Towards Zero-Shot Document Image Classification

Artigo WPOS

Lucas de Almeida Bandeira Macedo

Universidade de Brasília
Departamento de Ciência da Computação

João Paulo Costa
João Pedro Felix de Almeida
Marina Silva Lyra
Daniel Alves de Santana
Pedro Garcia Freitas
Li Weigang

Novembro de 2025

Capítulos

1. Introdução
2. Metodologia
3. Resultados
4. Conclusão

Introdução

Contexto - Documentos e Compliance

- Ambiente Bancário
- Classificação de documentos
- Exemplos:

MINISTÉRIO DA FAZENDA — IMPOSTO SOBRE A RENDA - PESSOA FÍSICA
SECRETARIA DA RECEITA FEDERAL DO BRASIL — EXERCÍCIO 2017 — ANO-CALENDÁRIO 2016

RECIBO DE ENTREGA DA DECLARAÇÃO DE AJUSTE ANUAL - OPÇÃO PELO DESCONTO SIMPLIFICADO
DECLARAÇÃO ORIGINAL

IDENTIFICAÇÃO DO DECLARANTE

CPF do declarante	Nome do declarante	Telefone
352.659.378-08	ERIKA TOMAZELLA	(11) 43791221

Endereço	Número	Complemento
RUA RUA JUSTINO ALVES BATISTA	89	AP 04 BL P

Bairro/Distrito	CEP	Município	UF
VILA YOLANDA	06126-120	OSASCO	SP

(Valores em Reais)

TOTAL RENDIMENTOS TRIBUTÁVEIS	62.269,42
IMPOSTO DEVIDO	3.562,58
IMPOSTO A RESTITUIR	781,16

Evolução dos Documentos

Zero-Shot Learning

Permite que o modelo reconheça elementos de classes nunca vistas no treinamento

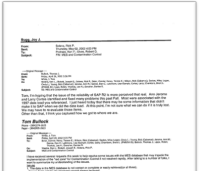
Sobre a Pesquisa

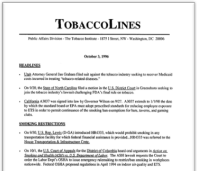
Contribuições

1. **Novo dataset LA-CDIP**
   - Classificação ZSL
   - Derivado do RVL-CDIP
2. **Abordagem de Visual Document Matching (VDM)**
   - Similaridade de documentos
   - Metric Learning
   - Generalização Zero-Shot
3. **Avaliação sistemática**
   - Benchmark extensivo

Metodologia

Dataset - Motivação





TOBACCOLINES

Public Affairs Division - The Tobacco Institute - 1875 I Street, NW - Washington, DC 20006

October 3, 1996

HEADLINES

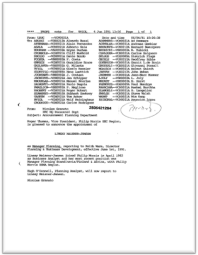
- Utah Attorney General Jan Graham filed suit against the tobacco industry seeking to recover Medicaid costs incurred in treating "tobacco-related diseases."
- On 9/30, the State of North Carolina filed a motion in the U.S. District Court in Greensboro seeking to join the tobacco industry's lawsuit challenging FDA's final rule on tobacco.
- California ARB3 was signed into law by Governor Wilson on 9/27. ARB3 extends to 1/98 the date by which tobacco manufacturers must be in compliance with the unaffiliated vendors law that was passed in 1994 in order to permit continuation of the standing free samples/coupons for bars, taverns, and gaming clubs.

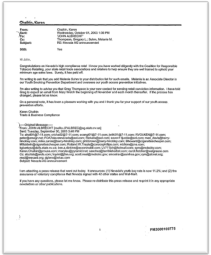
SMOKING RESTRICTIONS

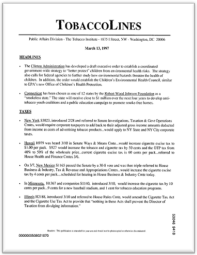
- On 9/30, U.S. Rep. Lewis (D-GA) introduced HR4151, which would prohibit smoking in any transportation facility for which federal financial assistance is provided. HR4151 was referred to the House Transportation & Infrastructure Cmte.
- On 9/9, the U.S. Court of Appeals for the District of Columbia heard oral arguments in Action on Smoking and Health(ASH) vs. U.S. Department of Labor. The ASH lawsuit requests the Court to order OSHA to issue a "comprehensive" workplace smoking standard, or alternatively, to ban smoking nationwide. Federal OSHA proposed regulations in April 1994 on indoor air quality and ETS.
- The Mayor of Friendship Heights, MD is proposing an ordinance that would ban the use or discharging of tobacco products in public. The ordinance is scheduled to be discussed during a public hearing on 10/15. If approved, the measure would bend the control of the Montgomery County Council, which is currently considering a county-wide smoking ban of its own.
- On 9/5, the Sun Work TV City Council received closer presentations from each group in the competition for the contract to operate the TV station. The Council is considering whether the groups presented their viewpoints on a proposed city ordinance that would ban smoking in restaurants, public buildings and places of employment.

Source: The information is intended for use as an internal or proprietary memo to members.

0000000000TTOBACCO







TOBACCOLINES

Public Affairs Division - The Tobacco Institute - 1875 I Street, NW - Washington, DC 20006

March 15, 1997

HEADLINES

- The Clinton Administration has developed a draft resource order to establish a coordinated Administration "Tobacco Team" to work on the issues of tobacco production, use and health. The Tobacco Team will be federal agencies to further trade issues (environmental hazards, business tax, health, of tobacco, the trade embargo, and the Cuban's Environmental Health Council, similar to EPA's Asbestos Tiger Team.
- Congressman Jim Ross (Arkansas) is one of 110 members of the Million Mom Marches Foundation's "mandatory dues." The state will receive close to $1 million in dues, which will be used to fund tobacco-related coalitions and a public education campaign to promote smoke-free homes.

TAXES

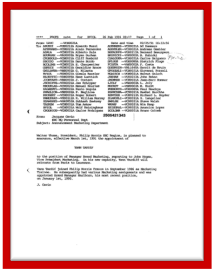
- New York S9023, introduced 2/28 and referred to Senate Investigations, Taxation & Govt Operations Cmte, would require corporate taxpayers to add back to their adjusted gross income amounts deducted from income as costs of advertising tobacco products, would apply to NY State and NYC City corporate taxes.
- Hawaii HB79 was heard 3/10 in Senate Ways & Means Cmte; would increase cigarette excise tax to $1.00 per pack. HS17 would increase the tobacco and cigarette tax by 50 cents and the ESP tax from 40% to 70% of the wholesale price, current cigarette excise tax is 60 cents per pack, referred to House Health and Human Services Cmte.
- The KY, New Mexico H 910 passed the Senate by a 30-8 vote and was then triple referred to House Appr, Ways & Means, and Health & Welfare. The bill would increase the state cigarette excise tax by 4 cents per pack, ...scheduled for hearing on House Revenue & Industry Cmte today.
- In Minnesota, HS787 and companion S2143, introduced 3/10, would increase the cigarette tax by 10 cents per pack. 2 cents for a new Youth Coalition, and 1 cent for tobacco education programs.
- Illinois HS186, introduced 3/10 and referred to House Rules Cmte, would amend the Cigarette Tax Act and the Cigarette Use Act to provide that "nothing in these Acts shall prevent the Director of Taxation from divulging information."

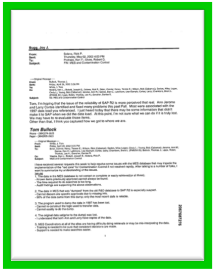
Source: The information is intended for use as an internal or proprietary memo to members.

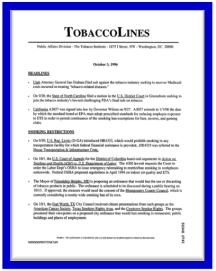
0000000000TTOBACCO

Lucas de Almeida Bandeira Macedo (UnB) Towards ZS-DIC Novembro de 2025 9 / 21

Dataset - Motivação







TOBACCOLINES

Public Affairs Division · The Tobacco Institute · 1875 I Street, NW · Washington, DC 20006

October 3, 1996

HEADLINES

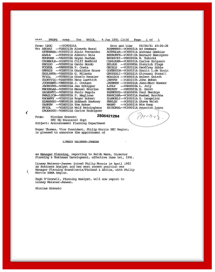
- Utah Attorney General Jan Graham filed suit against the tobacco industry seeking to recover Medicaid costs incurred in treating "tobacco-related diseases."
- On 9/30, the State of North Carolina filed a motion in the U.S. District Court in Greensboro seeking to join the tobacco industry's lawsuit challenging FDA's final rule on tobacco.
- California A.B.97 was signed into law by Governor Wilson on 9/27. A.B.97 extends to 1/1/98 the date by which the standard based on EPA's most stringent prescribed standards for reducing employee exposure to ETS in order to permit continuance of the smoking ban exemption for bars, taverns, and gaming cards.

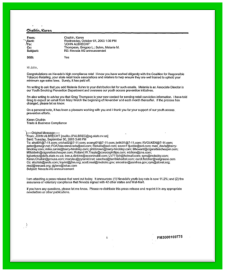
SMOKING RESTRICTIONS

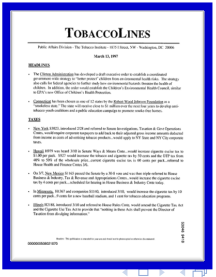
- On 9/30, U.S. Rep. James Oberstar (D-Minn) introduced HR2355, which would prohibit smoking in any transportation facility for which federal financial assistance is provided. HR2355 was referred to the House Transportation and Infrastructure Committee.
- On 10/1, the U.S. Court of Appeals for the District of Columbia heard oral arguments in Action on Smoking and Health v. Department of Labor. The plaintiff contends that the ASH local lawsuit filed under the Labor Dept's OSHA to issue emergency rulemaking to establish smoking in workplaces as a "serious health hazard" is "ripe" for a decision. The Labor Dept. has not ruled on the petition.
- The Mayor of Vinnedale Beach, MD is proposing an ordinance that would ban the use or discarding of tobacco products in public. The ordinance is scheduled to be discussed during a public hearing on 10/21. If approved, the measure would need the consent of the Montgomery County Council, which is currently considering a county-wide smoking ban of its own.
- On 10/1, the Oak Woods TX City Council received citizen presentations from each group in the community that favor Smoking Rights Day and the Clearwater Smoke Filters. The press presented their viewpoints on a proposed smoking ban ordinance that would ban smoking in restaurants, public buildings and places of employment.

Source: The preceding is a tabulation by Lexis-Nexis of recent items of tobacco-related interest.

0000000000/TTOA 1.0/0







TOBACCOLINES

Public Affairs Division · The Tobacco Institute · 1875 I Street, NW · Washington, DC 20006

March 13, 1997

HEADLINES

- The Clinton Administration has developed a draft executive order to establish a coordinated government-wide strategy to "better protect" children from environmental health risks. The strategy would "ensure that Federal agencies take into account environmental health risks to children. In addition, the order would establish the Children's Environmental Health Council, similar to EPA's new Office of Children's Health Protection.
- Connecticut has been chosen as one of 12 states by the Robert Wood Johnson Foundation as a "tobacco free." The state will receive a check for $1 million over the next four years to develop anti-tobacco youth coalitions and a public education campaign to promote smoke-free homes.

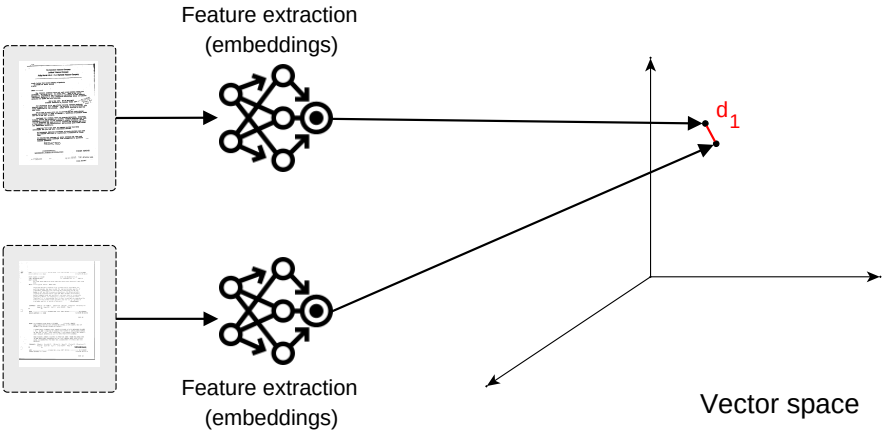
TAXES

- New York S.9821, introduced 2/28 and referred to Senate Investigations, Taxation & Govt Operations Comm, would require corporate taxpayers to add back to their adjusted gross income amounts deducted from their income as costs of activities tobacco products, would apply to NY State and NYC City corporate taxes.
- Hawaii H.B.9 was heard 3/10 in Senate Ways & Means Comm. would increase cigarette excise tax to $2.00 per pack, would increase tax on tobacco and cigarette tax by 50 cents and the ETF tax from 40% to 50% of the wholesale price of cigarettes, would increase tax on 1.90 cents per pack, referred to House Health and Finance Comm 3/6.
- On NY Tax Matters 3/10 issued the benefits of a $0.50 cents and tax free triple referred to House Business & Industry. Tax & Revenue and Appropriations Comm. would increase the cigarette excise tax by 0.cents per pack, scheduled for hearing in House Business & Industry Comm today.
- In Minnesota, HF.707 and companion S.510, introduced 3/10, would increase the cigarette tax by 10 cents per pack. .0 cents for a new baseline stadium, and .1 cent for tobacco education programs.
- Illinois HB.586, introduced 3/10 and referred to House Rules Comm, would amend the Cigarette Tax Act to increase the tax by 0.cents per pack "holding in Illinois Act shall pursuant the Illinois of Taxation from dealing information."

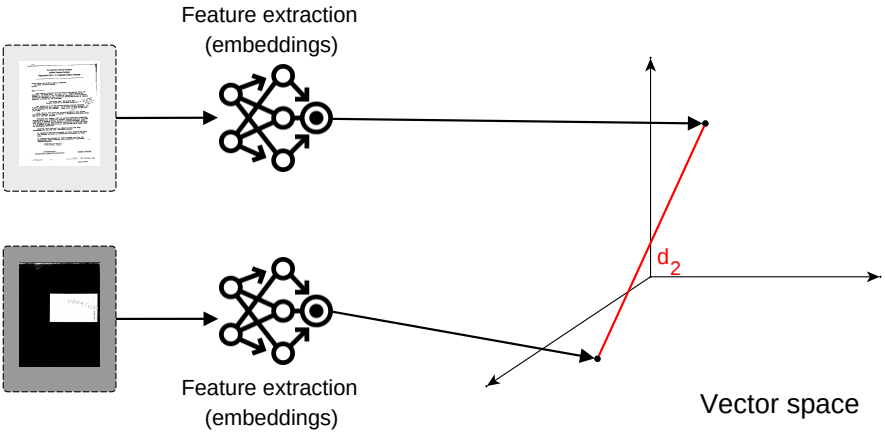
Source: The preceding is a tabulation by Lexis-Nexis of recent items of tobacco-related interest.

0000000000/BBT 1.0/0

Visual Document Matching



Visual Document Matching

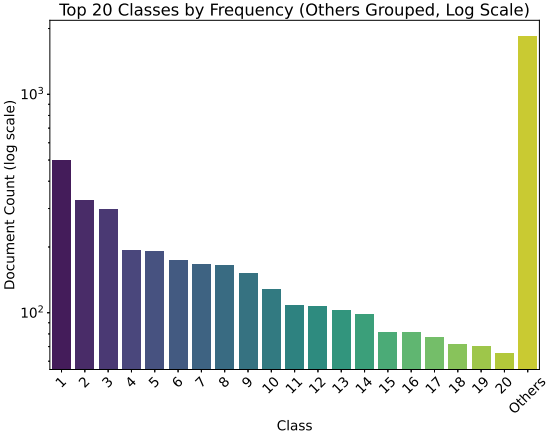


Resultados

LA-CDIP Dataset - Estatísticas

Composição:

- 4.993 documentos
- 144 classes diferentes
- Min: 2 documentos/classe
- Max: 497 documentos/classe
- Mediana: 13 documentos/classe



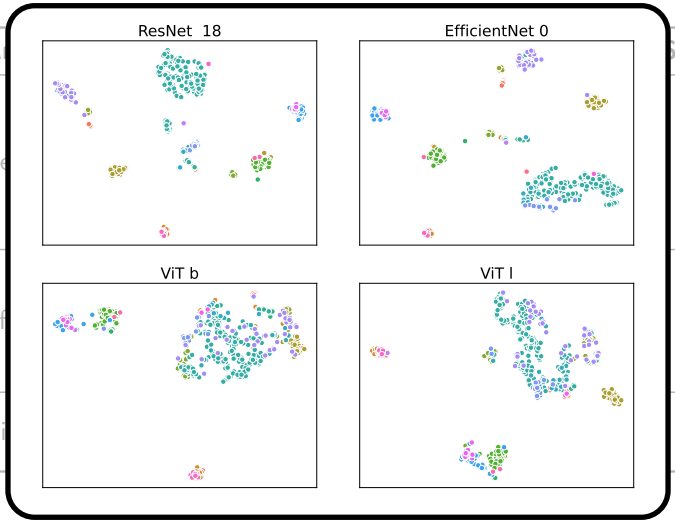
Resultados - Equal Error Rate (%)

Architecture	Edition	Params	ZSL	Test ZSL
VGG	11	129M	7.47	14.24
	13	129M	7.03	9.30
	16	134M	8.29	14.74
	19	139M	7.30	17.08
ResNet	18	11M	5.03	4.98
	34	21M	4.32	4.13
	50	23M	6.90	10.34
	101	42M	8.20	11.31
	152	58M	9.44	12.70
MobileNetV3	Small	1M	7.98	12.74
	Large	4M	8.16	8.45
EfficientNet	0	4M	4.41	6.02
	1	6M	**3.93**	8.88
	2	7M	5.73	7.29
	3	10M	5.65	7.37
ViT	Base	87M	12.43	19.72
	Large	305M	13.16	19.88

Resultados - Equal Error Rate (%)

Architecture	Edition	Params	ZSL	Test ZSL
ResNet	18	11M	5.03	4.98
	34	21M	4.32	**4.13**
	50	23M	6.90	10.34
	101	42M	8.20	11.31
	152	58M	9.44	12.70
EfficientNet	0	4M	4.41	6.02
	1	6M	**3.93**	8.88
	2	7M	5.73	7.29
	3	10M	5.65	7.37
ViT	Base	87M	12.43	19.72
	Large	305M	13.16	19.88

Resultados - Visualização TSNE



Conclusão

Conclusões

1. **LA-CDIP Dataset**

   Categorizado por layout visual para Zero-Shot Learning

2. **Metodologia com VDM**

   Enforçando comparação de documentos como forma de classificação

3. **Benchmarking Sistemático**

   Comparação entre modelos visuais

Publicações

Visual Document Matching for Zero-Shot Document Classification

Lucas De Almeida Bandeira Macedo [0000−0004−0052−4702], João Paulo Vieira, Gaetia Ze [0000−0001−7570−9802], João Pedro Felix De Almeida [0000−0006−3267−9032], Pedro Garcia Freitas [0000−0003−0866−4583], and Li Weigang [0000−0003−1826−5056]

University of Brasília, UnB - Brasília, Brasília - Federal District, 70910-900

Abstract. Accurate document identification is crucial for ensuring compliance and maintaining consistency across various applications. Consequently, document classification has been extensively studied. However, the dynamic nature of documents often renders simple classification inefficient, necessitating frequent model retraining. To mitigate this constant maintenance, comparing incoming documents against known references is a more suitable approach. To address the zero-shot classification challenge, we introduce LA-CDIP, a document dataset. This dataset focuses on verifying whether two documents share an identical visual layout structure, a particularly effective approach in scenarios where training classes do not align with those encountered during inference. Despite its significant potential, the Zero-Shot Learning (ZSL) approach remains largely underexplored in document layout understanding. To support our study, we introduce Layout-Aware Complex Document Information Processing (LA-CDIP), a dataset comprising 4,061 documents across 134 classes. We reorganized this dataset from the RVL-CDIP database to facilitate zero-shot classification evaluations. Our proposed methodology is benchmarked using a siamese network within a contrastive learning framework across multiple backbone architectures, including ResNet, EfficientNet, and Vision Transformer (ViT). In zero-shot scenarios, our method achieves an Equal Error Rate (EER) below 5% in 1-vs-1 verification with cross-validation. Furthermore, our VGG approach outperforms fighter Large Language Models (LLMs) and rivals GPT-4o. These findings highlight the superiority of specialized VDM techniques over general-purpose multimodal models, demonstrating high accuracy with significantly fewer parameters, making our approach more practical for real-world applications.

Keywords: Zero-shot learning · Metric Learning · Document Understanding · Document Image Classification · Visual Document Matching

1 Introduction

Documents are central to many organizations, acting as official records for contracts, identities, and evidence, and playing a crucial role in legal, financial, and

ICDAR

Towards Zero-Shot Document Image Classification

Lucas Macedo, João Paulo Costa, João Pedro Felix de Almeida, Pedro Freitas, and Li Weigang
Department of Computer Science, University of Brasília, Brasília, Brazil

Abstract—Classification is a fundamental tool to automate the process of categorizing documents in many real-world applications, such as information management, financial document processing, healthcare records management, news categorization, fraud detection, regulatory compliance, and many others. Because of the diversity of document layouts, accurate document classification is of paramount importance for various companies. However, documents often change in terms of format and their visual patterns, which may impair a single classification model. Moreover, model continuous and retraining often demands substantial resources, including computational power and manpower, annoying data. Therefore, techniques capable of classifying documents by simply observing new data, without necessarily requiring retraining the classifier, are of immense importance for a wide variety of applications. In this context, Zero-Shot Learning (ZSL) offers a promising approach to document classification because it handles diverse and ever-changing document content. In this work, we tackle the gap involving Zero-Shot Document Image Classification (ZS-DIC), where we classify documents that have not been seen by the model during training. To achieve this, we introduce Layout-Aware Complex Document Information Processing (LA-CDIP), a dataset tailored for this problem. LA-CDIP prioritizes structural consistency, allowing models to classify documents correctly under a ZSL scenario. To benchmark this dataset, we developed a series of Siamese Neural Networks (SNN) with various backbone architectures, including ResNet, variants such as ResNet, EfficientNet, ViT and others. As a result, the proposed ZSL-based method achieves Equal Error Rates (EERs) under 5%. The code of the proposed method is available at https://github.com/AIMLab-uhd/zsl.

I. Introduction

Document understanding, vital for businesses, automates information extraction and interpretation from documents [1]. The prominence of this field within the business sector stems from the universal necessity for companies to manage documents. Therefore, an automated pipeline for document processing is widely adopted and often essential. For instance, companies that process personal documents, such as identification documents, frequently encounter low-quality images. Nevertheless, an automated pipeline remains more efficient in both cost and time than a manual verification pipeline. Consequently, most companies find machine learning a suitable technology for automation, as deterministic methods may struggle to account for the high variability of document quality.

In this context, the prevalent approach to address this problem involves traditional classification methods, often utilizing Deep Learning (DL). The problem has been widely studied in the field of document understanding as a simple classification task, mainly supported by the Ryerson Vision Lab Complex Document Information Processing (RVL-CDIP) dataset [2]. This dataset comprises 400,000 documents divided into 16 classes based on their purpose, such as letters and emails. RVL-CDIP has been a subject of extensive research for traditional classification, with some works achieving over 97% accuracy [3], [4] in specific scenarios. However, existing DL techniques frequently encounter changes in form and structure, and new document types are continuously generated. In such dynamic environments, new data must be acquired to retrain the classification model.

To overcome the limitations imposed by the need for labeled training data, ZSL emerges as the most appropriate paradigm. This approach defines a task where the classes in the test set are disjoint from those in the training set, thus forcing the model to generalize effectively to novel categories [5], [6]. Despite being a considerably under-researched area within document analysis, ZS-DIC provides valuable contributions.

Although RVL-CDIP is the most widely used dataset for document classification, other datasets are available. DocVQA [7] focuses on Visual Question Answering (VQA). DocLayNet [8] is specialized in layout segmentation, and CORD [9], SROIE [10], FUNSD [11], and XFUND [12] were created for information extraction. While these datasets advance their respective fields, their label characteristics prevent their effective use in ZS-DIC.

To address these challenges, we introduce LA-CDIP, a dataset designed for the ZS-DIC challenge. LA-CDIP is a document-image dataset that re-purposes the images from RVL-CDIP under a new classification scheme. Unlike RVL-CDIP, which classifies documents by their purpose (e.g., emails, letters), our dataset arranges them by their visual structure, grouping together documents that share a similar layout. This reorganization enables ZS-DIC by ensuring that each class exhibits high consistency in both visual and textual elements.

We propose a Visual Document Matching (VDM) framework for this dataset. This framework leverages Contrastive Learning (CL) [13] to generate a similarity score between two documents. A predefined similarity threshold is used to determine if documents belong to the same class. In a practical application, by providing a reference for each class, the framework can be deployed in two modes: as a binary classifier for verification (comparing each incoming document with a single class reference) or as a multiclass classifier for identification (comparing each incoming document with all available references to select the most similar one).

The remainder of this paper is structured as follows. Section II details our LA-CDIP dataset. Section III describes the proposed Siamese network modeling. Sections IV and Sec-

SIBGRAPI

Obrigado!

Lucas de Almeida Bandeira Macedo
lucasabmacedo@hotmail.com

Orientador: Prof. Dr. Pedro Garcia Freitas
Coorientador: Prof. Dr. Bruno Luiggi Macchiavello Espinoza