

Layout-Aware Zero-Shot Learning for Visual Document Matching

Qualificação de Mestrado

Lucas de Almeida Bandeira Macedo

Universidade de Brasília
Departamento de Ciência da Computação

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

Outubro de 2025

1 Introdução

2 Metodologia

3 Resultados

4 Conclusão

Introdução

Contexto - Documentos e Compliance

- Documentos físicos
- Imagens de documentos
- Exemplos:

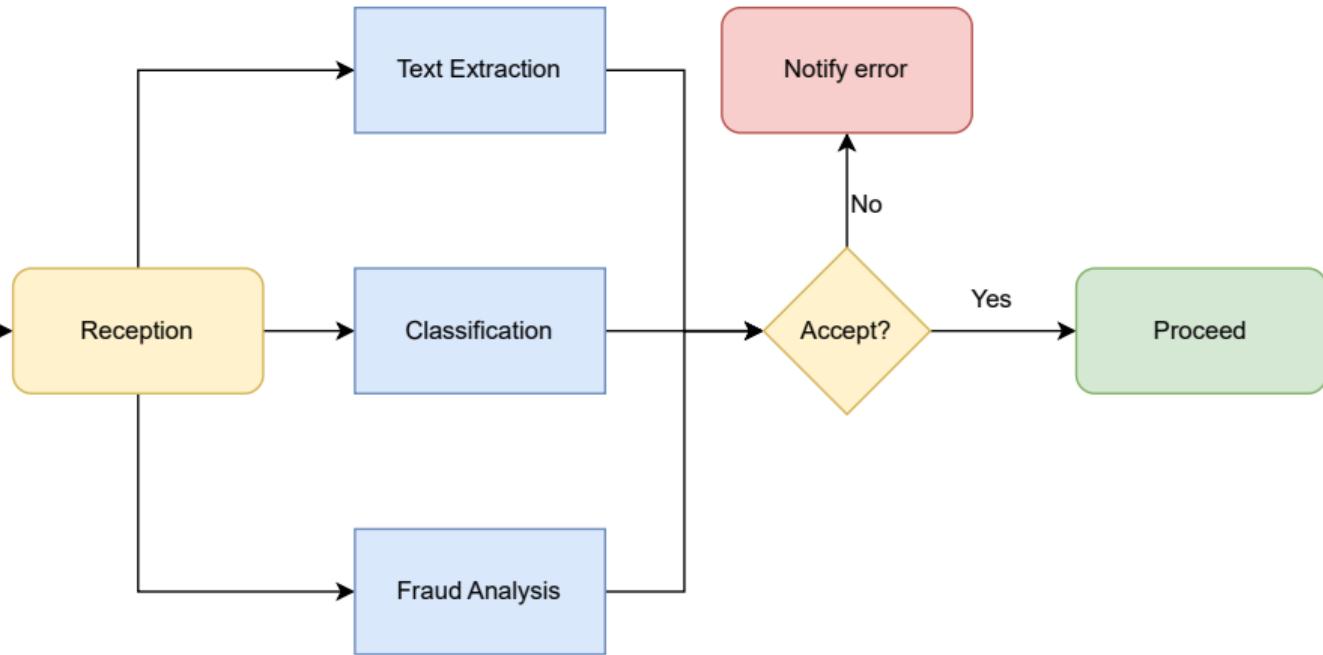


MINISTÉRIO DA FAZENDA SECRETARIA DA RECEITA FEDERAL DO BRASIL IMPOSTO SOBRE A RENDA - PESSOA FÍSICA EXERCÍCIO 2017 ANO-CALENDÁRIO 2016			
RECEBO DE ENTREGA DA DECLARAÇÃO DE AJUSTE ANUAL - OPÇÃO PELO DESCONTO SIMPLIFICADO DECLARAÇÃO ORIGINAL			
IDENTIFICAÇÃO DO DECLARANTE			
CPF do declarante 332.653.378-08	Nome do declarante ERIKA TOMAZELLA	Telefone (11) 43791221	
Endereço RUA RUA JUSTINO ALVES BATISTA VILA VILANDA		Número 89	Complemento AP 64 BL P
Bairro/ Distrito VILA VILANDA	CEP 06199-120	Município OGASCO	UF / SP (Valores em Reais)
TOTAL RENDIMENTOS TRIBUTÁVEIS		62.260,42	
IMPOSTO DEVIDO		3.560,59	
IMPOSTO A RESTITUIR		781,16	

Contexto - Processamento de Documentos



Actor

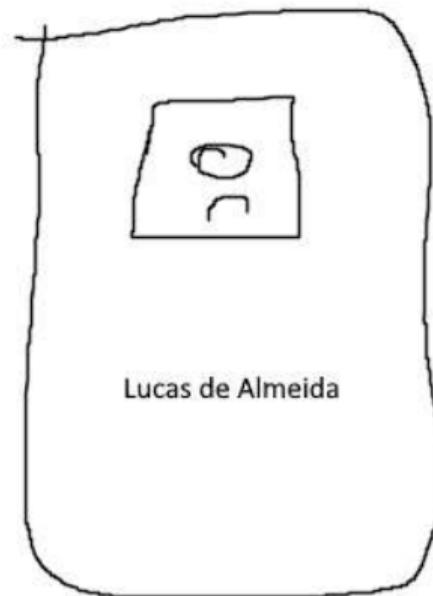


Contexto - Importância da Classificação da Imagem

- Assegurar que o documento está correto
- Documentos não-digitais
- Evita fraudes

Contexto - Importância da Classificação da Imagem

- Assegurar que o documento está correto
- Documentos não-digitais
- Evita fraudes



Classificação Tradicional:

- Categorização em classes predefinidas
- Cross-Entropy Loss

Desempenho Atual:

- Bakkali et al. (2021): 97.70% de acurácia no RVL-CDIP

Classificação Tradicional:

- Categorização em classes predefinidas
- Cross-Entropy Loss

Desempenho Atual:

- Bakkali et al. (2021): 97.70% de acurácia no RVL-CDIP

O Problema:

- Novos layouts de documentos
- Classes completamente novas
- Necessidade de retreinamento
- Semanas/meses de engenharia de dados e treinamento

Zero-Shot Learning

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

Desafios

- Falta de dataset especializado
 - Imagens de Documento
 - Generalização
 - Divisão treino e teste zero-shot
- Ausência de metodologia estado-da-arte
 - Paradigma ZSL
 - Capacidade de classificar

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
- Comparação com LLM

Metodologia

Datasets Disponíveis

- PubLayNet, DocLayNet
- DocVQA
- CORD, SROIE
- RVL-CDIP

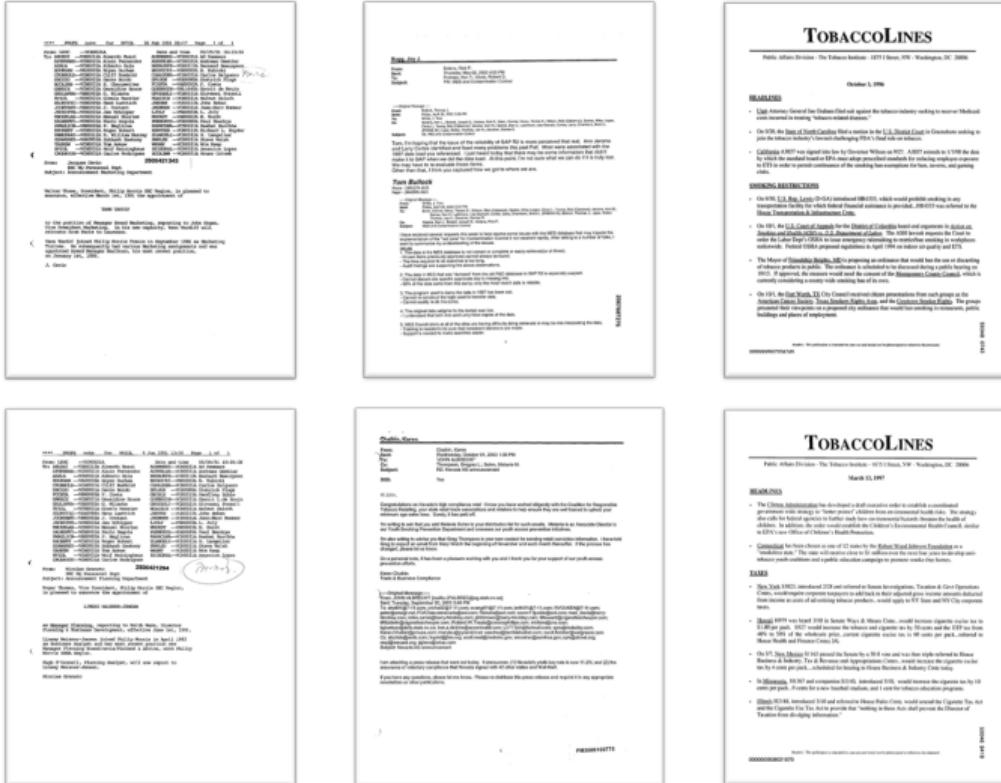
Datasets Disponíveis

- PubLayNet, DocLayNet
- DocVQA
- CORD, SROIE
- RVL-CDIP

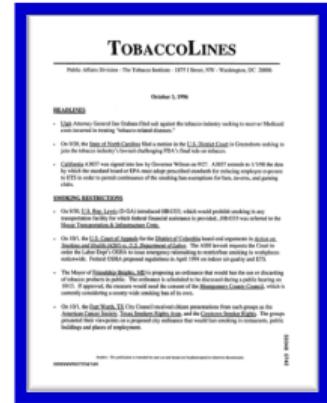
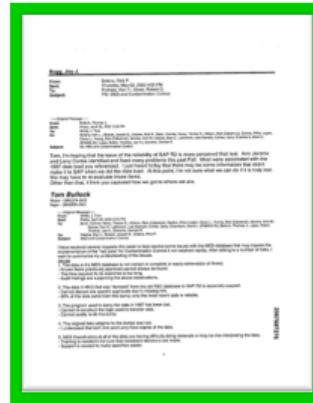
RVL-CDIP

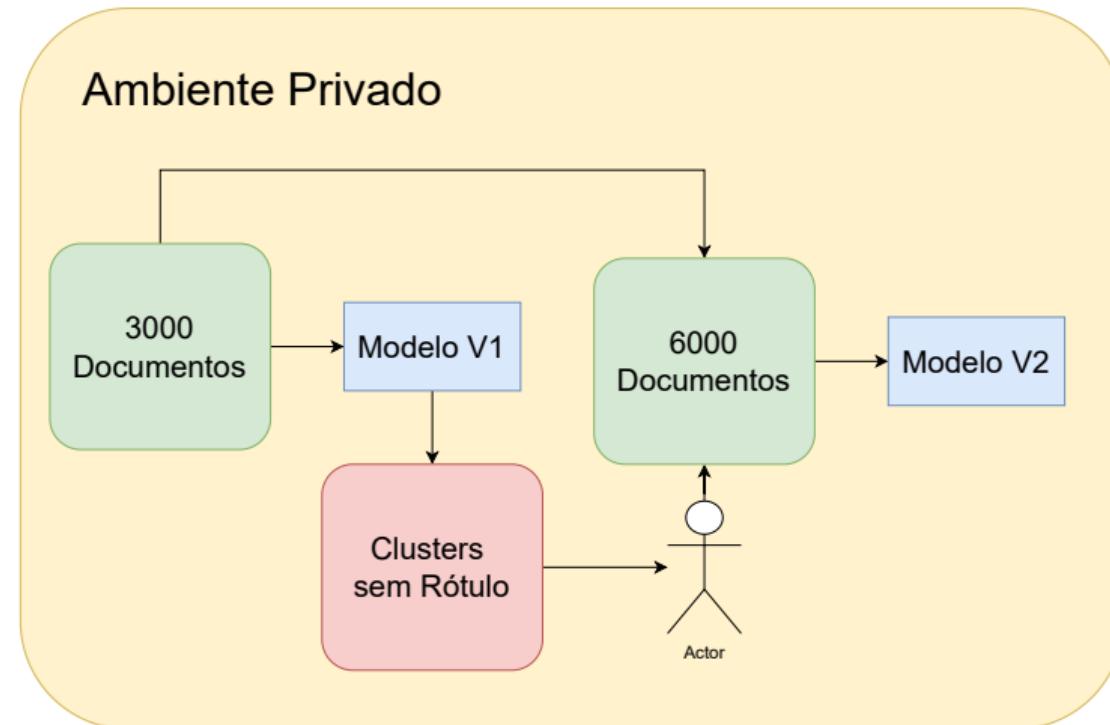
- 400.000 documentos
- 16 classes
- email, formulário, carta...
- Separado por função

Dataset - Motivação

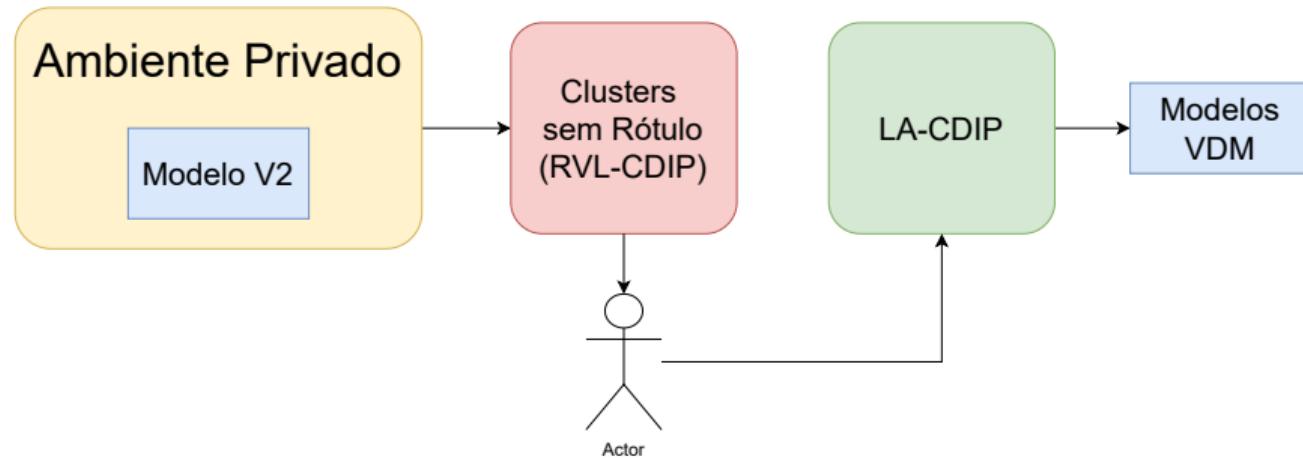


Dataset - Motivação

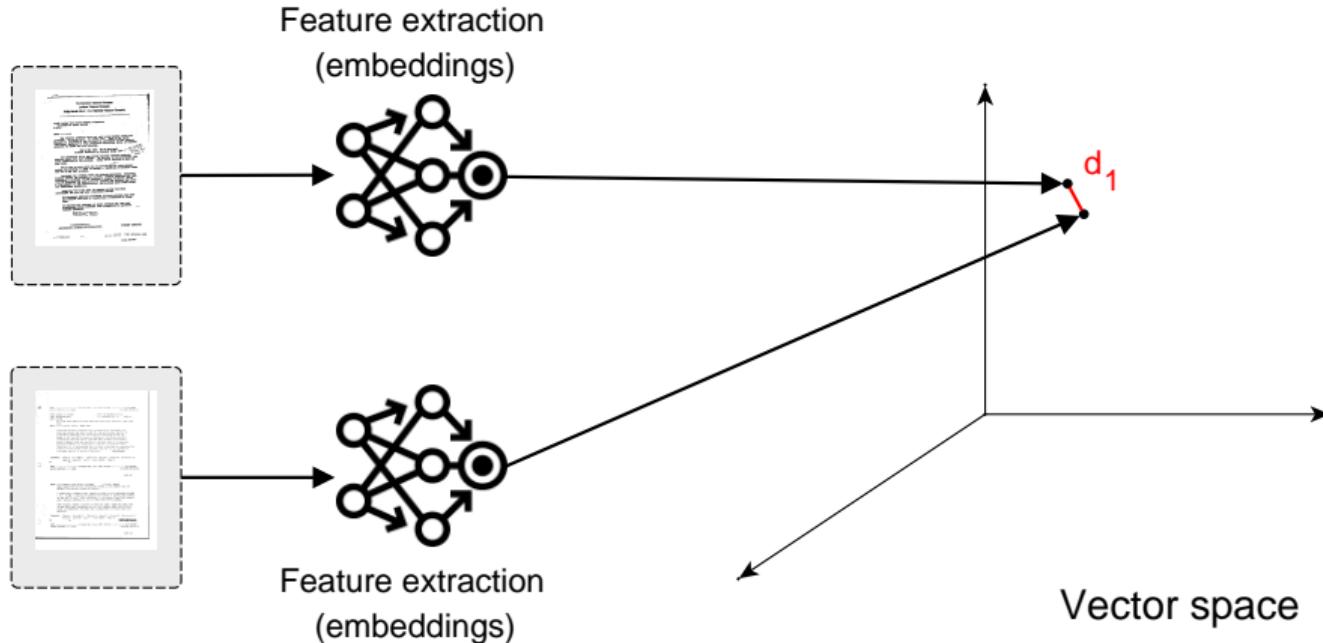




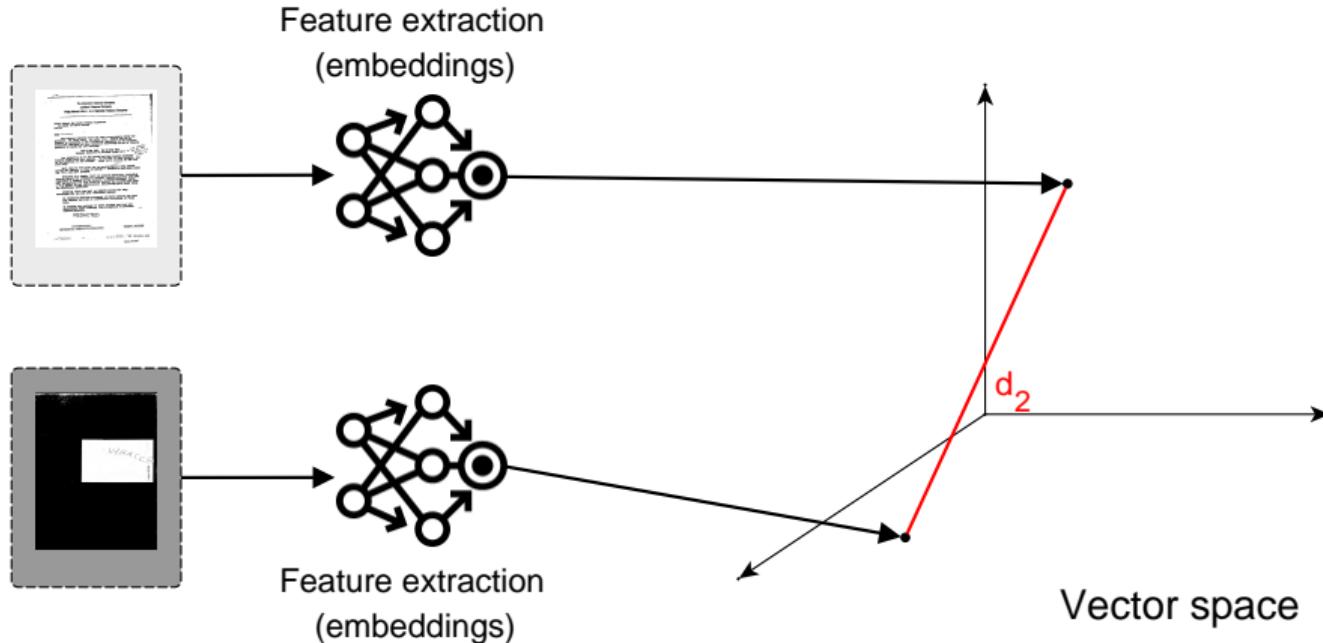
LA-CDIP Dataset - Processo de Rotulação

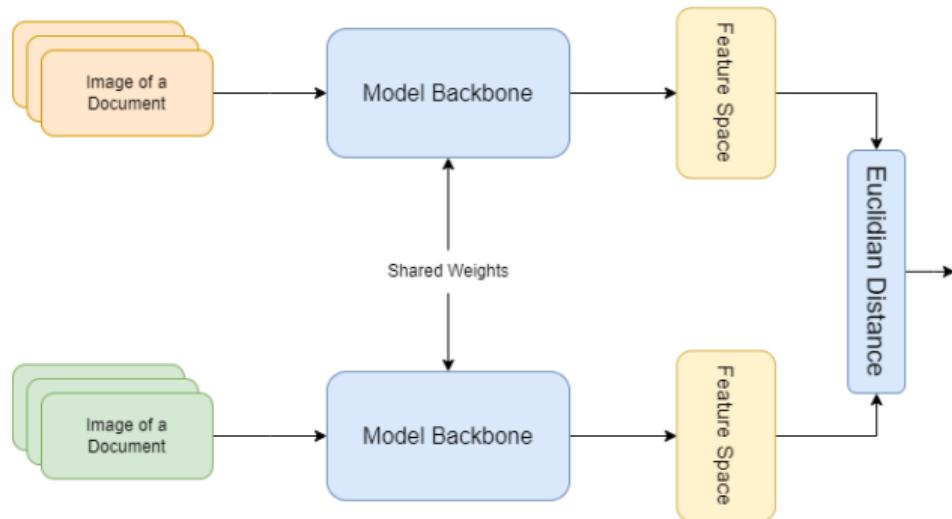


Visual Document Matching - Ilustração



Visual Document Matching - Ilustração





$$\mathcal{L} = \begin{cases} d(x_1, x_2)^2, & \text{if } y = 1 \\ (m - d(x_1, x_2))^2, & \text{otherwise.} \end{cases}$$

Model Backbone

Model Backbone



- AlexNet
- ResNet
- VGG
- EfficientNet
- MobileNetV3
- Vision Transformer (ViT)

Modelos Avaliados:

- LLaVA 3.2 Vision
- InternVL 2.5
- Qwen2.5-VL
- GPT-4o (2024-11-20)
- GPT-4o-mini (2024-07-18)

Avaliação:

- Zero-shot (sem fine-tuning)
- Pontuação de similaridade 0–100
- 5 níveis de categorização

Benchmarking com LLMs - Exemplos

Reference Image

2020 MARLBORO BAR PROGRAM
CONTRACT TOP SHEET

GMM/SSM: Jimmy Durick
MARKET: Durles
VENUE NAME: Blackberry's

VENUE ID
DAL-0121-03

Please check the appropriate box that will identify the type of club and the appropriate contract executed by club owner/manager:
 EVENT VISIBILITY MUSIC RNB

Please check the appropriate box regarding venue admission policy:
 AQ18-P AQ21-P PAO-P
 AD18-V AQ21-V

GMM/SSM SIGNATURE: Jimmy Durick DATE: 2-3-00

SELL-IN APPROVAL: X DATE: 2-3-00

Image to Compare

2020 MARLBORO BAR PROGRAM
CONTRACT TOP SHEET

GMM/SSM: Lindy Jenkins
MARKET: Charlotte
VENUE NAME: Clock's Sports Bar & Grill

VENUE ID
LHA-0080-01

Please check the appropriate box that will identify the type of club and the appropriate contract executed by club owner/manager:
 EVENT VISIBILITY MUSIC RNB

Please check the appropriate box regarding venue admission policy:
 AQ18-P AQ21-P PAO-P
 AD18-V AQ21-V

GMM/SSM SIGNATURE: Lindy Jenkins DATE: 2-20-00

SELL-IN APPROVAL: J DATE: 3-13-00

Similarity Score: 98
Category: Nearly Identical

Reference Image

2020 MARLBORO BAR PROGRAM
CONTRACT TOP SHEET

GMM/SSM: Jimmy Durick
MARKET: Durles
VENUE NAME: Blackberry's

VENUE ID
DAL-0121-03

Please check the appropriate box that will identify the type of club and the appropriate contract executed by club owner/manager:
 EVENT VISIBILITY MUSIC RNB

Please check the appropriate box regarding venue admission policy:
 AQ18-P AQ21-P PAO-P
 AD18-V AQ21-V

GMM/SSM SIGNATURE: Jimmy Durick DATE: 2-3-00

SELL-IN APPROVAL: X DATE: 2-3-00

Image to Compare

Action T:N Request

Reference Action Network, 1875 Eye Street, N.W., Washington, D.C. 20006 800-424-9870

March 9, 1981
C W-P-Please give O.K. to Kelly office
TO: TAN Corporate Coordinator Bob Shaw to A.D.C.
Mr. Charles I. McCarty Mr. Shepard P. Pollio TAC S-10-R1
Mr. V.E.S. Bey, Jr. Mr. Edward A. Morrison, Jr.
Mr. Curtis H. Judge Mr. Manuel Leitao, Jr.

FROM: Jack Kelly done
RE: Maine Legislation - L.D. 195, L.D. 509

BACKGROUND
Two bills currently remain under consideration in the State of Maine: L.D. 395 and L.D. 509. L.D. 246 was withdrawn by its sponsor, Senator Barbara Gill. A third piece of legislation may be added in the next month.

On February 17, 1981 a public hearing was held by the Joint Health & Institutional Services Committee on the remaining bills. After a lengthy debate on the action taken at the March 5, 1981 in a work session, the committee voted 4 to 3 against L.D. 395. However, a minority report written by the committee chairman, Senator Barbara Gill, will be presented to the legislature.

You have previously approved action by TAN members in the state to impact on this legislation. This second request is designed to seek an amendment to our efforts to defeat the pending legislation in the State of Maine.

ACTION REQUESTED
At this time, we request that the State Director be given permission to take your organization, TAN members in the state and request that they take the following action as their own issue:

1. Personally solicit a minimum of 60 signatures on a public smoking petition (See Attachment A) and return petition to the State Director by March 31, 1981.

Similarity Score: 15
Category: Completely Different

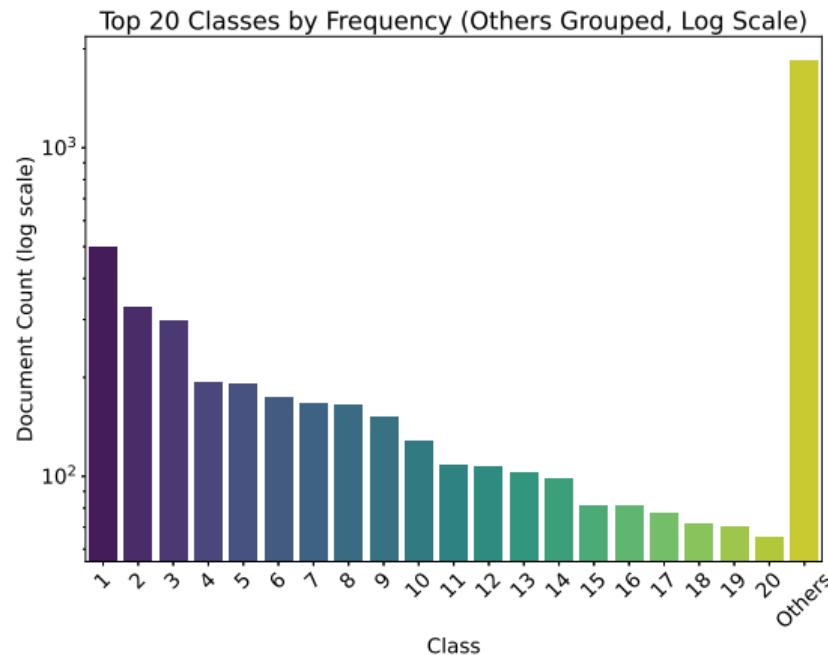
Resultados

Composição:

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

Splits:

- ZSL: separação completa treino/teste
- GZSL: 50% overlap de classes
- 5-fold cross-validation



Equal Error Rate (EER)

- Ponto onde FAR = FRR
- FAR: False Acceptance Rate
- FRR: False Rejection Rate

$$FAR(\tau) = \frac{\text{False Acceptances}}{\text{Total Negatives}}$$

$$FRR(\tau) = \frac{\text{False Rejections}}{\text{Total Positives}}$$

Protocolo de Teste:

Para cada documento: 1 par similar + 1 par dissimilar

Resultados - Visão Geral

Architecture	Edition	Params	ZSL	GZSL	Test ZSL	Test GZSL
AlexNet		57M	8.92	5.45	17.33	6.31
VGG	11	129M	7.47	5.01	14.24	3.95
	13	129M	7.03	4.79	9.30	3.95
	16	134M	8.29	5.23	14.74	4.82
	19	139M	7.30	4.57	17.08	3.90
ResNet	18	11M	5.03	1.54	4.98	1.51
	34	21M	4.32	2.10	4.13	1.53
	50	23M	6.90	3.39	10.34	2.21
	101	42M	8.20	2.72	11.31	1.98
	152	58M	9.44	3.38	12.70	2.39
MobileNetV3	Small	1M	7.98	5.06	12.74	5.26
	Large	4M	8.16	4.27	8.45	4.43
EfficientNet	0	4M	4.41	2.27	6.02	0.95
	1	6M	3.93	3.54	8.88	2.70
	2	7M	5.73	2.61	7.29	2.14
	3	10M	5.65	3.64	7.37	2.34
ViT	Base	87M	12.43	7.97	19.72	5.19
	Large	305M	13.16	7.57	19.88	5.26
Llama	3.2	11B	—	—	13.95	21.90
InternVL	2.5	8B	—	—	8.58	10.40
Qwen-VL	2.5	7B	—	—	6.61	4.20
GPT-4o mini	2024-07-18	*	—	—	4.70	4.07
GPT-4o	2024-11-20	*	—	—	2.75	1.33

Resultados - ResNet e EfficientNet

Architecture	Edition	Params	ZSL	GZSL	Test ZSL	Test GZSL
ResNet	18	11M	5.03	1.54	4.98	1.51
	34	21M	4.32	2.10	4.13	1.53
	50	23M	6.90	3.39	10.34	2.21
	101	42M	8.20	2.72	11.31	1.98
	152	58M	9.44	3.38	12.70	2.39
EfficientNet	0	4M	4.41	2.27	6.02	0.95
	1	6M	3.93	3.54	8.88	2.70
	2	7M	5.73	2.61	7.29	2.14
	3	10M	5.65	3.64	7.37	2.34

Resultados - ResNet, EfficientNet e ViT

Architecture	Edition	Params	ZSL	GZSL	Test ZSL	Test GZSL
ResNet	18	11M	5.03	1.54	4.98	1.51
	34	21M	4.32	2.10	4.13	1.53
	50	23M	6.90	3.39	10.34	2.21
	101	42M	8.20	2.72	11.31	1.98
	152	58M	9.44	3.38	12.70	2.39
EfficientNet	0	4M	4.41	2.27	6.02	0.95
	1	6M	3.93	3.54	8.88	2.70
	2	7M	5.73	2.61	7.29	2.14
	3	10M	5.65	3.64	7.37	2.34
ViT	Base	87M	12.43	7.97	19.72	5.19
	Large	305M	13.16	7.57	19.88	5.26

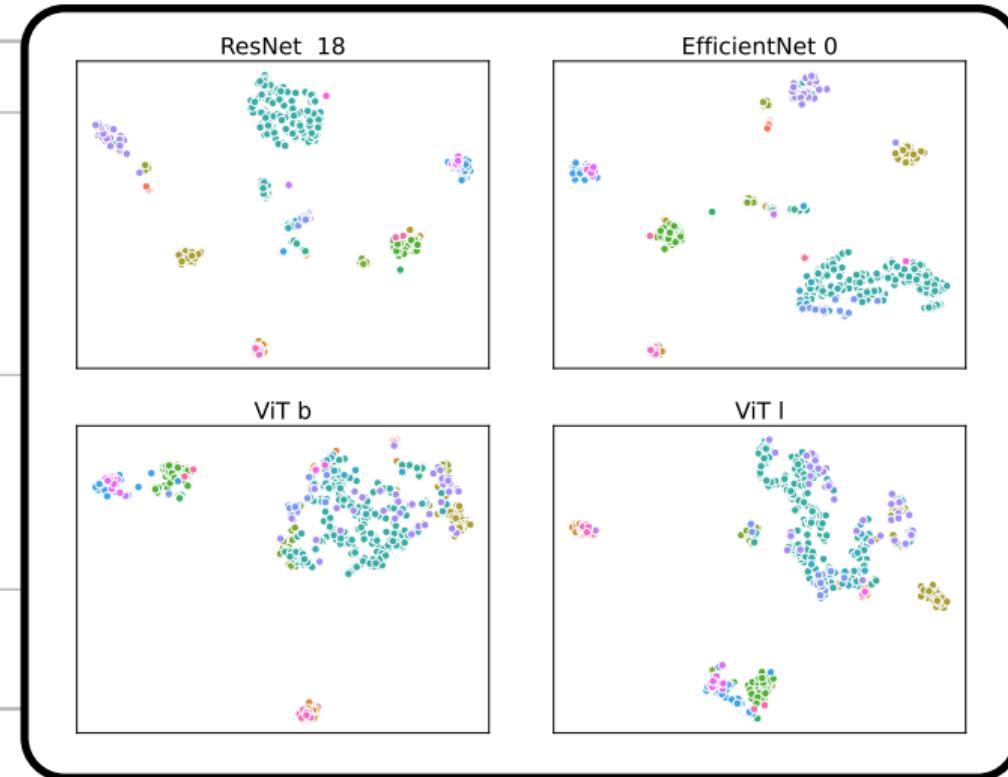
Resultados - Visualização TSNE

Architecture

ResNet

EfficientNet

ViT



Test GZSL

1.51
1.53
2.21
1.98
2.39

0.95
2.70
2.14
2.34

5.19
5.26

Resultados - Large Language Models

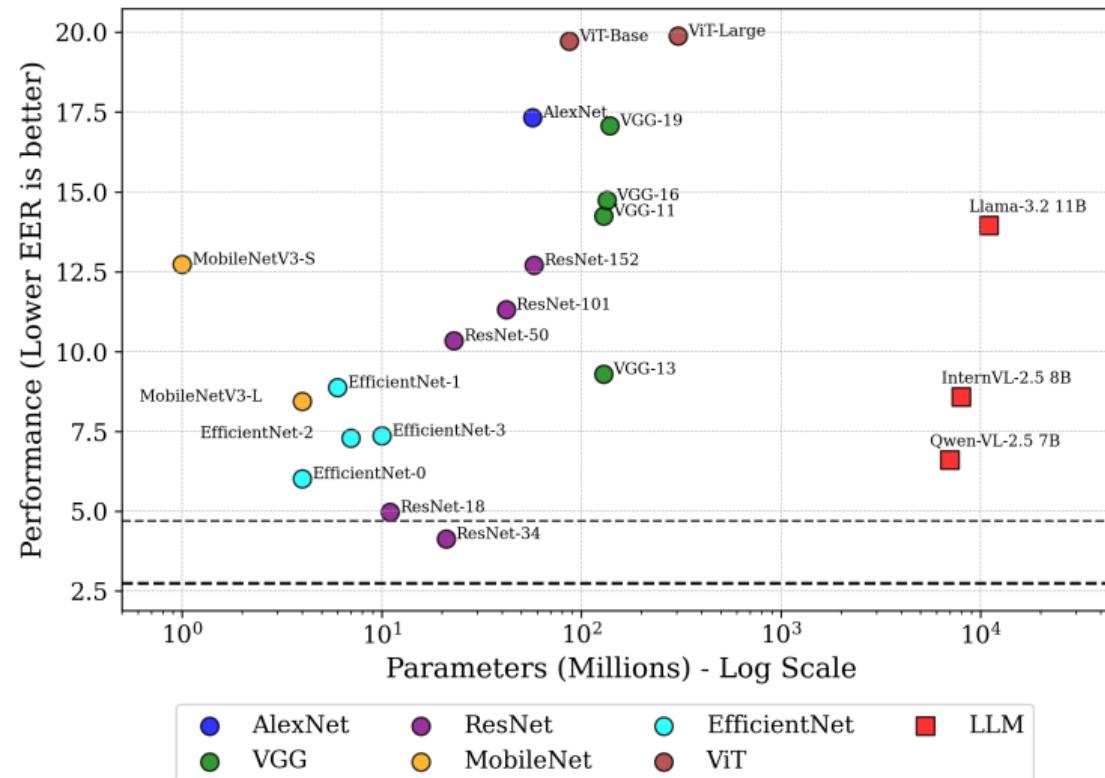
Model	Version	Params	Test ZSL	Test GZSL
Llama	3.2	11B	13.95	21.90
InternVL	2.5	8B	8.58	10.40
Qwen-VL	2.5	7B	6.61	4.20
GPT-4o mini	2024-07-18	*	4.70	4.07
GPT-4o	2024-11-20	*	2.75	1.33

* Parameter count not publicly disclosed

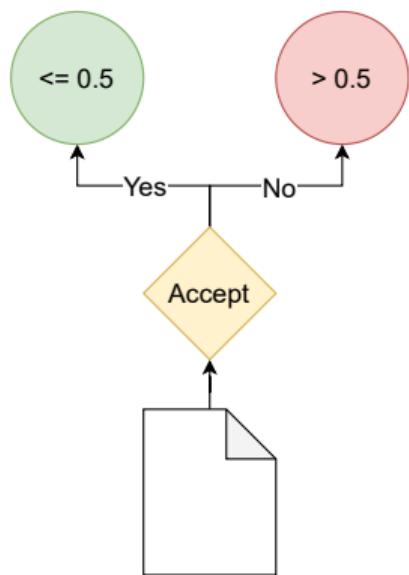
Resultados - Melhores Modelos Visuais vs LLMs

Architecture	Edition	Params	ZSL	GZSL	Test ZSL	Test GZSL
ResNet	18	11M	5.03	1.54	4.98	1.51
	34	21M	4.32	2.10	4.13	1.53
	50	23M	6.90	3.39	10.34	2.21
	101	42M	8.20	2.72	11.31	1.98
	152	58M	9.44	3.38	12.70	2.39
EfficientNet	0	4M	4.41	2.27	6.02	0.95
	1	6M	3.93	3.54	8.88	2.70
	2	7M	5.73	2.61	7.29	2.14
	3	10M	5.65	3.64	7.37	2.34
Llama	3.2	11B	–	–	13.95	21.90
InternVL	2.5	8B	–	–	8.58	10.40
Qwen-VL	2.5	7B	–	–	6.61	4.20
GPT-4o mini	2024-07-18	*	–	–	4.70	4.07
GPT-4o	2024-11-20	*	–	–	2.75	1.33

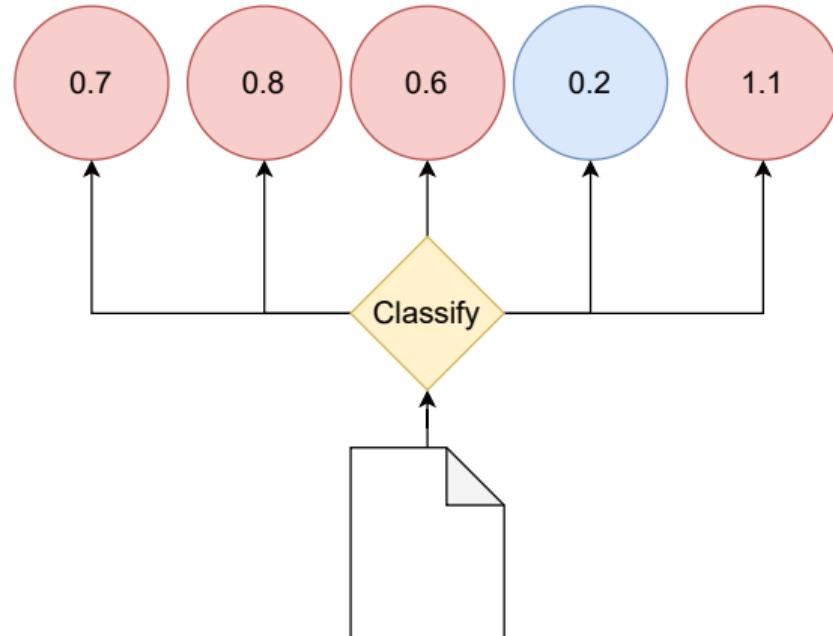
Comparação: Visual Models vs LLMs



Comparação: Visual Models vs LLMs



Sistema de Verificação



Sistema de Identificação

Conclusão

Principais Contribuições

① LA-CDIP Dataset

- Dataset categorizado exclusivamente por layout
- Alternativa Zero-Shot para classificação de documentos

② Benchmarking Sistemático

- Diversos backbones visuais estabelecidos
- Comparação com LLMs populares

③ Resultados Práticos

- Modelos visuais menores superam LLMs (exceto GPT-4o)
- GPT-4o melhor, mas custo-benefício desfavorável
- Alternativas open-source viáveis

Limitações Conhecidas:

- Dataset relativamente pequeno
- Complexidade de arquiteturas limitada
- Apenas documentos do RVL-CDIP

Trabalhos Futuros:

- Aumentar número de amostras
- Aumentar número de classes
- Incluir fontes adicionais de documentos
- Data augmentation
- Permitir modelos mais complexos

Timeline da Pesquisa

Tarefa	5º	6º	7º	8º	9º–10º
Literatura	✓				
Framework	✓	✓			
Dataset Labeling		✓			
Experimentos		✓	✓		
Produção de Paper				✓	✓
Extra Labeling					✓
Qualificação					8º trim
Defesa de Mestrado					10º trim

Status atual: **8º trimestre (Qualificação)**

Obrigado!

Perguntas?

Lucas de Almeida Bandeira Macedo

lucasabmacedo@hotmail.com

Orientador: Prof. Dr. Pedro Garcia Freitas

Coorientador: Prof. Dr. Bruno Luiggi Macchiavello Espinoza