

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 Luigi Macchiavello Espinoza

Outubro de 2025

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

Introdução

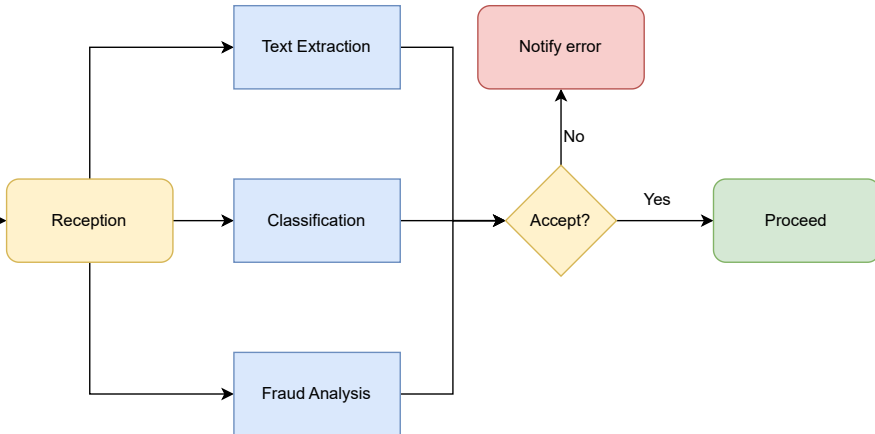
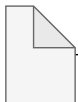
Contexto - Documentos e Compliance

- Ambiente Bancário
- Documentos físicos
- Imagens de documentos
- Exemplos:



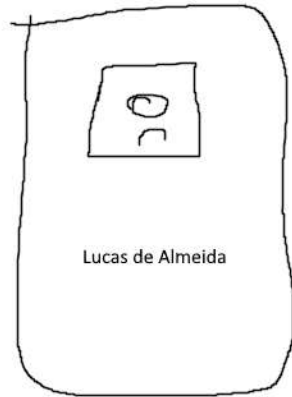
MINISTÉRIO DA FAZENDA		IMPOSTO SOBRE A RENDA - PESSOA FÍSICA	
SECRETARIA DA RECEITA FEDERAL DO BRASIL		EXERCÍCIO 2017 ANO-CALENDÁRIO 2016	
RECIBO DE ENTREGA DA DECLARAÇÃO DE AJUSTE ANUAL - OPÇÃO PELO DESCONTO SIMPLIFICADO DECLARAÇÃO ORIGINAL			
IDENTIFICAÇÃO DO DECLARANTE			
CPF do declarante 352.056.378-08		Nome do declarante ERIKA TOMAZELLA	
Endereço RUA RUA JUSTINO ALVES BATISTA		Número 99	Complemento AP 64 Bl. P
Bairro/Cidade VILA YOLANDA	CEP 06126-120	Município DASCO	UF SP
TOTAL RENDIMENTOS TRIBUTÁVEIS			(Valores em Reais) 62.260,42
IMPOSTO DEVIDO			3.562,56
IMPOSTO A RESTITUIR			781,16

Contexto - Fluxo de Compliance



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

- 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

Lucas de Almeida Bandeira Macedo (UnB)

Walter Thomas, President, World Health Organization, is pleased to announce, effective March 1st, 1961, the appointment of

DAVID DAVIES

to the position of Managing General Marketing, replacing to John Hughes-Jones, Marketing, in his new capacity, Sam Woodard will relocate from Kuala Lumpur.

Sam Woodard joined Health Service France in September 1958 as Marketing and Administration staff member Marketing Department. He has been and is appointed General Manager Marketing, his most recent position, on January 1st, 1961.

A. Davies

an Strategic Planning, reporting to David Mann, Director Planning & Business Development, effective June 1st, 1991.

George Delaney-Jones joined Philip Morris in April 1992 as Business Analyst, and has been given primary responsibility for Strategic Planning Development & Financial & Admin. with Philip Morris Sales Region.

High Potential, Planning Analyst, will now report to George Delaney-Jones.

Michael Stroud

[illegible][illegible]

Public Affairs Division, The National Institute, 14711 Route 520, Washington, DC 20004

October 5, 1998

REFERENCES

- [illegible]

Notas: Se utilizó una α nominal de 0.05 en todos los tests estadísticos.

Public Affairs Division - The Tobacco Institute - 1175 I Street, SW - Washington, DC - 20005

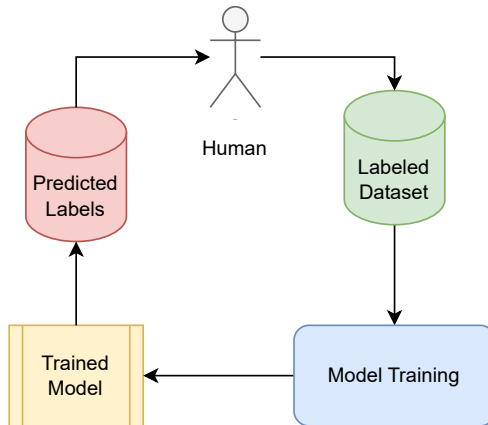
March 13, 1997

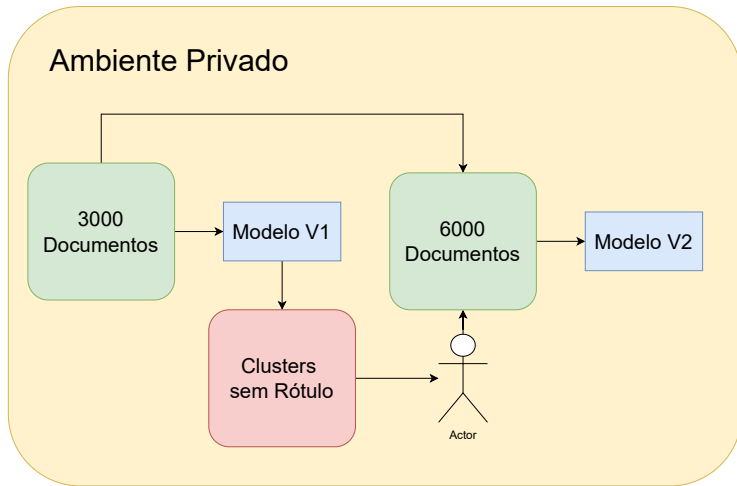
REFERENCES

- [illegible]

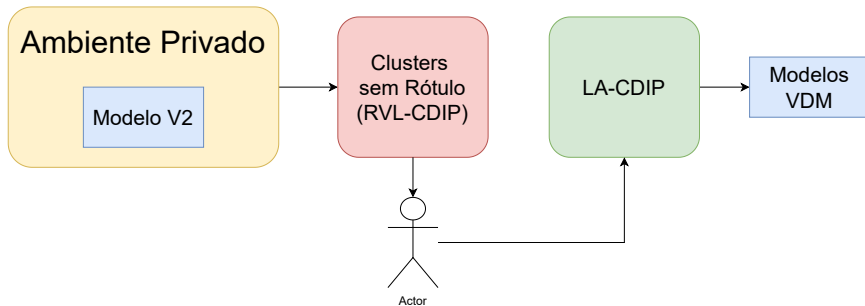
Source: The authors's qualitative research and their own photographs of the landscape.

LA-CDIP Dataset - Active Learning

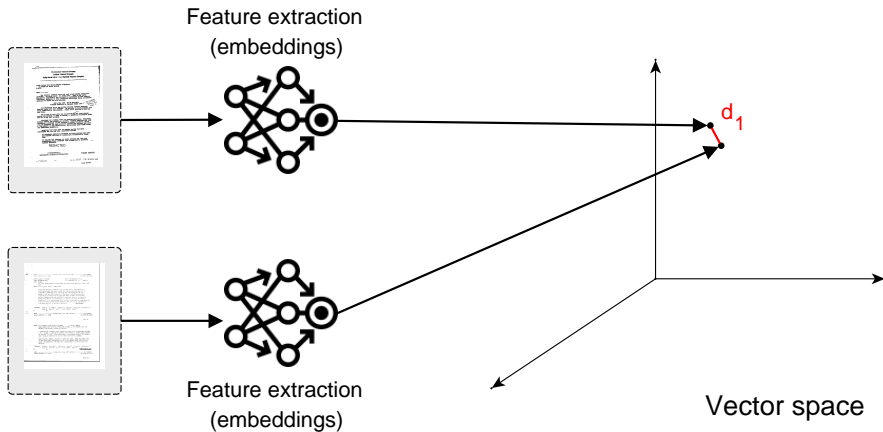




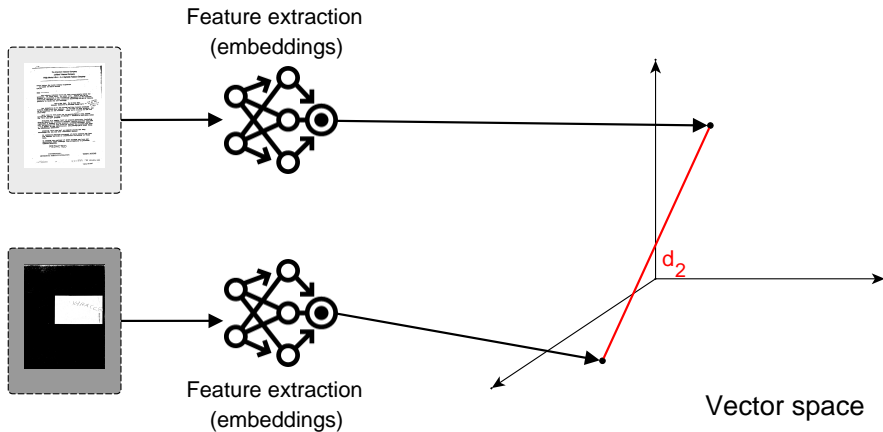
LA-CDIP Dataset - Rotulação



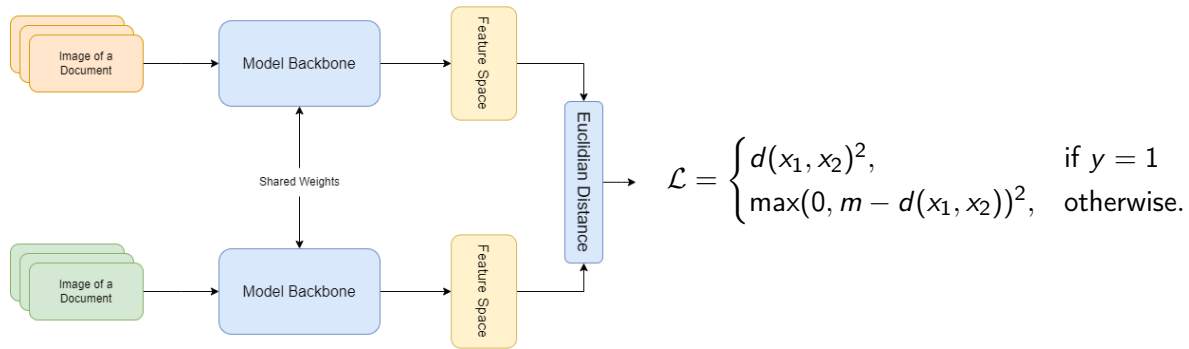
Visual Document Matching



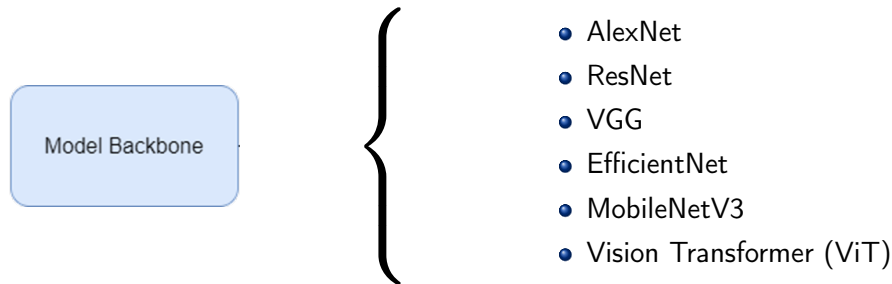
Visual Document Matching



Visual Document Matching - Arquitetura



Visual Document Matching - Backbone



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

Reference Image

2000 MARLBORO BAR PROGRAM
CONTRACT TOP SHEET

GMM/SSM: Ameyrick
MARKET: Dulles
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:
☐ AO18-P ☐ AO21-P ☐ PAQ-P
☐ AO18-V ☒ AO21-V

GMM/SSM SIGNATURE: [Signature] DATE: 2-2-00
SELLER APPROVAL: X DATE:

20080319001

Similarity Score: 98
Category: Nearly Identical

Image to Compare

2000 MARLBORO BAR PROGRAM
CONTRACT TOP SHEET

GMM/SSM: Cindy Jenkins
MARKET: Charlotte
VENUE NAME: Bocky's Sports Bar & Grill

VENUE ID
CHA-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:
☐ AO18-P ☐ AO21-P ☒ PAQ-P
☐ AO18-V ☐ AO21-V

GMM/SSM SIGNATURE: Cindy Jenkins DATE: 2-21-00
SELLER APPROVAL: [Signature] DATE: 3-13-00

2008032177

Reference Image

2000 MARLBORO BAR PROGRAM
CONTRACT TOP SHEET

GMM/SSM: [Signature]
MARKET: [Signature]
VENUE NAME: [Signature]

VENUE ID
[Signature]

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:
☐ AO18-P ☐ AO21-P ☐ PAQ-P
☐ AO18-V ☐ AO21-V

GMM/SSM SIGNATURE: [Signature] DATE: [Signature]
SELLER APPROVAL: [Signature] DATE: [Signature]

2008032177

Similarity Score: 15
Category: Completely Different

Image to Compare

Action T-N Request

Rebacco Action Network, 1875 Eye Street, N.W., Washington, D.C. 20006 800-424-5870

March 9, 1981

TO: TAN Corporate Coordinator Donna A. R.T.S.
Mr. Charles J. McCarty Mr. Shepard P. Pollack
Mr. K.V.B. Day, Jr. Mr. Edward A. Morrison, Jr.
Mr. Curtis H. Judge Mr. Manuel Leitao, Jr.

FROM: Jack Kelly [Signature]

RE: Maine Legislation - L.D. 395, 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 on February 17, 1981. 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. The Committee took no action at that time. On March 5, 1981 in a work session, the committee voted 4 to 3 against L.D. 509. However, a minority report written by the committee chairman, Senator Barbara Gili, 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 add an additional dimension 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 contact your company's TAN members in the state and request that they take the following action on their own time:

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.

CONCISE

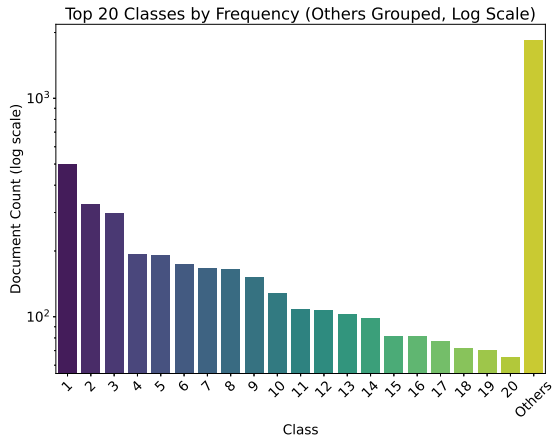
Resultados

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	<u>1.54</u>	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	<u>0.95</u>
	1	6M	<u>3.93</u>	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	*	—	—	<u>2.75</u>	<u>1.33</u>

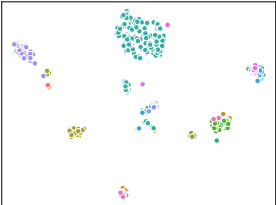
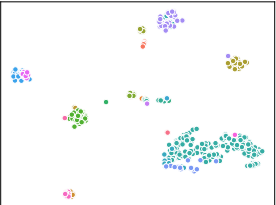
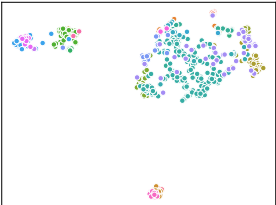
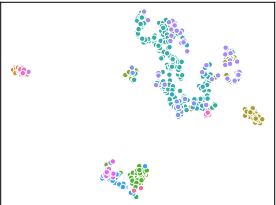
Resultados - ResNet e EfficientNet

Architecture	Edition	Params	ZSL	GZSL	Test ZSL	Test GZSL
ResNet	18	11M	5.03	<u>1.54</u>	4.98	1.51
	34	21M	4.32	2.10	<u>4.13</u>	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	<u>0.95</u>
	1	6M	<u>3.93</u>	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	<u>1.54</u>	4.98	1.51
	34	21M	4.32	2.10	<u>4.13</u>	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	<u>0.95</u>
	1	6M	<u>3.93</u>	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			Test GZSL
ResNet			1.51
			1.53
			2.21
			1.98
			2.39
EfficientNet			0.95
			2.70
			2.14
			2.34
			5.19
ViT			5.26

Resultados - Large Language Models

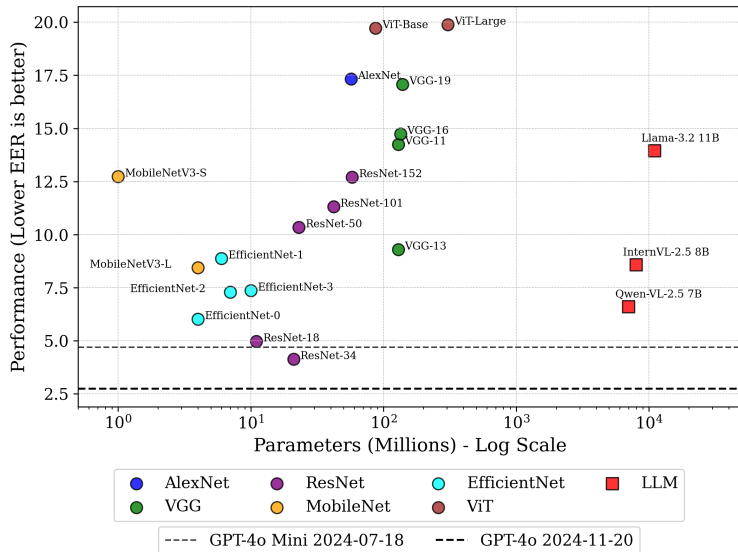
Model	Version	Params	Test ZSL	Test GZSL
Llama	3.2	11B	13.95	21.90
InternVL	2.5	8B	8.58	10.40
Qwen-VL	2.5	7B	6.61	4.20
GPT-4o mini	2024-07-18	*	4.70	4.07
GPT-4o	2024-11-20	*	<u>2.75</u>	<u>1.33</u>

* Parameter count not publicly disclosed

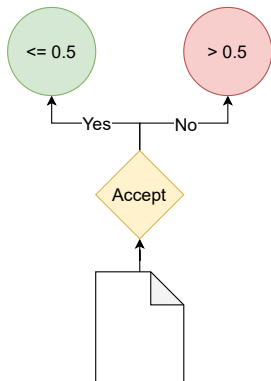
Resultados - Melhores Modelos Visuais vs LLMs

Architecture	Edition	Params	ZSL	GZSL	Test ZSL	Test GZSL
ResNet	18	11M	5.03	<u>1.54</u>	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	<u>0.95</u>
	1	6M	<u>3.93</u>	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	*	—	—	<u>2.75</u>	1.33

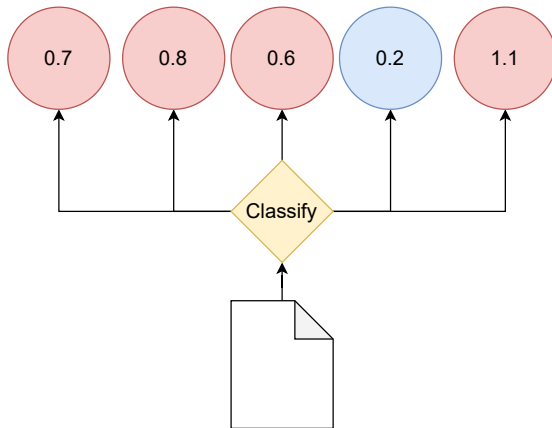
Comparação: Visual Models vs LLMs



Comparação: Visual Models vs LLMs



Sistema de Verificação



Sistema de Identificação

Conclusão

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 e LLMs

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

Obrigado!

Lucas de Almeida Bandeira Macedo
lucasabmacedo@hotmail.com

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