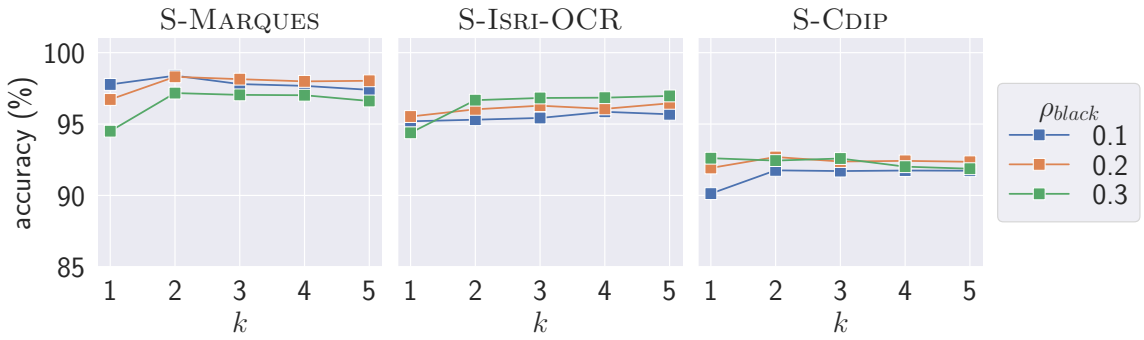


Figure 18: Investigation of the accuracy sensitivity w.r.t. the parameter $\rho_{\mathtt{black}}$.

Figure 19: Investigation of the accuracy sensitivity w.r.t. the size of training samples.

not mean that the same behavior will invariably be observed for documents with similar layout/features. The reconstruction quality also depends on where the cuts take place. In S-CDIP, for example, there are an email and a budget document visually similar to those in Figure 17a for which the accuracy was 80 and 84%, respectively.

3.3.2 *Experiment 2: Ablation Study*

The results for the three investigated parameters are summarized in Figures 18, 19, and 21. Figure 18 shows the accuracy sensitivity with respect to the parameter $\rho_{\mathtt{black}}$. Ideally, the system is expected to be robust to changes in this parameter. From the results, it can be observed a wider variation range for $k = 1$. The performance difference becomes less noticeable as k increases, which represents a more realistic scenario for the reconstruction application.

Figure 19 shows the impact of the size of training samples on the final reconstruction performance. In general, the system generalizes better across the datasets for samples with reduced width, *i. e.*, 32×32 and 64×32. By keeping the samples narrower, visual ambiguity (illustrated in Figure 20) can be explored in compatibility evaluation of scarce/unseen patterns in the training data. For instance, the model can perceive a "wo" association as valid (as in "**wo**rld") if samples with "vo" (as in "**vo**xel", "**vo**lume", and "reser**vo**ir") were observed during the training. The results for 64×64 samples were competitive in

(a) (b)

Figure 20: The visual ambiguity between "wo" and "vo" is illustrated in (a) for a 32×32 input window highlighted in red. Such ambiguity is not seen (b) after increasing the width of the input window (highlighted in blue).

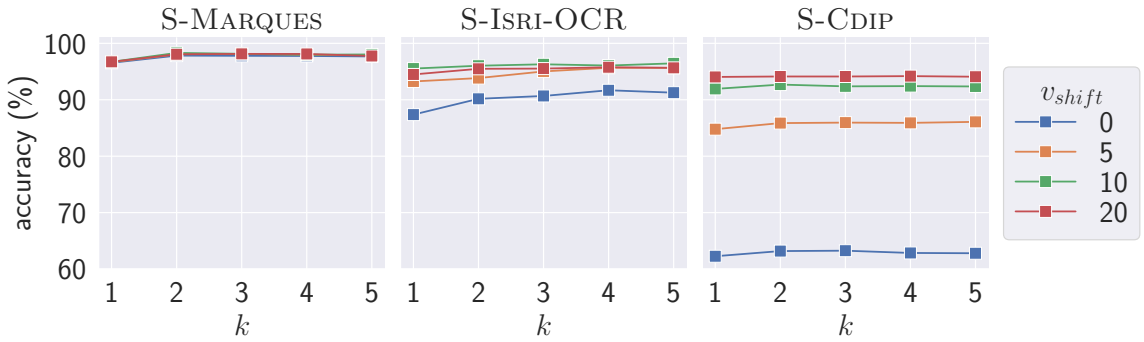


Figure 21: Investigation of the accuracy sensitivity w.r.t. the parameter $v_{\texttt{shift}}$.

terms of accuracy on the S-Isri-OCR, where the documents have primarily textual content. However, the performance dropped significantly for documents with a higher density of graphical elements (*e. g.*, forms and budgets) present in S-Marques and S-Cdip.

Finally, Figure 21 shows the influence of the vertical shift range parameter ($v_{\texttt{shift}}$) on the reconstruction performance. In practice, no sensitivity to this parameter was observed for S-Marques since the shreds for this collection are (practically) vertically aligned, as exemplified in Figure 13a. In contrast, the results on the S-Isri-OCR and S-Cdip datasets, which better depict real-world conditions, show the relevance of properly treating the misalignment between shreds. The misalignment degree is higher for S-Cdip, which explains the consistent accuracy improvement with the increase of $v_{\texttt{shift}}$.

3.3.3 *Experiment 3: Comparative Evaluation*

Figure 22 shows the comparative performance with the literature. The average accuracy of the proposed method using Concorde (Deeprec-CL) was consistently superior to the compared methods. Additionally, it demonstrated greater robustness, which is mainly evidenced by the stability of the accuracy curve with the increase of k.

Unlike the proposed method, the modified version (Deeprec-CL-NN) – intended for comparison with Marques – presented a decay in accuracy with the increase of k. Nevertheless, it greatly outperformed Marques, which also uses the same optimization approach, and Paixão, which leverages Concorde. In fact, Marques struggles with black-white documents since it is based on color features. Moreover, it is very sensitive to the

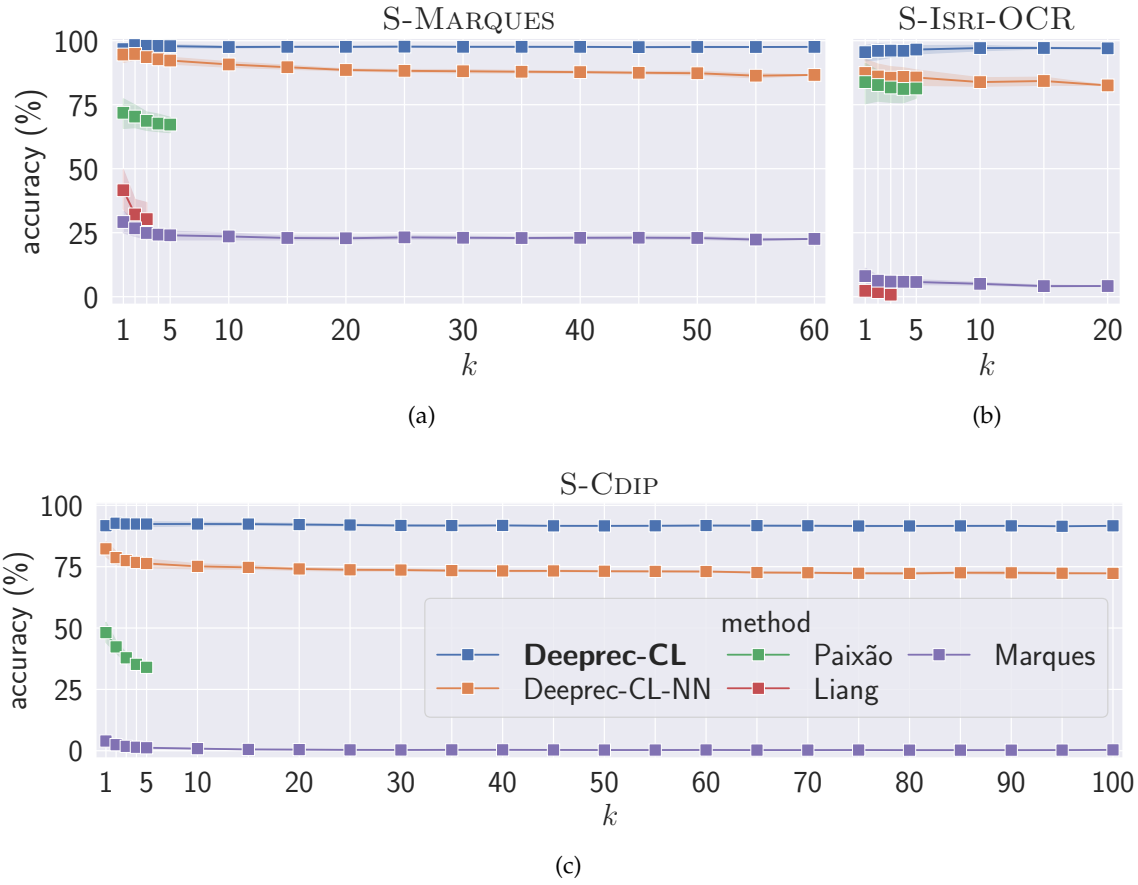
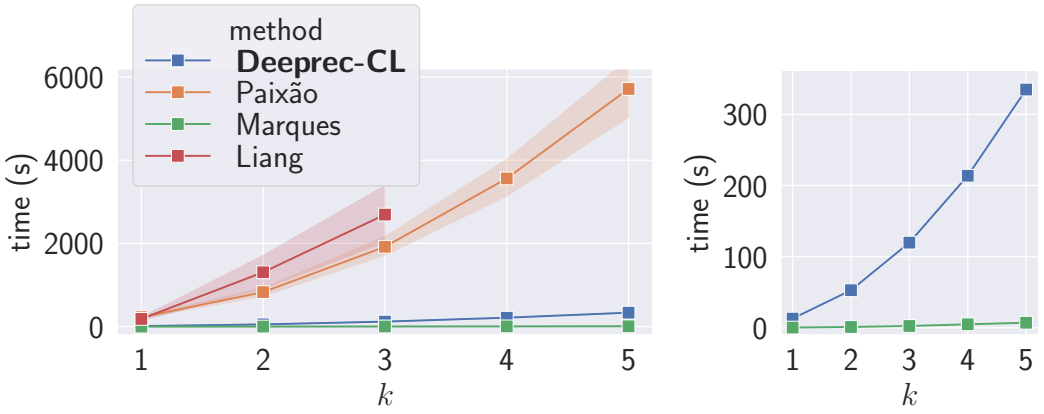


Figure 22: Comparative accuracy performance: the proposed approach achieved the highest accuracy for all three test sets.

damage on the shreds' borders caused by the mechanical fragmenting process, and to the vertical misalignment of the shreds. Both issues are accentuated in the S-Isri-OCR and S-Cdip datasets, resulting in a significant drop in performance when compared to S-Marques. It can also be observed that the accuracy of Paixão degrades more sharply for S-Cdip, which is explained by the large presence of pictorial elements (as depicted in Figure 17b), and also by a greater diversity of symbols in different font types, sizes, and styles (including handwritten characters). When mixing documents, such diversity becomes a critical factor since Paixão assumes a fixed-size alphabet in which each symbol has a unique representative. For single-reconstruction ($k = 1$), Liang was capable of reconstructing 7 pages of S-Marques (in a total of 60) with 100% of accuracy. These instances have a great concentration of text and no pictorial content. Nonetheless, the average accuracy considering all the 60 pages was under 50% with a sharp decay as k increases. The observed decay corroborates the scalability issue raised by the authors and mentioned in Section 2.2.2. Like Marques, the accuracy was dramatically worse for S-Isri-OCR than for S-Marques. This is because Liang strongly relies on the OCR capability of recognizing full words on composition of shreds (visually similar to Figure 15), and such capability is substantially affected by geometric distortion between shreds.



(a) (b)

Figure 23: Comparative time performance (measured on S-MARQUES and S-ISRI-OCR). The difference between Marques and the proposed method is better noticed in (b). Despite Marques is very efficient, it does not deliver accurate solutions. The proposed approach is a feasible alternative since it reaches accurate solutions and is significantly more scalable than Paixão and Liang.

Besides reaching better accuracy, the proposed approach is also remarkably more scalable in terms of time performance than Paixão and Liang, as seen in Figure 23. Time scalability is a critical issue in real scenarios because it is expected much more than 5 shredded pages to be reconstructed. Note, in particular, that the time performance of Liang was the worst even leveraging heavy parallelism (240 threads). This is due to the overhead introduced by successive calls to the OCR software, which is the core of their method. Conversely, Marques is very time efficient, but, as shown in Figure 22, it delivered low accuracy reconstruction. Nonetheless, Marques' time performance will serve as the lower bound of efficiency for future optimizations of the proposed method.

3.4 CONCLUDING REMARKS

This chapter addressed the classification-based approach for reconstruction of mixed text documents focusing on the central problem of evaluating the compatibility between shreds. The proposed segmentation-free deep learning approach has enabled faster and more robust reconstruction of strip-shredded documents in more realistic scenarios and it also has the benefit of self-supervised learning, which facilitates scaling the training data.

To enable a better and more extensive evaluation, we introduced a dataset comprising 100 mechanically-shredded documents (2,292 shreds) with diverse layout. Despite the challenging scenarios, real-world cross-database experiments showed that our method achieved average accuracy superior to 90% for different quantities of mixed documents. Nevertheless, the absence or scarcity of some patterns may hamper the proper reconstruction of the documents. A possible way to solve this problem is fine-tuning the model with samples from the inner region of the shreds belonging to the test documents themselves.

The ablation study evidenced that small and local samples are more effective for learning the compatibility between shreds. It is important, though, to consider this result in view of the limited diversity of the training data produced from relatively – in the context of deep learning – few documents. Additionally, the study showed the relevance of treating the misalignment between shreds at test time. An alternative approach for this issue is augmenting training data by simulating vertical misalignment. This would save processing time during the online reconstruction stage but could increase the complexity of the problem.

Comparative experiments showed that the accuracy of the proposed method (even in the modified version) was superior to the current state-of-the-art. When compared to Paixão [61], for instance, our method generalized better for documents with a more diverse layout and appearance, and also scaled more time-efficiently for the multi-page scenario. Furthermore, the time savings obtained by Marques [53] (based on the naive dissimilarity between edge pixels) were shown at the price of low reconstruction accuracy. Finally, the recently published Liang method [47] performed significantly worse than the proposed method in terms of accuracy, in addition to a limited time-scalability to real-world scenarios comprising several documents.

In addition to the mentioned directions, there is an important issue to be addressed: the time performance when scaling up to larger instances, *i. e.*, when there are more shreds to be analyzed. This motivated us to develop a novel deep learning approach that can still benefit from the self-supervised learning paradigm, however with significant effort reduction in processing the pairwise compatibilities, therefore improving the time performance of the overall pipeline. This novel approach is presented in the following chapter.

4

DEEP RECONSTRUCTION: AN ASYMMETRIC METRIC-LEARNING APPROACH

The critical issue for the time performance of the classification-based approach is the need for inference whenever each pair of shreds is evaluated. In other words, considering a network inference as the time unit cost, we can say that such an approach scales quadratically with the number of shreds.

To deal with this issue, we discuss in this chapter an approach 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 [14] in signature verification, to an application of the triplet loss [98] in face verification [85], to cite a few. Unlike most of these works, however, the proposed method does not employ the same model for semantically different samples. In our case, right and left shreds are (asymmetrically) projected by two different models onto a common space so that the measured distance between them are interpreted as cost in the optimal reconstruction framework.

The following contributions are covered in this chapter:

1. A compatibility evaluation method leveraging metric learning and the asymmetric nature of the problem;
2. As the classification-based approach, the proposed method does not require manual labels (trained in a self-supervised way) nor real data (the model is trained with artificial data);
3. 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 a higher number of shreds.

The following sections cover, respectively, (i) the reconstruction method, (ii) the experimental assessment, (iii) the discussion of the results, and (iv) the final remarks.

4.1 THE RECONSTRUCTION METHOD

The novelty in the proposed reconstruction method is the metric-learning approach for compatibility evaluation whose general intuition is illustrated in Figure 24 (real embed-

dings are shown in Appendix A.3). 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 $\boldsymbol{e}$) by projecting them onto a common embedding space $\mathbb{R}^d$. Projection is accomplished by two models (CNNs): f_{left} and f_{right}, $f_\bullet : \mathbf{x} \mapsto \boldsymbol{e}$, specialized on the left and right boundaries, respectively.

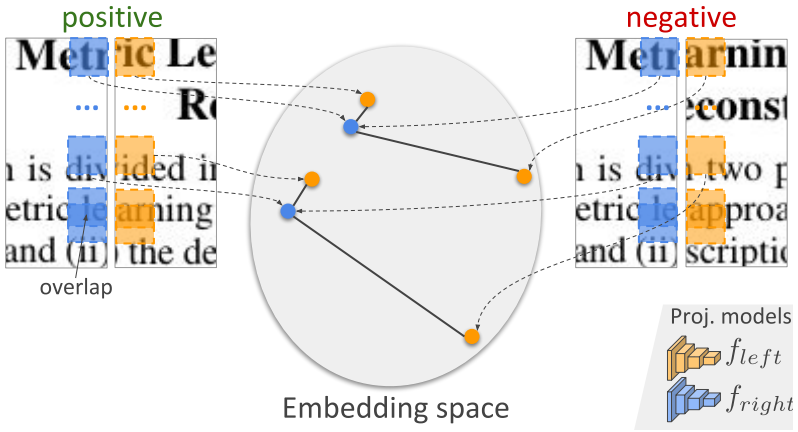


Figure 24: Metric learning approach for shreds' compatibility evaluation. Embeddings generated from compatible regions are expected to be closer in the embedding space, whereas those from non-fitting regions are expected to be mapped far from each other.

Assuming that these models are properly trained, boundary samples (indicated by the orange and blue regions in Figure 24) 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:

$$\phi(s_i, s_j) \propto \text{dist}(\boldsymbol{e}_i, \boldsymbol{e}_j), \tag{5}$$

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

The interesting property of this evaluation process is that the projection step (network inference) 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 the evaluation will be presented, including the formal definition of a cost function that composes the global objective function (Equation (1)).

4.1.1 *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 [35] architecture (*i. e.*, until the *fire3* block).

SqueezeNet has been effectively used in distinguishing between valid and invalid symbol patterns in the context of compatibility evaluation [62, 63], as discussed in the previous chapter. 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.

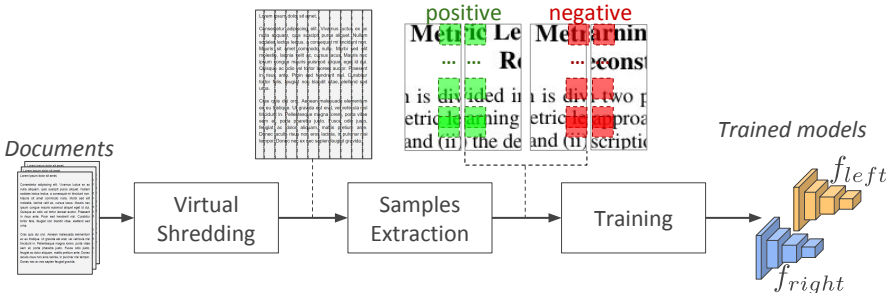


Figure 25: Self-supervised learning of the models with samples extracted from digital documents.

Figure 25 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 (x_r, x_l, y), where x_r and 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

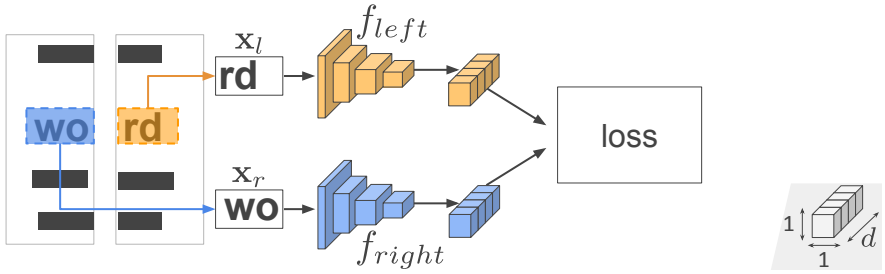


Figure 26: Learning projection models for shreds' compatibility evaluation. The models are jointly trained with sample pairs guided by the contrastive loss function. The input vectors for the loss function are encoded as $1 \times 1 \times d$ tensors.

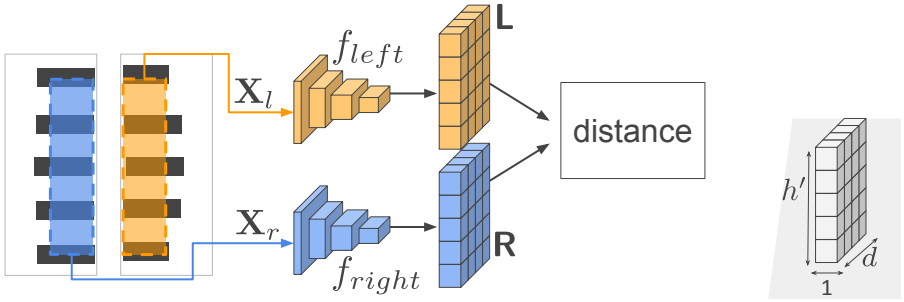


Figure 27: Compatibility evaluation of a pair of shreds. Local embeddings, represented by the $h' \times 1 \times d$ tensors **L** and **R**, are extracted along the boundary regions. Compatibility is a real value given by the squared Euclidean distance between **L** and **R** (computed over the flattened tensors).

sample pair is positive, and $y = 0$, otherwise. Training is driven by the contrastive loss function [19]:

$$\mathcal{L}(f_{left}, f_{right}, x_l, x_r, y) = \frac{1}{2}\{y \cdot dist^2 + (1-y) \cdot [\max(0, m - dist)]^2\}, \tag{6}$$

where $dist = \left\| f_{left}(x_l) - f_{right}(x_r) \right\|_2$, and m is the margin parameter. For better understanding, an illustration is provided in Figure 26. 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 Section 4.2.

4.1.2 *Pairwise Compatibility Evaluation*

For compatibility evaluation, shreds' embedding and distance computation are two decoupled steps. Figure 27 presents a joint view of these two steps for a 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 inter-

est 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 24.

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 [62, 63]. 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 $\mathbf{T}_{a:b} = (T_{i,j,k})_{a \leqslant i \leqslant b,\ j=1,\ 1 \leqslant k \leqslant d}$ 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, the pairing cost is defined by

$$\phi_{\uparrow}(s_i, s_j) = \min_{0 \leqslant \delta \leqslant \delta_{max}} \left\| \mathbf{R}^{(i)}_{1:1+n_{rows}} - \mathbf{L}^{(j)}_{1+\delta:1+n_{rows}+\delta} \right\|_2, \tag{7}$$

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

$$\phi_{\downarrow}(s_i, s_j) = \min_{0 \leqslant \delta \leqslant \delta_{max}} \left\| \mathbf{R}^{(i)}_{1+\delta:1+n_{rows}+\delta} - \mathbf{L}^{(j)}_{1:1+n_{rows}} \right\|_2. \tag{8}$$

Finally, the proposed cost function is a straightforward combination of Equations (7) and (8):

$$\phi(s_i, s_j) = \min(\phi_{\uparrow}(s_i, s_j), \phi_{\downarrow}(s_i, s_j)). \tag{9}$$

Note that, if δ_{max} is set to 0 (*i.e.*, no shifts), then $n_{rows} = h'$, therefore

$$\phi(s_i, s_j j) = \phi_{\uparrow}(s_i, s_j) = \phi_{\downarrow}(s_i, s_j) = \left\| \mathbf{R}^{(i)} - \mathbf{L}^{(j)} \right\|_2. \tag{10}$$

4.2 EXPERIMENTAL ASSESSMENT

The experiments aim to evaluate the accuracy and time performance of the proposed metric learning approach – referred to as DEEPREC-ML (**Deep rec**onstruction based on **m**etric **l**earning) –, as well as to compare with the literature in document reconstruction focusing on the classification-based approach method presented in the previous section [62, 63] (referred to as DEEPREC-CL). For this purpose, we followed the basic protocol proposed in [61] in which the methods are coupled to an exact optimizer (as described

in Section 2.2.3) and tested on two datasets: S-MARQUES and S-ISRI-OCR (introduced in Section 3.2.2). Two different scenarios are considered here: single- and multi-page reconstruction. The accuracy measure in Equation (2) is applied to evaluate the quality of the reconstructions.

4.2.1 *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 [44], *i.e.*, the same documents comprising CDIP (Section 3.2.1). For sampling, the pages are split longitudinally into 30 virtual shreds (estimated from usual A4 paper shredders). Next, the shreds are individually thresholded with Sauvola's algorithm [83] 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 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 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 [20] 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 $\mathrm{SMD}(dist^{+}, dist^{-})$.

DEEPREC-CL. The classification-based approach (DEEPREC-CL) was published in two works – [62, 63] – with slight changes in the experimental setup. By the time the experiments described in this chapter were performed, only [62] was available, being [63] a work in progress. Therefore, the experiments with DEEPREC-CL reflect majorly the implementation described in [62], however, we incorporated some features introduced in the consolidated approach described in the previous chapter [63] to enhance comparison: (i) the samples' size was set to 32×32; (ii) the optimization formulation was set to the optimal one described in Section 2.2.3; and, finally, (iii) the training dataset was changed to CDIP's documents to enable cross-database experiments.

4.2.2 *Experiments*

The experiments rely on the trained models f_{left} and f_{right}, as well as on the DEEPREC-CL's model. As aforementioned, the latter was retrained on the CDIP's 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 [83] to keep consistency with training samples. The default parameters of DEEPREC-ML 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.

EXPERIMENT 1: SINGLE-PAGE RECONSTRUCTION. This experiment aims to show whether the proposed method is able to individually reconstruct pages with accuracy similar to DEEPREC-CL, 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 S-MARQUES and S-ISRI-OCR were individually reconstructed with DEEPREC-ML and DEEPREC-CL methods, first using their default configuration, and after disabling the vertical shifts (in DEEPREC-ML, 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 DEEPREC-ML–, pairwise compatibility evaluation (pw), and optimization process (opt).

EXPERIMENT 2: MULTI-PAGE RECONSTRUCTION. This experiment focuses on 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 S-MARQUES and the 505 mixed shreds of S-ISRI-OCR. Each instance was reconstructed with DEEPREC-ML and DEEPREC-CL 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 S-ISRI-OCR.

EXPERIMENT 3: SENSITIVITY ANALYSIS. The last experiment assesses how DEEPREC-ML is affected (time and accuracy) by testing with different embedding dimensions: $d = 2, 4, 8, \ldots, 512$. Note that this experiment demands the retraining of f_{left} and f_{right} for each d. After training, the S-MARQUES and S-ISRI-OCR instances were individually reconstructed, and then accuracy and time processing were measured.

4.2.3 *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 was written in Python 3.5 using Tensorflow for training and inference, and OpenCV for basic image manipulation. The code, pre-trained models, and datasets are publicly available at https://github.com/thiagopx/deeprec-cvpr20.

4.3 RESULTS AND DISCUSSION

4.3.1 *Experiment 1: Single-page Reconstruction*

Method	S-MARQUES ∪ S-ISRI-OCR	S-MARQUES	S-ISRI-OCR
DEEPREC-ML	93.71 ± 11.60	93.14 ± 12.93	$\mathbf{95.39 \pm 6.02}$
DEEPREC-CL [62, 63]	$\mathbf{96.28 \pm 5.15}$	$\mathbf{96.78 \pm 4.44}$	94.78 ± 6.78
Paixão et al. [61]	74.85 ± 22.50	71.85 ± 23.14	83.83 ± 18.12
Marques and Freitas [53]	23.90 ± 17.95	29.18 ± 17.43	8.05 ± 6.60

Table 1: Single-page reconstruction performance: average accuracy ± standard deviation (%). The highest average value in each column is highlighted in bold.

A comparison with the literature on 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 DEEPREC-CL. The box-plots in Figure 28 show the accuracy distribution obtained with DEEPREC-ML and DEEPREC-CL for single-page reconstruction. Likewise DEEPREC-CL, we also observe for DEEPREC-ML that vertical shifts affect only the S-ISRI-OCR's accuracy since the shreds in S-MARQUES are practically aligned (vertical direction). The methods did not present a significant difference in accuracy for the dataset S-ISRI-OCR. For S-MARQUES, however, DEEPREC-CL slightly outperformed DEEPREC-ML: the latter – in its default configuration (vertical shift "on") – yielded accuracy of $93.14 \pm 12.88\%$ (arithmetic mean ± standard deviation), while DEEPREC-CL achieved $96.78 \pm 4.44\%$. The higher variability in DEEPREC-ML is mainly ex-

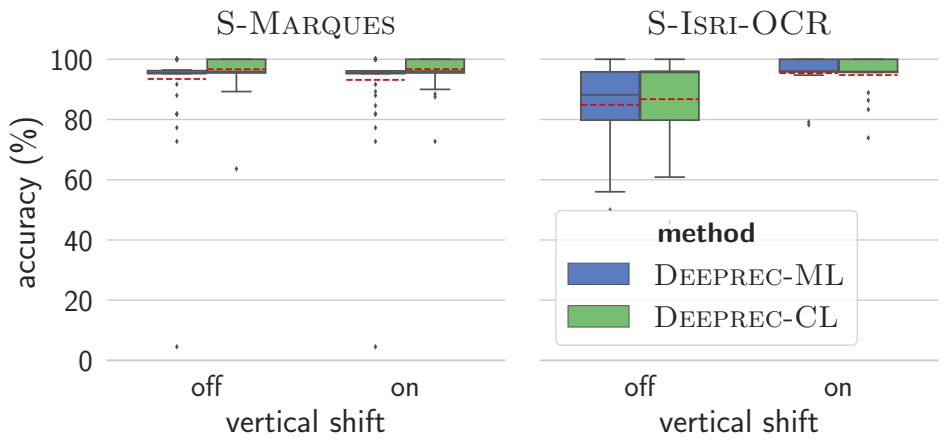


Figure 28: Accuracy distribution for single-page reconstruction with the proposed and DEEPREC-CL methods. Accuracies are calculated document-wise and the average values are represented by the red dashed lines.

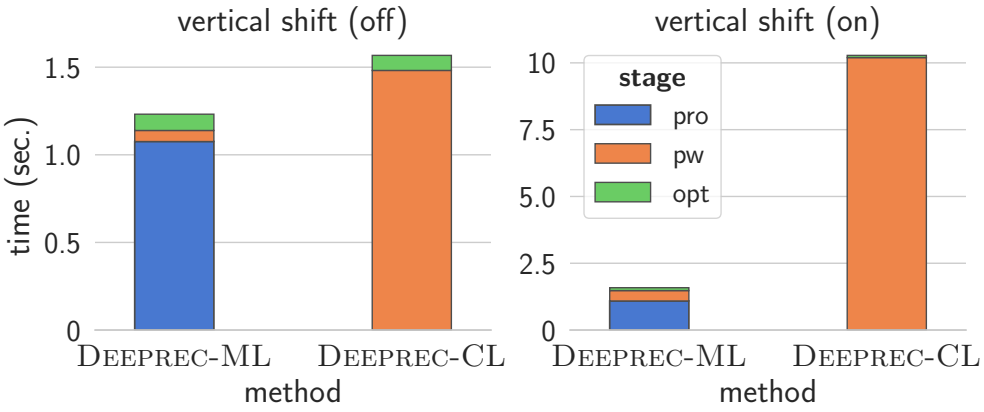


Figure 29: Time performance for single-page reconstruction. The stacked bars represent the average elapsed time for each reconstruction stage: projection (pro), pairwise compatibility evaluation (pw), and optimization process (opt).

plained 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 29. The stacked bars represent the average elapsed time in seconds (s) for each reconstruction stage: projection (pro), pairwise compatibility evaluation (pw), and optimization process (opt). With vertical shift disabled (left chart), DEEPREC-ML spent much more time producing the embeddings (1.075s) than in pairwise evaluation (0.063s) and optimization (0.092s). Although DEEPREC-CL does not have the cost of embedding projection, pairwise evaluation took 1.481s, about 23 times the time elapsed in the same stage for DEEPREC-ML. This difference becomes more significant (in absolute values) when the number of pairwise evaluations increases, as can be seen with the enabling of vertical shifts (right chart). In this scenario, pairwise evaluation took 0.389s

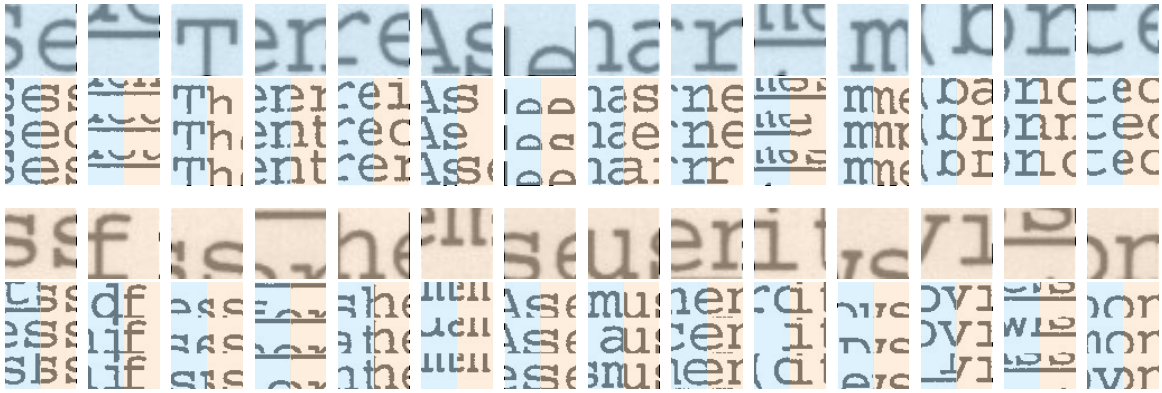


Figure 30: Local samples nearest neighbors. In the top row, the largest square is the "query" sample (before binarization) followed, below, by its binary version and its three nearest neighbors side-by-side (with the closest in the top row). The blue and orange samples were projected by f_{right} and f_{left}, respectively. The bottom row shows some examples in which the "query" is projected by the f_{left} instead.

in our method, against the 10.197s spent in DEEPREC-CL ($\approx$ 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 DEEPREC-CL would drop from 94.77 to 86.74% in S-ISRI-OCR.

Finally, we provide an insight into what the embedding space using DEEPREC-ML might look like by showing a local sample and its three nearest neighbors. As shown in Figure 30, 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 Appendix A).

4.3.2 *Experiment 2: Multi-page Reconstruction*

For multi-page reconstruction, DEEPREC-ML achieved 94.81 and 97.22% of accuracy for S-MARQUES and S-ISRI-OCR, respectively, whereas DEEPREC-CL achieved 97.08 and 95.24%. Overall, both methods yielded high-quality reconstructions with a low difference in accuracy (approx. $\pm$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 differently, as evidenced in Figure 31. The left chart shows the average elapsed time of each stage to process the 505 shreds of S-ISRI-OCR. In this context, with a larger number of shreds, the optimization cost became negligible when compared to the time required for pairwise evaluation. Remarkably, DEEPREC-CL demanded more than 80 minutes to complete the evaluation stage, 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

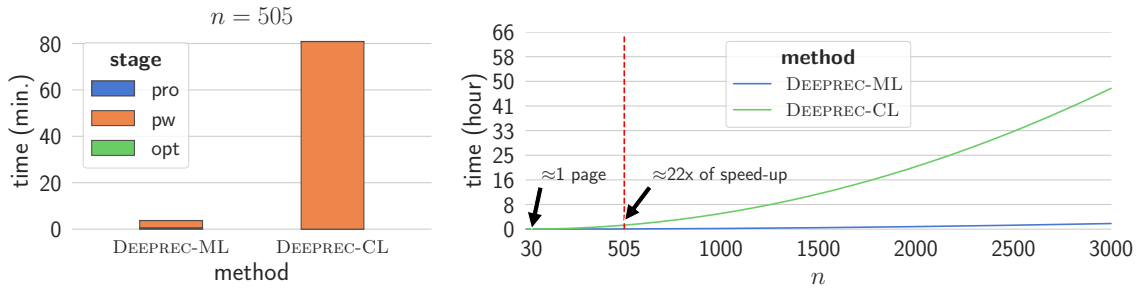


Figure 31: Time performance for multi-page reconstruction. Left: the time demanded in each stage to reconstruct S-ISRI-OCR entirely ($n = 505$ shreds). Right: predicted processing time in function of the number of shreds.

of shreds (n). The last n value was 30K, which corresponds nearly to 100 pages. Viewed comparatively, the growth of the proposed method's curve (in blue) seems to be linear, although pairwise evaluation time (not the number inferences) grows quadratically with n.

Equation 11 describes the machine-independent speed-up ratio between DEEPREC-ML and DEEPREC-CL:

$$\begin{aligned} \text{speed-ratio} &= \frac{n(n-1)t_{inf}}{2n \cdot t_{inf} + n(n-1)t_{dist}} \\ &= \frac{1}{2/(n-1) + t_{dist}/t_{inf}}, \end{aligned} \tag{11}$$

where t_{inf} and t_{dist} stand for, respectively, inference time and distance computation time (used to compute DEEPREC-ML). The assumption of our method is that $t_{inf} \gg t_{dist}$, *i.e.*, $t_{inf}/t_{dist} \gg 1$. For n values such that $t_{dist}/t_{inf} \ll 2/(n-1)$, the speed-up can be approximated by $(n-1)/2 = O(n)$. For $n \to +\infty$, in its turn, the speed-ratio tends to the constant t_{inf}/t_{dist}, which sets a theoretical limit. In practice, the greater the number of shreds are, the higher the speed-up ratio is.

4.3.3 *Experiment 3: Sensitivity Analysis*

Figure 32 shows, for single-page reconstruction, how accuracy and time processing (mean values over pages) are 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 S-MARQUES and S-ISRI-OCR, 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 in accuracy and processing time by focusing on small values of d, which will be better investigated in future work.

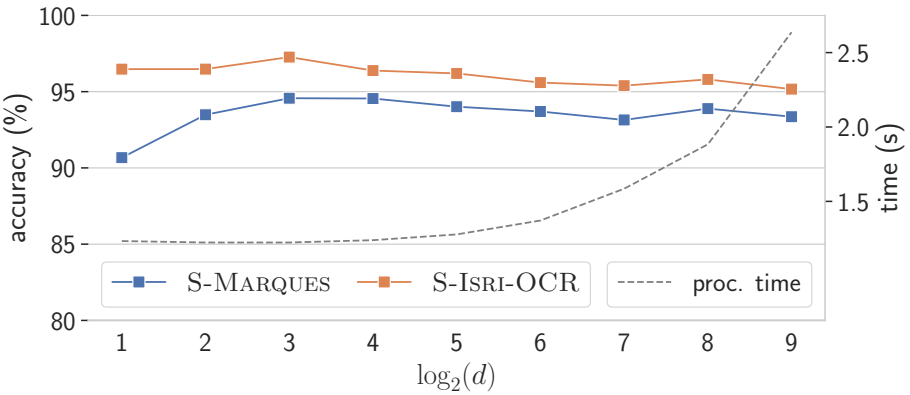


Figure 32: Sensitivity analysis w.r.t. embeddings dimension (d). The best accuracy was observed for $d = 8$: 94.57 and 97.27% for S-MARQUES and S-ISRI-OCR, respectively. This reduced embedded size yielded a reduction of 23% on processing time.

4.4 CONCLUDING REMARKS

This chapter addressed the improvement of scalability (time performance) for reconstruction of shredded text documents. The core of the proposal is the metric learning-based approach for shreds' compatibility evaluation in which the number of inferences scales linearly rather than quadratically [62, 63] with the number of shreds of the reconstruction instance. As the alternative deep learning method, this new 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 manual 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. Experiments were also conducted in a more realistic scenario: 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 DEEPREC-CL (*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.

Despite the achieved results, we showed that the current metric learning loses accuracy when dealing with pictorial content instead of text. Therefore, future work will investigate how to explore more rich graphic content in a self-supervised way to enhance the training of the models.

5

A HUMAN-IN-THE-LOOP RECONSTRUCTION FRAMEWORK

The previous chapters discussed the two deep learning approaches for automatic reconstruction developed in the context of this thesis. Despite the remarkable results, it should be considered that the proposed models may fail in coherently measuring the fitness of the shreds, limiting the reconstruction accuracy. A particular way to obtain better solutions is to introduce active human supervision (semi-automatic reconstruction). Inspired by the active learning literature [21, 80, 86], the reconstruction process can be modeled as a loop where, in each iteration, the human is queried to provide inputs, and a new solution is attained.

This chapter discusses our human-in-the-loop (HIL) framework for reconstruction of mechanically strip-shredded documents. The core of the framework is the recommender module, which is responsible for selecting pairs of adjacent shreds of a solution for annotation. The conducted experiments considered different workloads (number of pairs to be annotated) and different numbers of loop iterations.

In summary, the main contributions covered in this chapter are:

- A human-in-the-loop recommendation-based framework (or simply HIL framework) for the reconstruction of strip-shredded documents;
- Four query strategies for recommending pairs of shreds to be annotated;
- A novel experimental methodology that assesses the impact of human labor on the quality of the reconstructions: results have shown that a user workload of 25% can lead to more than 4 p.p. of accuracy improvement ($>$ 40% of error reduction) on the deep learning methods.

The chapter is organized into four main sections addressing the (i) HIL framework, (ii) the experimental assessment, (iii) the discussion of the results, and (iii) the concluding remarks.

5.1 HIL FRAMEWORK

The proposed HIL reconstruction framework works iteratively, as illustrated in Figure 33 (the superscript $k \geqslant 0$ indicates the current iteration). The automatic part of the framework (above the dashed line) comprises three elements: a cost matrix $\mathbf{\Phi}^{(k)}$, an optimization solver that computes a solution $\pi_s^{(k)}$, and the recommender module, which determines a query set $Q^{(k)}$ comprising the pairs of adjacent shreds of $\pi_s^{(k)}$ to be analyzed. The human role represented below the dashed line is to split the query into positive ($Q_+^{(k)}$)

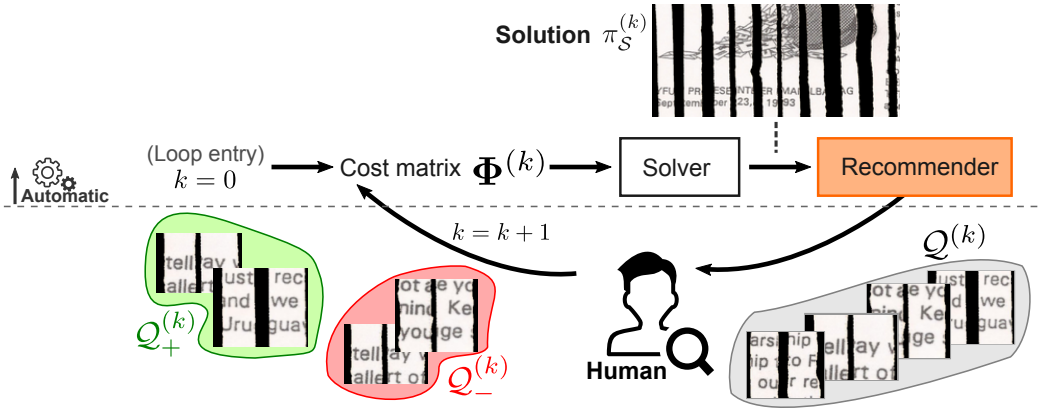


Figure 33: Overview of the proposed HIL reconstruction framework. The automatic part locates above the dashed line. A reconstruction loop works as follows. First, the cost function $\mathbf{\Phi}^{(k)}$ is computed/edited. Then an optimization solver produces a solution $\pi_S^{(k)}$ based on such costs. Next, the recommender module determines a query set comprising pairs of shreds for annotation (positive or negative). Finally, these labels are used to edit the cost matrix. The process repeats until a predefined number of iterations is reached.

and negative ($\mathcal{Q}_-^{(k)}$) pairs. The positive pairs are those to be grouped (locked), and the negatives are those to be set apart (forbidden). Initially, when $k = 0$, the costs $\mathbf{\Phi}^{(0)}$ are fully defined by a third-party procedure, usually the cost computation module of the reconstruction algorithms. From this perspective, $\pi_S^{(0)}$ represents the solution obtained without interaction with the user. For the next iteration, the cost matrix is edited to reflect the human intervention according to Equation (12):

$$\Phi_{i,j}^{(k+1)} = \begin{cases} -\infty, & \text{if } (s_i, s_j) \in \mathcal{Q}_+^{(k)} \\ +\infty, & \text{if } (s_i, s_j) \in \mathcal{Q}_-^{(k)} \\ +\infty, & \text{if } \exists j'[(s_i, s_{j'}) \in \mathcal{Q}_+^{(k)}] \\ +\infty, & \text{if } \exists i'[(s_{i'}, s_j) \in \mathcal{Q}_+^{(k)}] \\ \Phi_{i,j}^{(k)}, & \text{otherwise.} \end{cases} \tag{12}$$

The first case represents the "lock" operation, while the following three cases represent the "forbid" operation. Cases three and four check whether there exists a shred out of the pair that should be locked to s_i or s_j, implying that (s_i, s_j) is negative. If none of the above situations is met, nothing can be said about the pair. Therefore, the cost remains unaltered for the next iteration.

Ideally, the annotation effort (workload) for improving a solution should be minimal, making the query strategy a critical element of the framework. We conjecture that it is more relevant to repel the negative pairs than lock the positives. Therefore, the developed strategies focus on finding potentially wrong pairs to compose the queries. These strategies are described in the rest of this section. For simplicity of notation, the iteration superscript k will be omitted.

5.1.1 *Optimality-based Strategies*

The base (deterministic) algorithm for the optimality-based query strategies (OPT-R and OPT-RL) is presented as follows. Let $\mathcal{P}_S = \{(s_{\pi_1}, s_{\pi_2}), (s_{\pi_2}, s_{\pi_3}), \ldots, (s_{\pi_{n-1}}, s_{\pi_n})\}$ denote the consecutive pairs of a solution π_S:

1. Predict (using some criteria) each pair $(s_{\pi_i}, s_{\pi_{i+1}}) \in \mathcal{P}_S$ as positive or negative following strictly the left-to-right order ($i = 1, 2, \ldots, n-1$);
2. If the current pair was predicted negative, insert it into the beginning of a list L;
3. If positive, append it onto the end of L;
4. Finally, select the first n_{query} pairs of L for user annotation (*i.e.*, verify whether each pair is in fact negative).

In this algorithm, the prediction step relies on two desirable properties of a well-designed cost function:

- Right-shred optimality: $\phi(s_i, s_{i+1}) < \phi(s_i, s_{j'})$, for all $j' \neq i+1$;
- Left-shred optimality: $\phi(s_{j-1}, s_j) < \phi(s_{i'}, s_j)$, for all $i' \neq j-1$.

Right-shred optimality states that, given a shred s_i, the correct right shred (s_{i+1}) should yield the lowest cost when compared to any other possible matches on the right side. Similarly, in left-shred optimality, given a shred s_j, the correct left neighbor (s_{j-1}) should be the minimum-cost candidate. Two query strategies are derived from incorporating these properties into the base algorithm. The first, OPT-R, predicts a pair as positive iff it fulfills the right-shred optimality property. The second, OPT-RL, is more restrictive since it assumes that a pair is positive iff both left- and right-shred optimality properties are met.

5.1.2 *Uncertainty-based Strategies*

The two proposed query uncertainty-based strategies (UNC-R and UNC-RL) correlate potentially negative pairs with a high degree of uncertainty, which, in this paper, relies on the entropy measure [88]. The (deterministic) algorithm that implements both strategies consists in sorting the pairs of the solution by decreasing order of uncertainty and selecting the first n_{query} pairs for user annotation.

In our framework, the entropy measure quantifies the uncertainty degree in predicting the true neighbor of a shred based on the relative costs to the other shreds. The more uniform the relative costs are (high entropy), the less certain the true neighbor prediction is. Two probability distributions are considered for entropy calculations. The first, $\Pr_r(s_j \mid s_i)$, defines the probability for a right-shred s_j conditioned on a left-shred s_i as left shred. Analogously, fixing a right shred s_j, $\Pr_l(s_i \mid s_j)$ defines the probabilities for candidate

left shreds s_i. The distributions rely on the softmax function, as seen in Equations (13) and (14):

$$\Pr_r(s_j \mid s_i) = \frac{\exp(-\Phi'_{i,j})}{\sum_{j'=1}^{n} \exp(-\Phi'_{i,j'})} \tag{13}$$

$$\Pr_l(s_i \mid s_j) = \frac{\exp(-\Phi'_{i,j})}{\sum_{i'=1}^{n} \exp(-\Phi'_{i',j})}, \tag{14}$$

where

$$\boldsymbol{\Phi}' = \lambda \frac{\boldsymbol{\Phi}}{\max_{\tilde{\infty}}(\boldsymbol{\Phi})}. \tag{15}$$

The denominator of Equation (15) is defined as the maximum non-infinity ($\max_{\tilde{\infty}}$) cost value of $\boldsymbol{\Phi}$, therefore, $\Phi'_{i,j} \mapsto [0, \lambda]$ for non-infinite values. This is convenient to establish a common input range for the softmax function since it yields different values for scaled inputs.

Based on these distributions, the entropies $E_r(s_i)$ and $E_l(s_j)$ are calculated by Equations (16) and (17), which denote the uncertainty of which shred is the right neighbor of s_i, and of which shred is the left neighbor of s_j, respectively.

$$E_r(s_i) = -\sum_{j'=1}^{n} \Pr_r(s_{j'} \mid s_i) \log \Pr_r(s_{j'} \mid s_i) \tag{16}$$

$$E_l(s_j) = -\sum_{i'=1}^{n} \Pr_l(s_{i'} \mid s_j) \log \Pr_l(s_{i'} \mid s_j) \tag{17}$$

Finally, for the first strategy (Unc-R), the uncertainty for the pair (s_i, s_j) considers only the entropy $E_r(s_i)$, while for the second (Unc-RL), the uncertainty is calculated as $E_r(s_i) + E_l(s_j)$.

5.2 EXPERIMENTAL ASSESSMENT

Two experiments were conducted to assess the impact of incorporating human interaction in the reconstruction process. The *workload experiment* investigates, for a single iteration (until $k = 1$), the effect of increasing the workload (*i.e.*, the number of pairs of shreds queried to the user) upon the accuracy of the solution. The *multi-iteration experiment* investigates how splitting the workload into iterations can affect the accuracy of the solutions.

Initial costs ($\boldsymbol{\Phi}^{(0)}$) and solution ($\pi_S^{(0)}$) are obtained according to the the deep learning methods here named as Deeprec-ML in [65] and Deeprec-CL [63]. Following the eval-

uation protocol in [63], the models are trained[1] on the document collections ISRI-OCR (a subset of the ISRI-Tk OCR collection [58]) and CDIP (subset of RVL-CDIP [34]), and evaluated on the shredded datasets S-MARQUES [53], S-ISRI-OCR [61], and S-CDIP [63]. To reflect a more realistic scenario, we also adopted a cross-database approach where the models trained on ISRI-OCR are used to reconstruct S-CDIP, while CDIP is used to reconstruct both S-MARQUES and S-ISRI-OCR. Furthermore, the experiments address only the multi-page reconstruction scenario given that, in a real-world context, shreds of different pages/documents are mixed and the system is not aware of which page each shred belongs to. The particularities of the experiments are detailed next.

5.2.1 *Experiments*

EXPERIMENT 1: WORKLOAD EXPERIMENT. The intuition behind this experiment is that accuracy should improve as the user is demanded to analyze more pairs of shreds, and that, ideally, significant improvement should be achieved with low human effort. From the $n-1$ pairs of a solution, the human effort is quantified by number of pairs to be annotated $n_{query} = \alpha_{load}\,(n-1)$, where the factor α_{load} denotes the workload. That been said, this experiment consisted of running a single iteration of the framework for each $\alpha_{load} = 0.1$, 0.15, 0.2, and 0.25 and for each one of the query strategies defined in Section 5.1: OPT-R, OPT-RL , UNC-R , and UNC-RL. DEEPREC-ML and DEEPREC-CL were used to compute initial costs and solutions. In addition to the proposed query strategies, a baseline strategy that randomly selects pairs for human annotation was evaluated. Note that this is equivalent (in terms of performance) to the manual selection in [76] (discussed in the introduction) since there are no criteria guiding the selection. Since the training of the deep models (which plays the role of cost functions) is non-deterministic, the experiment was run with five different models generated for each reconstruction method, thus enabling a more robust analysis.

EXPERIMENT 2: MULTI-ITERATION EXPERIMENT. This experiment hypothesizes that splitting the workload into a few iterations may yield a faster improvement rate. This analysis focused on the state-of-the-art reconstruction method DEEPREC-ML– considering both accuracy and time performance – and on the query strategy OPT-R, which achieved the most consistent performance for DEEPREC-ML in the workload experiment, as seen in Figure 36 (discussed in the next section). The performance was evaluated for $n_{iter} =$ 1, 2, and 3 iterations, being n_{query}/n_{iter} pairs analyzed in each iteration, $n_{query} = \alpha_{load}\,(n-1)$. Again, we tested with $\alpha_{load} = 0.1$, 0.15, 0.2, and 0.25. The analysis focus on the metric learning-based reconstruction algorithm in [65]. As in the workload experiment, we report the average performance of five runs in each framework configuration.

1 For implementation details of each deep learning method (samples extraction, architectures, training parameters, etc.), the reader is referred to the respective works.

5.2.2 *Implementation Details.*

The human interaction in the experiments was simulated by using the ground-truth order available for the test datasets. We assume a "perfect" annotator/oracle, which means that the pairs of shreds are always correctly labeled. To enable the solver execution, infinite costs (Equation (12)) were replaced by real values following the implementation in [33]. For experiments with DEEPREC-CL, we convert the initial compatibilities into costs by doing $\mathbf{\Phi} = \max(\mathbf{\Gamma}) - \mathbf{\Gamma}$, where $\max(\mathbf{\Gamma})$ is the maximum value (excluding the diagonal) of the compatibility matrix $\mathbf{\Gamma}$ [63]. Finally, the λ normalization factor in Equation (15) was set to 100. Preliminary empirical investigation showed that $\lambda \in [50, 150]$ results in reasonable error detection.

5.2.3 *Experimental Platform*

The experiments were conducted in an Intel Core i7-4770 CPU @ 3.40GHz with 16GB of RAM running Linux Ubuntu 18.04. A Nvidia TITAN Xp GPU (12GB) was used for fast deep learning training/inference. The software was implemented in Python 3.6, being Tensorflow 2.6 used for training/inference of the models, and OpenCV leveraged for basic image processing[2].

	DEEPREC-ML		DEEPREC-CL	
Dataset	Avg.	Med.	Avg.	Med.
S-MARQUES	93.51 ± 0.49	93.50	$\mathbf{96.74 \pm 0.50}$	97.08
S-ISRI-OCR	$\mathbf{96.04 \pm 1.69}$	**96.83**	95.76 ± 0.86	96.24
S-CDIP	$\mathbf{90.39 \pm 0.74}$	**90.18**	81.29 ± 11.73	88.26

Table 2: Multi-page reconstruction performance: average accuracy $\pm$ standard deviation (%) and median for five executions. The highest average/median value in each row is highlighted in bold.

5.3 RESULTS AND DISCUSSION

The results for the multi-page reconstruction without human intervention for five executions are shown in Table 2. Comparatively, DEEPREC-CL performed better only for the dataset S-MARQUES, which represents a less realistic scenario compared to the other two datasets [63]. The most significant performance discrepancy, considering the average accuracy, was observed for S-CDIP, although the median difference is less than 2 p.p.. The lowest accuracies not only reinforce that S-CDIP is the most challenging dataset but also inform us that it has the most significant margin for improvement. The following subsections discuss the impact of our HIL framework in producing better reconstructions for the three datasets.

2 The code, pre-trained models, and datasets will soon be publicly available.

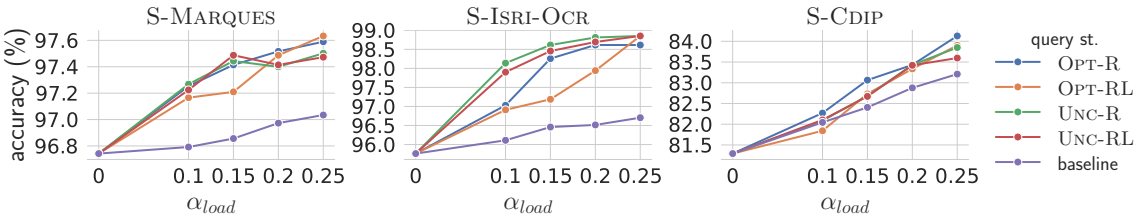


Figure 34: Reconstruction accuracy w.r.t. workload (DEEPREC-CL). Each curve is associated with a distinct query strategy.

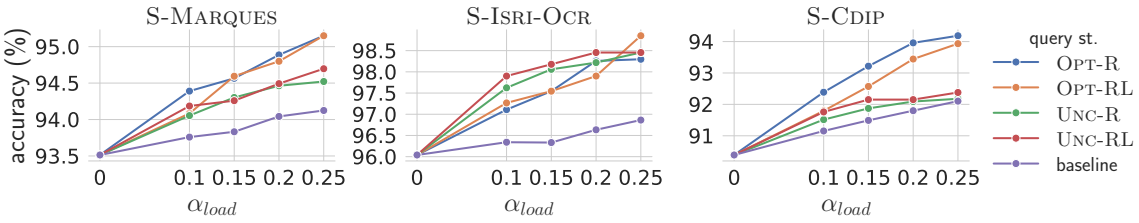


Figure 35: Reconstruction accuracy w.r.t. workload (DEEPREC-ML). Each curve is associated with a distinct query strategy.

5.3.1 *Experiment 1: Workload Experiment*

The results for the workload experiment are shown in Figures 34 and 35. As expected, the accuracy increases roughly linearly with the human workload, even for the random baseline. Nonetheless, it is noticeable that the proposed query strategies outperform the baseline [76], which means that our simple selection criteria yield better results than random selection. Moreover, the results show very similar performance for the uncertainty-based strategies (UNC-R and UNC-RL): the maximum accuracy difference was nearly 0.22 p.p. (on average).

Dataset	OPT-R	OPT-RL	UNC-R	UNC-RL
S-MARQUES	$\mathbf{1.64 \pm 0.28}$	$\mathbf{1.64 \pm 0.34}$	1.01 ± 0.35	1.18 ± 0.22
S-ISRI-OCR	2.26 ± 1.54	$\mathbf{2.81 \pm 1.25}$	2.42 ± 0.90	2.42 ± 0.90
S-CDIP	$\mathbf{3.80 \pm 0.53}$	3.54 ± 0.55	1.79 ± 0.51	1.99 ± 0.18

Table 3: Accuracy improvement w.r.t. the query strategies (DEEPREC-ML, $\alpha_{load} = 0.25$): average accuracy difference $\pm$ standard deviation (p.p.). The highest value in each row is highlighted in bold. OPT-R yielded an increase of ≈ 3.80 p.p for S-CDIP ($\approx 39.50\%$ of error reduction)

Particularly for the DEEPREC-ML reconstruction method, the uncertainty-based strategies achieved competitive performance with the optimality-based strategies for the dataset S-ISRI-OCR (20 documents), while for larger datasets – S-MARQUES (60 documents) and S-CDIP (100 documents) – OPT-R/OPT-RL significantly outperforms UNC-R/UNC-RL. This can be seen in Table 3, which displays the accuracy improvement for each dataset ($\alpha_{load} = 0.25$). The performance difference becomes more noticeable with the increase of the work-

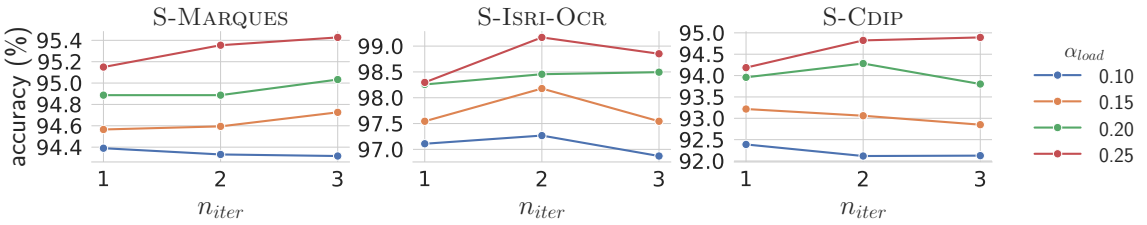


Figure 36: Reconstruction accuracy w.r.t. the number of iterations (DEEPREC-ML, OPT-R). Each curve represents a different workload (α_{load}).

Dataset	1	2	3
S-MARQUES	1.64 ± 0.28	1.84 ± 0.43	$\mathbf{1.91 \pm 0.24}$
S-ISRI-OCR	2.26 ± 1.54	$\mathbf{3.13 \pm 1.50}$	2.81 ± 1.64
S-CDIP	3.80 ± 0.53	4.43 ± 0.88	$\mathbf{4.50 \pm 0.91}$

Table 4: Accuracy improvement w.r.t. the number of iterations (DEEPREC-ML, $\alpha_{load} = 0.25$, OPT-R): average accuracy difference $\pm$ standard deviation (p.p.). The highest value in each row is highlighted in bold. Two iterations yielded an increase of ≈ 4.43 p.p for S-CDIP ($\approx 46.10\%$ of error reduction).

load, which suggests that the optimality-based strategies are more effective for DEEPREC-ML. OPT-R and OPT-RL performed similarly to each other, being the maximum accuracy difference (on average) between them of ≈ 0.5 p.p.. Remarkably, OPT-R was able to increase the original solution accuracy of the S-CDIP dataset on ≈ 3.80 p.p. for $\alpha_{load} = 0.25$: 87 pairs were corrected from a total of 220 mistakes ($\approx 39.50\%$ of error reduction).

For DEEPREC-CL, the uncertainty- and optimality-based strategies give closer performance compared to the results for DEEPREC-ML. We verified that DEEPREC-ML yielded more uniform values for the measures $E_r(s_i)$ and $E_l(s_j)$ (Equations (16) and (17)) which implies an uncertainty in discriminate potentially negative pairs. This fact serves to emphasize a relationship between the performance of the strategies and the cost functions.

5.3.2 *Experiment 2: Multi-iteration Experiment*

The results for the multi-iteration experiment where the user workload (effort) is split into iterations are shown in Figure 36. As commented in the previous section, this experiment focused on the state-of-the-art method DEEPREC-ML with the query strategy OPT-R given its superior performance in the previous experiments DEEPREC-ML. At first glance, the most consistent improvement was obtained by increasing the number of iterations (n_{iter}) from one to two. Nonetheless, it can be observed that adopting $n_{iter} \geqslant 2$ is interesting when compared to a single iteration for higher workload values. The results for $\alpha_{load} = 0.25$ (the red curves in Figure 36) is presented in Table 4. When comparing the top accuracies (highlighted in the table) with those obtained for $n_{iter} = 1$, we see that the maximum improvement is 0.70 p.p.. This value drops to 0.08 p.p. when n_{iter} is increased

from two to three. This reinforces the decision of not running more than two iterations, including the fact that the higher the number of iterations, the higher the computation burden of multiple solver runs is (a single run can take ≈ 3 minutes for S-CDIP). Compared to Table 2, the accuracy on S-CDIP for two iterations increased around 4.43 p.p. (≈ 46.10% of error reduction).

5.4 CONCLUDING REMARKS

This chapter investigated the impact on the performance of introducing a human user in the process of reconstructing shredded documents. We proposed a human-in-the-loop reconstruction framework where the user is queried to verify whether every two adjacent shreds in the solution are, in fact, adjacent in the original document. Four query strategies were proposed for the recommender module to select pairs of shreds for human verification.

The workload experiment showed that our framework consistently improves solutions as the user takes more part in the process and that the accuracy increases roughly linearly with the human workload. Furthermore, the proposed query strategies outperformed the baseline strategy [76], and, among the proposed strategies, those based on optimality criteria yielded the most consistent performance. This can be due to the cost assignment that results in high uncertainty when the correct pair meets the optimality criteria (*e. g.*, it is assigned the lowest cost). Alternatively, low uncertainty may arise in scenarios where the pairs are wrong. In particular, the OPT-R strategy increased by nearly 3.80 p.p. the accuracy for the S-CDIP dataset (≈ 39.50% of error reduction).

We also investigated whether the solutions could be improved by splitting the workload into iterations. The general conclusion is that the increase in iterations is effective for higher workloads. Furthermore, considering the evaluated scenarios, we concluded that two iterations are reasonable for the framework. For such a value, the OPT-R yielded an increase of nearly 4.43 p.p. in the accuracy on S-CDIP (≈ 46.10% of error reduction).

Future work could investigate using an ensemble of query strategies to provide more relevant queries to the user. Also, the framework could be adapted for cross-cut documents. However, it requires the availability of realistic datasets as there are available for strip-shredded documents. Finally, new query strategies driven to our problem should be investigated/adapted with a more in-depth review of literature in active learning.

6

CONCLUDING REMARKS AND FUTURE WORK

This thesis presented a corpus of contributions for (semi-)automatic reconstruction of mechanically-shredded documents. Besides the relevance of the topic, this research was motivated by the need of the literature for methods capable to deal with real shredded data. Our effort was initially towards robust compatibility evaluation between shreds for fully automatic reconstruction (Chapters 3 and 4), so that the optimization process might yield improved reconstructions. Once the reconstructed documents were available, they could then be analyzed – by document examiners, for example – as they seek relevant/sensitive information. In a second moment, it was investigated the introduction of the human as part of the reconstruction process (Chapter 5).

Chapter 3 described our deep learning approach based on classification (DEEPREC-CL) and its application on multi-page reconstruction. Results have shown accuracy superior to 90% for S-CDIP, the most challenging dataset comprising 2,292 shreds. In Chapter 4, it was presented a metric learning reconstruction approach (DEEPREC-ML) where the number of network inferences linearly rather than quadratically as in DEEPREC-CL. The theoretical result in Section 4.3.2 shows that the speed-up can grow linearly with the number of input shreds. For the tested scenarios, it can be highlighted a speed-up of ≈ 22 times for the 505 shreds of S-ISRI-OCR.

Later, in Chapter 5, it was proposed an interactive reconstruction framework (HIL framework) inspired in the field of Active Learning (AL) that takes valuable human feedback to improve solutions. The focus of our contribution was on the recommender module responsible for automatically selecting data to be analyzed by the user. Results have shown that $> 40\%$ of error reduction can be achieved in certain cases given a user workload of 25%, *i.e.*, by looking at nearly 1/4 of the (pairs of) shreds providing positive/negative labels for them. Although the tests were performed with the deep learning methods, the framework is generic and can be used with different cost functions and solvers.

Additional contributions include:

- An experimental methodology for multi-page reconstruction;
- A novel experimental methodology that assesses the impact of human labor on the quality of the reconstructions;
- A dataset comprising 120 shredded documents, totalling 2,797 shreds.

In future work, it would be helpful to extend the proposed methodologies to cross-cut documents. This is not straightforward since it adds complexity to the compatibility evaluation and to the optimization process. For compatibility evaluation, it requires models

that perform well with less useful content given the reduced dimensions of the shreds and that is fast enough to couple with the increase in the number of shreds, which also affects the time performance of the optimizer. Concerning human-assisted reconstruction, a promising direction is the development/adaptation of query strategies to the reconstruction application, including the use of ensemble of strategies for more valuable user feedback. Finally, from a generalization perspective, there are correlated problems that should benefit from our findings. Currently, in the literature, it could be highlighted the generic problem of solving jigsaw puzzles with eroded borders [13, 46, 68, 79] and the application of reassembling fragments of ancient papyrus [1, 71].

Part I

APPENDIX

A

APPENDIX: AN ASSYMETRIC METRIC-LEARNING APPROACH

A.1 LOCAL SAMPLES NEAREST NEIGHBORS

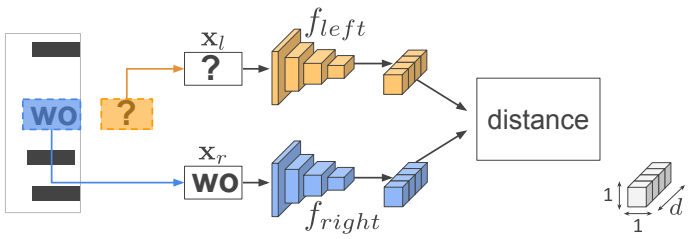


Figure 37: Querying x_l samples by fixing x_r.

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 37, one can select x_r as a query sample and try to recover the top-1 x_l's, *i.e.*, that the sample of the left boundary which minimizes the distance to the anchor in the embedding space. Figure 30 (in Chapter 4) shown some queries for both x_r and 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 38: 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.

A.2 RECONSTRUCTION OF S-ISRI-OCR

The dataset S-Isri-OCR comprises 20 single-page documents, totaling 505 shreds. Figure 39 shows the reconstruction of the entire S-Isri-OCR 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 39: Reconstruction of S-Isri-OCR. The generated image was split into 4 parts for better visualization.

A.3 EMBEDDING SPACE

Figure 24 (in Chapter 4) illustrates the embedding space onto which the local samples are projected. For a more concrete view of this space, four charts (Figures 40 to 43) 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 [52, 97]. 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.

A.3.1 *Case 1*

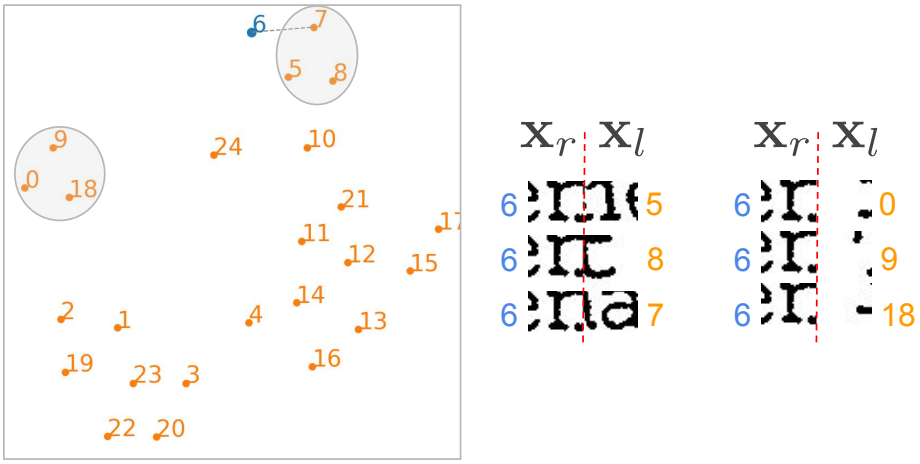


Figure 40: Case 1.

In Figure 40, 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.

A.3.2 *Case 2*

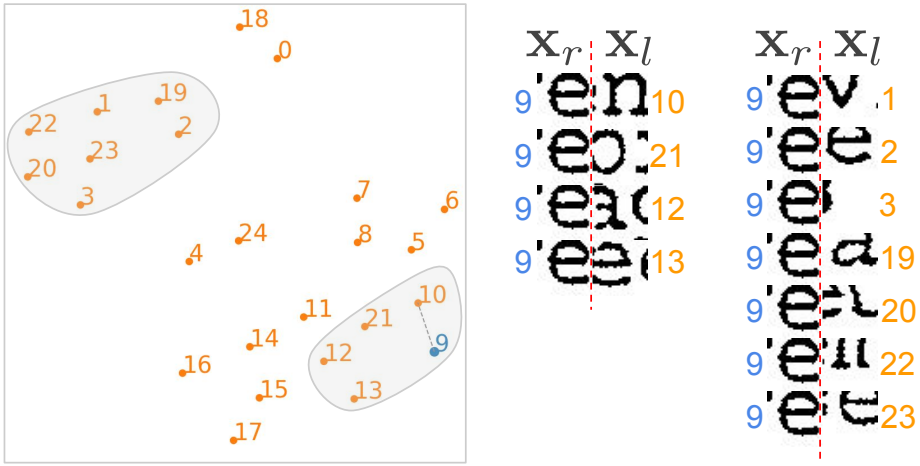


Figure 41: Case 2.

In Figure 41, 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 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).

A.3.3 *Case 3*

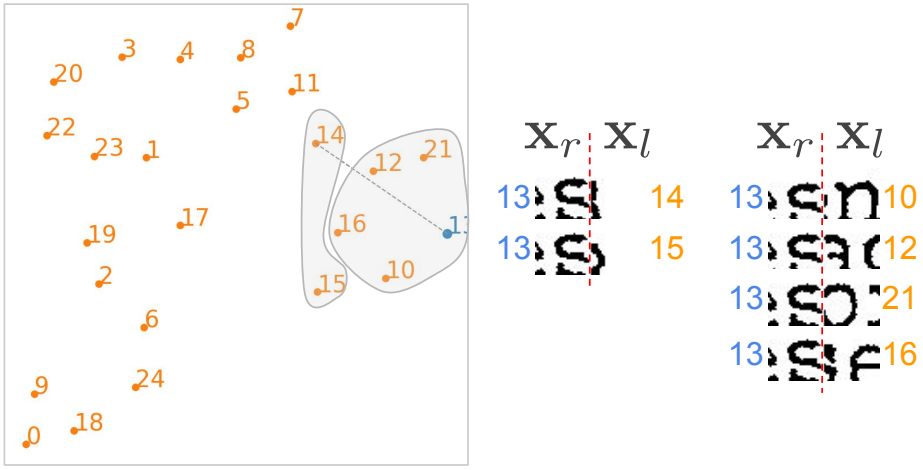


Figure 42: Case 3.

The third case, illustrated in Figure 42, depicts a situation where a couple of matchings are better evaluated than the corrected one: $(13, 14)$. 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 x_l (and analogously in 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 x_l is practically blank.

A.3.4 *Case 4*

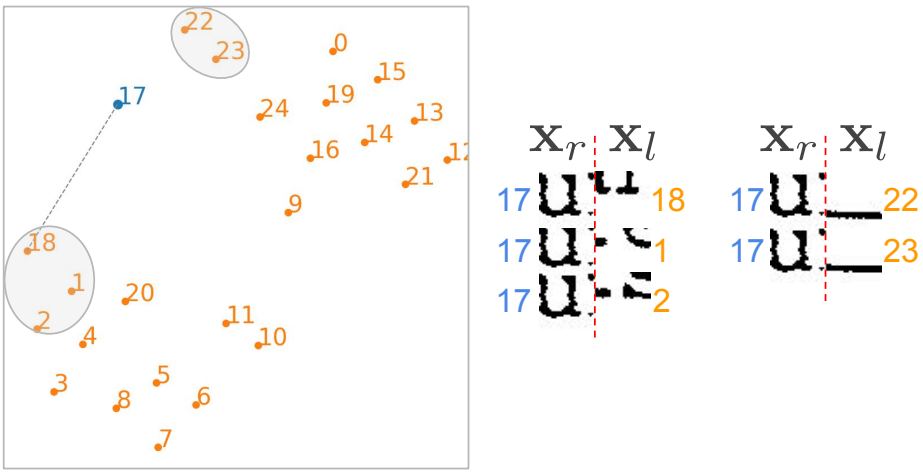


Figure 43: Case 4.

The last selected case, Figure 43, emphasizes the relevance of the vertical alignment stage of the metric-learning approach (Section 3.1.2.1). 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 x_l.

A.4 SENSITIVITY ANALYSIS W.R.T. SAMPLE SIZE

As in classification-based approach, this method use small samples (32×32) to explore features at text line (local) level since we assume weak feature correlation across text lines. In [63], we observed that the reconstruction accuracy decreases for larger samples. This is also verified in the proposed metric learning approach when the sample height (s_y) is increased, as seen in Figure 44.

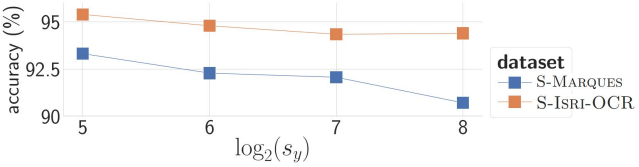


Figure 44: Reconstruction accuracy w.r.t. to the sample height (s_y).

A.5 STATISTICAL TEST

Considering a threshold of 5%, the proposed method was statistically equivalent to DEEPREC-CL in S-ISRI-OCR and superior to Paixão et al. [61] in both datasets. Table 5 shows the p-values of the page-wise paired t-test.

	S-MARQUES ∪ S-ISRI-OCR	S-MARQUES	S-ISRI-OCR
Deeprec-ML vs. DEEPREC-CL	1.6%	0.7%	**52.2%**
Deeprec-ML vs. Paixão et al. [61]	0%	0%	0.4%

Table 5: Page-wise paired t-test.

B

APPENDIX: OTHER PUBLICATIONS

JOURNAL

- C. Badue, R. Guidolini, R. V. Carneiro, P. Azevedo, V. B. Cardoso, A. Forechi, L. Jesus, R. Berriel, T. M. Paixão, F. Mutz, et al. "Self-driving cars: A survey." In: *Expert Syst. with Appl.* (2020), p. 113816
- R. Sarcinelli, R. Guidolini, V. B. Cardoso, T. M. Paixão, R. F. Berriel, P. Azevedo, A. F. De Souza, C. Badue, and T. Oliveira-Santos. "Handling pedestrians in self-driving cars using image tracking and alternative path generation with Frenét frames." In: *Comp. & Graph.* 84 (2019), pp. 173–184
- J. P. V. de Mello, L. Tabelini, R. F. Berriel, T. M. Paixão, A. F. De Souza, C. Badue, N. Sebe, and T. Oliveira-Santos. "Deep traffic light detection by overlaying synthetic context on arbitrary natural images." In: *Comp. & Graph.* (2020)
- V. F. Arruda, R. F. Berriel, T. M. Paixão, C. Badue, A. F. De Souza, N. Sebe, and T. Oliveira-Santos. "Cross-domain object detection using unsupervised image translation." In: *Expert Syst. with Appl.* 192 (2022), p. 116334
- L. Tabelini, R. Berriel, T. M. Paixão, A. F. De Souza, C. Badue, N. Sebe, and T. Oliveira-Santos. "Deep traffic sign detection and recognition without target domain real images." In: *Mach. Vision and Appl.* 33 (2022), p. 50

CONFERENCE

- L. C. Possatti, R. Guidolini, V. B. Cardoso, R. F. Berriel, T. M. Paixão, C. Badue, A. F. De Souza, and T. Oliveira-Santos. "Traffic light recognition using deep learning and prior maps for autonomous cars." In: *Int. Joint Conf. on Neural Networks*. IEEE. 2019, pp. 1–8
- V. F. Arruda, T. M. Paixão, R. F. Berriel, A. F. De Souza, C. Badue, N. Sebe, and T. Oliveira-Santos. "Cross-domain car detection using unsupervised image-to-image translation: From day to night." In: *Int. Joint Conf. on Neural Networks*. IEEE. 2019, pp. 1–8
- L. T. Torres, T. M. Paixão, R. F. Berriel, A. F. De Souza, C. Badue, N. Sebe, and T. Oliveira-Santos. "Effortless Deep Training for Traffic Sign Detection Using Templates and Arbitrary Natural Images." In: *Int. Joint Conf. on Neural Networks*. IEEE. 2019, pp. 1–7

- L. S. Paulucio, T. M. Paixão, R. F. Berriel, A. F. De Souza, C. Badue, and T. Oliveira-Santos. "Product Categorization by Title Using Deep Neural Networks as Feature Extractor." In: *Int. Joint Conf. on Neural Networks*. IEEE. 2020, pp. 1–7
- L. Tabelini, R. Berriel, T. M. Paixao, C. Badue, A. F. De Souza, and T. Oliveira-Santos. "Polylanenet: Lane estimation via deep polynomial regression." In: *Int. Conf. on Pattern Recognit.* IEEE. 2021, pp. 6150–6156
- J. P. V. de Mello, T. M. Paixão, R. Berriel, M. Reyes, C. Badue, A. F. De Souza, and T. Oliveira-Santos. "Deep Learning-based Type Identification of Volumetric MRI Sequences." In: *Int. Conf. on Pattern Recognit.* IEEE. 2021, pp. 1–8
- L. Tabelini, R. Berriel, T. M. Paixao, C. Badue, A. F. De Souza, and T. Oliveira-Santos. "Keep your eyes on the lane: Real-time attention-guided lane detection." In: *Conf. Comput. Vision and Pattern Recognit.* 2021, pp. 294–302

BIBLIOGRAPHY

[1] R. Abitbol, I. Shimshoni, and J. Ben-Dov. "Machine Learning Based Assembly of Fragments of Ancient Papyrus." In: *Journal on Comput.and Cultural Heritage* 14.3 (2021), pp. 1–21.

[2] F. A. Andaló, G. Carneiro, G. Taubin, S. Goldenstein, and L. Velho. "Automatic reconstruction of ancient Portuguese tile panels." In: *IEEE Comput. Graph. Appl* (2016).

[3] F. A. Andaló, G. Taubin, and S. Goldenstein. "PSQP: Puzzle solving by quadratic programming." In: *IEEE Trans. on Pattern Anal. and Mach. Intell.* 39.2 (2017), pp. 385–396.

[4] S. Antol, A. Agrawal, J. Lu, M. Mitchell, D. Batra, C. Lawrence Zitnick, and D. Parikh. "Vqa: Visual question answering." In: *Conf. Comput. Vision and Pattern Recognit.* 2015, pp. 2425–2433.

[5] D. Applegate, R. Bixby, V. Chvatal, and W. Cook. *Concorde: A code for solving traveling salesman problems*. http://www.math.uwaterloo.ca/tsp/concorde. accessed on: October 19, 2020. 2001.

[6] V. F. Arruda, R. F. Berriel, T. M. Paixão, C. Badue, A. F. De Souza, N. Sebe, and T. Oliveira-Santos. "Cross-domain object detection using unsupervised image translation." In: *Expert Syst. with Appl.* 192 (2022), p. 116334.

[7] V. F. Arruda, T. M. Paixão, R. F. Berriel, A. F. De Souza, C. Badue, N. Sebe, and T. Oliveira-Santos. "Cross-domain car detection using unsupervised image-to-image translation: From day to night." In: *Int. Joint Conf. on Neural Networks*. IEEE. 2019, pp. 1–8.

[8] C. R. Babcock. *Tongsun Park's Paper Jigsaw Puzzle Solved*. 1977. URL: https://www.washingtonpost.com/archive/politics/1977/05/13/tongsun-parks-paper-jigsaw-puzzle-solved/7cbe5c4c-285f-4bff-bccf-93e3004f0446.

[9] H. Badawy, E. Emary, M. Yassien, and M. Fathi. "Discrete Grey Wolf Optimization for Shredded Document Reconstruction." In: *Int. Conf. on Adv. Intell. System and Inf.* 2018, pp. 284–293.

[10] C. Badue, R. Guidolini, R. V. Carneiro, P. Azevedo, V. B. Cardoso, A. Forechi, L. Jesus, R. Berriel, T. M. Paixão, F. Mutz, et al. "Self-driving cars: A survey." In: *Expert Syst. with Appl.* (2020), p. 113816.

[11] J. Balme. "Reconstruction of shredded documents in the absence of shape information." In: *Working paper, Dept. of Comp. Sci., Yale Univ., USA, Tech. Rep.* (2007).

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

[13] D. Bridger, D. Danon, and A. Tal. "Solving jigsaw puzzles with eroded boundaries." In: *Conf. Comput. Vision and Pattern Recognit.* IEEE. 2020, pp. 3526–3535.

[14] 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.

[15] P. Butler, P. Chakraborty, and N. Ramakrishan. "The Deshredder: A visual analytic approach to reconstructing shredded documents." In: *IEEE Conf. on Vis. Analytics Sci. and Technol.* IEEE. 2012, pp. 113–122.

[16] G. Chen, J. Wu, C. Jia, and Y. Zhang. "A pipeline for reconstructing cross-shredded English document." In: *IEEE Int. Conf. on Image, Vision and Computing*. 2017, pp. 1034–1039.

[17] J. Chen, D. Ke, Z. Wang, and Y. Liu. "A high splicing accuracy solution to reconstruction of cross-cut shredded text document problem." In: *Multimedia Tools and Appl.* 77.15 (2018), pp. 19281–19300.

[18] J. Chen, M. Tian, X. Qi, W. Wang, and Y. Liu. "A solution to reconstruct cross-cut shredded text documents based on constrained seed K-means algorithm and ant colony algorithm." In: *Expert Syst. with Appl.* 127 (2019), pp. 35–46.

[19] S. Chopra, R. Hadsell, and Y. LeCun. "Learning a similarity metric discriminatively, with application to face verification." In: *IEEE/CVF Conf. on Comput. Vision and Pattern Recognit.* 2005, pp. 539–546.

[20] J. Cohen. "Statistical Power Analysis for the Behavioral Sciences." In: *Technometrics* 31.4 (1988), pp. 499–500.

[21] D. Cohn, L. Atlas, and R. Ladner. "Improving generalization with active learning." In: *Mach. learning* 15.2 (1994), pp. 201–221.

[22] Z. Daniels and M. Idrees. "Semi-automatic reconstruction of cross-cut shredded documents." In: *Center of Research in Computer Vision*. University of Central Florida, 2013.

[23] Darpa. *Darpa Shredder Challenge*. Acessed on: 2020, October 14th. 2011. URL: https://web.archive.org/web/20120121201439/http://www.shredderchallenge.com/.

[24] P. De Smet, J. De Bock, and W. Philips. "Semiautomatic reconstruction of strip-shredded documents." In: *Image and Video Commun. and Process.* Vol. 5685. Int. Soc. for Optics and Photonics. 2005, pp. 239–248.

[25] J. D. Derian. "Arms, Hostages, and the Importance of Shredding in Earnest: Reading the National Security Culture (II)." In: *Social Text* 22 (1989), pp. 79–91.

[26] C. Doersch, A. Gupta, and A. A. Efros. "Unsupervised visual representation learning by context prediction." In: *IEEE/CVF Int. Conf. on Comput. Vision*. 2015, pp. 1422–1430.

[27] H. Freeman and L. Garder. "Apictorial jigsaw puzzles: The computer solution of a problem in pattern recognition." In: 2 (1964), pp. 118–127.

[28] K. Fukushima. "Neocognitron: A hierarchical neural network capable of visual pattern recognition." In: *Neural Netw.* 1.2 (1988), pp. 119–130.

[29] Y.-F. Ge, Y.-J. Gong, W.-J. Yu, X.-M. Hu, and J. Zhang. "Reconstructing cross-cut shredded text documents: A genetic algorithm with splicing-driven reproduction." In: *Conf. on Genetic and Evol. Comput.* 2015, pp. 847–853.

[30] Y.-J. Gong, Y.-F. Ge, J.-J. Li, J. Zhang, and W. Ip. "A splicing-driven memetic algorithm for reconstructing cross-cut shredded text documents." In: *Appl. Soft Computing* 45 (2016), pp. 163–172.

[31] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. "Generative adversarial nets." In: *Adv. in Neural Inf. Process. Syst.* 2014, pp. 2672–2680.

[32] 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*. Springer. 2015, pp. 447–461.

[33] M. Hahsler and K. Hornik. "TSP-Infrastructure for the traveling salesperson problem." In: *Journal of Statistical Software* 23.2 (2007), pp. 1–21.

[34] A. W. Harley, A. Ufkes, and K. G. Derpanis. "Evaluation of deep convolutional nets for document image classification and retrieval." In: *IEEE Int. Conf. on Document Anal. and Recognit.* 2015, pp. 991–995.

[35] 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." In: *arXiv preprint arXiv:1602.07360* (2016).

[36] R. Jonker and T. Volgenant. "Transforming asymmetric into symmetric traveling salesman problems." In: *Operations Res. Lett.* 2.4 (1983), pp. 161–163.

[37] E. Justino, L. S. Oliveira, and C. Freitas. "Reconstructing shredded documents through feature matching." In: *Forensic Sci. Int.* 160.2-3 (2006), pp. 140–147.

[38] D. P. Kingma and J. Ba. "Adam: A Method for Stochastic Optimization." In: *Int. Conf. for Learn. Representations*. 2015.

[39] D. A. Kosiba, P. M. Devaux, S. Balasubramanian, T. L. Gandhi, and K. Kasturi. "An automatic jigsaw puzzle solver." In: *Int. Conf. on Pattern Recognit.* Vol. 1. IEEE. 1994, pp. 616–618.

[40] A. Krizhevsky, I. Sutskever, and G. E. Hinton. "ImageNet Classification with Deep Convolutional Neural Networks." In: *Adv. in Neural Inf. Process. Syst.* 2012, pp. 1097–1105.

[41] C. Le and X. Li. "JigsawNet: Shredded Image Reassembly using Convolutional Neural Network and Loop-based Composition." In: *arXiv preprint arXiv:1809.04137* (2018).

[42] Y. LeCun, Y. Bengio, et al. "Convolutional networks for images, speech, and time series." In: *The Handbook of Brain Theory and Neural Netw.* 3361.10 (1995).

[43] Y. LeCun, B. E. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. E. Hubbard, and L. D. Jackel. "Handwritten digit recognition with a back-propagation network." In: *Adv. in Neural Inf. Process. Syst. (NeurIPS)*. 1990, pp. 396–404.

[44] 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*. 2006, pp. 665–666.

[45] P. Li, X. Fang, L. Pan, Y. Piao, and M. Jiao. "Reconstruction of shredded paper documents by feature matching." In: *Math. Problems in Eng.* 2014 (2014).

[46] R. Li, S. Liu, G. Wang, G. Liu, and B. Zeng. "JigsawGAN: Auxiliary Learning for Solving Jigsaw Puzzles With Generative Adversarial Networks." In: *IEEE Trans. on Image Processing* 31 (2021), pp. 513–524.

[47] Y. Liang and X. Li. "Reassembling Shredded Document Stripes Using Word-path Metric and Greedy Composition Optimal Matching Solver." In: *IEEE Trans. on Multimedia* 22.5 (2020), pp. 1168–1181.

[48] H.-Y. Lin and W.-C. Fan-Chiang. "Reconstruction of shredded document based on image feature matching." In: *Expert Syst. with Appl.* 39.3 (2012), pp. 3324–3332.

[49] J. Long, E. Shelhamer, and T. Darrell. "Fully convolutional networks for semantic segmentation." In: *Conf. Comput. Vision and Pattern Recognit.* 2015, pp. 3431–3440.

[50] Loveday Morris. *Thirty years after the Berlin Wall fell, a Stasi spy puzzle remains unsolved*. accessed July 04, 2020. URL: https://www.washingtonpost.com/world/thirty-years-after-the-berlin-wall-fell-no-end-in-sight-for-stasi-spy-puzzle/2019/11/01/160d8ae2-fb29-11e9-9e02-1d45cb3dfa8f_story.html.

[51] A. A. Low. "Waste-paper receptacle." 929960. 1909.

[52] L. v. d. Maaten and G. Hinton. "Visualizing data using t-SNE." In: *J. of Mach. Learning Res.* 9.Nov (2008), pp. 2579–2605.

[53] M. Marques and C. Freitas. "Document Decipherment-restoration: Strip-shredded Document Reconstruction based on Color." In: *IEEE Latin America Trans.* 11.6 (2013), pp. 1359–1365.

[54] J. P. V. de Mello, T. M. Paixão, R. Berriel, M. Reyes, C. Badue, A. F. De Souza, and T. Oliveira-Santos. "Deep Learning-based Type Identification of Volumetric MRI Sequences." In: *Int. Conf. on Pattern Recognit.* IEEE. 2021, pp. 1–8.

[55] J. P. V. de Mello, L. Tabelini, R. F. Berriel, T. M. Paixão, A. F. De Souza, C. Badue, N. Sebe, and T. Oliveira-Santos. "Deep traffic light detection by overlaying synthetic context on arbitrary natural images." In: *Comp. & Graph.* (2020).

[56] R. Moore, L. Perdue, and J. N. Rowe. *The Washington Connection*. Condor, 1977.

[57] W. Morandell. "Evaluation and reconstruction of strip-shredded text documents." MA thesis. Inst. of Comput. Graph. and Algorithms, Vienna Univ. of Technol., 2008.

[58] T. A. Nartker, S. V. Rice, and S. E. Lumos. "Software tools and test data for research and testing of page-reading OCR systems." In: *Electron. Imag.* 2005, pp. 37–47.

[59] T. R. Nielsen, P. Drewsen, and K. Hansen. "Solving jigsaw puzzles using image features." In: *Pattern Recognit. Lett.* 29.14 (2008), pp. 1924–1933.

[60] M. Noroozi and P. Favaro. "Unsupervised learning of visual representations by solving jigsaw puzzles." In: *Eur. Conf. on Comput. Vision*. 2016, pp. 69–84.

[61] T. M. Paixão, M. C. S. Boeres, C. O. A. Freitas, and T. Oliveira-Santos. "Exploring Character Shapes for Unsupervised Reconstruction of Strip-shredded Text Documents." In: *IEEE Trans. Inf. Forensics Secur.* 14.7 (2019), pp. 1744–1754. ISSN: 1556-6013.

[62] T. M. Paixão, R. F. Berriel, M. C. S. Boeres, C. Badue, A. F. De Souza, and T. Oliveira-Santos. "A deep learning-based compatibility score for reconstruction of strip-shredded text documents." In: *Conf. on Graph., Patterns and Images*. 2018, pp. 87–94.

[63] T. M. Paixão, R. F. Berriel, M. C. S. Boeres, A. L. Koerich, C. Badue, A. F. De Souza, and T. Oliveira-Santos. "Self-supervised deep reconstruction of mixed strip-shredded text documents." In: *Pattern Recognit.* 107 (2020), p. 107535.

[64] T. M. Paixão, R. F. Berriel, M. C. S. Boeres, A. L. Koerich, C. Badue, A. F. De Souza, and T. Oliveira-Santos. "A human-in-the-loop recommendation-based framework for reconstruction of mechanically shredded documents." In: *Pattern Recognit. Letters* (under review).

[65] T. M. Paixão, R. F. Berriel, M. C. S. Boeres, A. L. Koerich, C. Badue, A. F. D. Souza, and T. Oliveira-Santos. "Fast(er) Reconstruction of Shredded Text Documents via Self-Supervised Deep Asymmetric Metric Learning." In: *IEEE/CVF Conf. on Comp. Vision and Pattern Recognit.* 2020, pp. 14343–14351.

[66] L. S. Paulucio, T. M. Paixão, R. F. Berriel, A. F. De Souza, C. Badue, and T. Oliveira-Santos. "Product Categorization by Title Using Deep Neural Networks as Feature Extractor." In: *Int. Joint Conf. on Neural Networks*. IEEE. 2020, pp. 1–7.

[67] M.-M. Paumard, D. Picard, and H. Tabia. "Jigsaw Puzzle Solving Using Local Feature Co-Occurrences in Deep Neural Networks." In: *IEEE Int. Conf. on Image Process.* 2018, pp. 1018–1022.

[68] M.-M. Paumard, D. Picard, and H. Tabia. "Deepzzle: Solving visual jigsaw puzzles with deep learning and shortest path optimization." In: *IEEE Trans. on Image Processing* 29 (2020), pp. 3569–3581.

[69] J. Perl, M. Diem, F. Kleber, and R. Sablatnig. "Strip shredded document reconstruction using optical character recognition." In: *Int. Conf. on Imag. for Crime Detection and Prevention*. 2011, pp. 1–6.

[70] T. Phienthrakul, T. Santitewagun, and N. Hnoohom. "A Linear Scoring Algorithm for Shredded Paper Reconstruction." In: *Int. Conf. on Signal-Image Tech. & Internet-Based Syst.* 2015, pp. 623–627.

[71] A. Pirrone, M. Beurton-Aimar, and N. Journet. "Self-supervised deep metric learning for ancient papyrus fragments retrieval." In: *International Journal on Document Analysis and Recognition (IJDAR)* 24.3 (2021), pp. 219–234.

[72] D. Pöhler, R. Zimmermann, B. Widdecke, H. Zoberbier, J. Schneider, B. Nickolay, and J. Krüger. "Content representation and pairwise feature matching method for virtual reconstruction of shredded documents." In: *9th IEEE Int. Symp. Image and Signal Process. and Anal.* 2015, pp. 143–148.

[73] D. Pomeranz, M. Shemesh, and O. Ben-Shahar. "A fully automated greedy square jigsaw puzzle solver." In: *IEEE Conf. Comput. Vision and Pattern Recognit.* 2011, pp. 9–16.

[74] L. C. Possatti, R. Guidolini, V. B. Cardoso, R. F. Berriel, T. M. Paixão, C. Badue, A. F. De Souza, and T. Oliveira-Santos. "Traffic light recognition using deep learning and prior maps for autonomous cars." In: *Int. Joint Conf. on Neural Networks*. IEEE. 2019, pp. 1–8.

[75] M. Prandtstetter and G. Raidl. "Meta-heuristics for reconstructing cross cut shredded text documents." In: *Conf. on Genetic and Evol. Comput.* 2009, pp. 349–356.

[76] M. Prandtstetter and G. R. Raidl. "Combining forces to reconstruct strip shredded text documents." In: *Int. Workshop on Hybrid Metaheuristics*. Springer. 2008, pp. 175–189.

[77] R. Ranca. "A modular framework for the automatic reconstruction of shredded documents." In: *Workshops AAAI Conf. on Artif. Intell.* 2013.

[78] S. Ren, K. He, R. Girshick, and J. Sun. "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks." In: *Adv. in Neural Inf. Process. Syst.* 2015, pp. 91–99.

[79] D. Rika, D. Sholomon, E. David, and N. S. Netanyahu. "TEN: Twin Embedding Networks for the Jigsaw Puzzle Problem with Eroded Boundaries." In: *arXiv preprint arXiv:2203.06488* (2022).

[80] N. Rubens, M. Elahi, M. Sugiyama, and D. Kaplan. "Active learning in recommender systems." In: *Recommender systems handbook*. Springer, 2015, pp. 809–846.

[81] P. Saboia and S. Goldenstein. "Assessing cross-cut shredded document assembly." In: *Iberoamerican Congr. on Pattern Recognit.* Springer. 2014, pp. 272–279.

[82] R. Sarcinelli, R. Guidolini, V. B. Cardoso, T. M. Paixão, R. F. Berriel, P. Azevedo, A. F. De Souza, C. Badue, and T. Oliveira-Santos. "Handling pedestrians in self-driving cars using image tracking and alternative path generation with Frenét frames." In: *Comp. & Graph.* 84 (2019), pp. 173–184.

[83] J. Sauvola and M. Pietikäinen. "Adaptive document image binarization." In: *Pattern Recognit.* 33.2 (2000), pp. 225–236.

[84] C. Schauer, M. Prandtstetter, and G. R. Raidl. "A memetic algorithm for reconstructing cross-cut shredded text documents." In: *Int. Workshop on Hybrid Metaheuristics*. Springer. 2010, pp. 103–117.

[85] F. Schroff, D. Kalenichenko, and J. Philbin. "FaceNet: A Unified Embedding for Face Recognition and Clustering." In: *IEEE/CVF Conf. on Comput. Vision and Pattern Recognit.* 2015, pp. 815–823.

[86] B. Settles. "Active learning literature survey. University of Wisconsin." In: *Computer Science Department* (2010).

[87] S. Shang, H. T. Sencar, N. Memon, and X. Kong. "A semi-automatic deshredding method based on curve matching." In: *2014 IEEE International Conference on Image Processing (ICIP)*. IEEE. 2014, pp. 5537–5541.

[88] C. E. Shannon. "A mathematical theory of communication." In: *The Bell system technical journal* 27.3 (1948), pp. 379–423.

[89] D. Sholomon, O. E. David, and N. S. Netanyahu. "DNN-Buddies: a deep neural network-based estimation metric for the jigsaw puzzle problem." In: *Int. Conf. on Art. Neural Netw.* 2016, pp. 170–178.

[90] A. Skeoch. "An investigation into automated shredded document reconstruction using heuristic search algorithms." In: *Unpublished Ph. D. Thesis in the Univ. of Bath, UK* (2006), p. 107.

[91] A. Sleit, Y. Massad, and M. Musaddaq. "An alternative clustering approach for reconstructing cross cut shredded text documents." In: *Telecommun. Syst.* 52.3 (2013), pp. 1491–1501.

[92] L. Tabelini, R. Berriel, T. M. Paixao, C. Badue, A. F. De Souza, and T. Oliveira-Santos. "Keep your eyes on the lane: Real-time attention-guided lane detection." In: *Conf. Comput. Vision and Pattern Recognit.* 2021, pp. 294–302.

[93] L. Tabelini, R. Berriel, T. M. Paixao, C. Badue, A. F. De Souza, and T. Oliveira-Santos. "Polylanenet: Lane estimation via deep polynomial regression." In: *Int. Conf. on Pattern Recognit.* IEEE. 2021, pp. 6150–6156.

[94] L. Tabelini, R. Berriel, T. M. Paixão, A. F. De Souza, C. Badue, N. Sebe, and T. Oliveira-Santos. "Deep traffic sign detection and recognition without target domain real images." In: *Mach. Vision and Appl.* 33 (2022), p. 50.

[95] L. T. Torres, T. M. Paixão, R. F. Berriel, A. F. De Souza, C. Badue, N. Sebe, and T. Oliveira-Santos. "Effortless Deep Training for Traffic Sign Detection Using Templates and Arbitrary Natural Images." In: *Int. Joint Conf. on Neural Networks*. IEEE. 2019, pp. 1–7.

[96] 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 Info. Technol.* 2004, pp. 334–337.

[97] L. Van Der Maaten. "Accelerating t-SNE using tree-based algorithms." In: *J. of Mach. Learn. Res.* 15.1 (2014), pp. 3221–3245.

[98] K. Weinberger and L. Saul. "Distance Metric Learning for Large Margin Nearest Neighbor Classification." In: *J. of Mach. Learn. Res.* 10.Feb (2009), pp. 207–244.

[99] A. R. Willis and D. B. Cooper. "Computational reconstruction of ancient artifacts." In: *IEEE Signal Process. Mag.* 25.4 (2008), pp. 65–83.

[100] N. Xing, S. Shi, and Y. Xing. "Shreds assembly based on character stroke feature." In: *Procedia Comput. Sci.* 116 (2017), pp. 151–157.

[101] N. Xing and J. Zhang. "Graphical-character-based shredded Chinese document reconstruction." In: *Multimedia Tools and Appl.* 76.10 (2017), pp. 12871–12891.

[102] 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." In: *Abstract and Appl. Anal.* Vol. 2014. Hindawi. 2014.

[103] H. Zhang, J. K. Lai, and M. Bächer. "Hallucination: A mixed-initiative approach for efficient document reconstruction." In: *Workshops AAAI Conf. on Artif. Intell.* 2012.