Projeto de Conclusão de Curso

Análise de Modelos zero-shot para Detecção de objetos.

Bruno Santa Cruz de Assunção

Área de Concentração: Visão Computacional
Orientador(a): Filipe Rolim Cordeiro

RECIFE, JULHO/2023.

Documento de Projeto de Pesquisa

1 Identificação

Aluno(a): Bruno Santa Cruz de Assunção(bruno.scassuncao@gmail.com)
Orientador(a): Filipe Rolim Cordeiro(filipe.rolim@ufrpe.br)
Título: Análise de Modelos zero-shot para Detecção de objetos.
Área de Concentração: Visão Computacional
Linha de Pesquisa: Processamento de imagem

2 Introdução

A detecção de objetos é uma área fundamental na visão computacional, que lida com instáncias de objetos de uma de diversas classes conhecidas, fazendo o trabalho de identificar e localizar de objetos em imagens e vídeos[1]. Esta tarefa, trás informações importantes para a análise de imagens, assim, se tornando base para diversas outras áreas da visão computacional, como reconhecimento de objetos, rastreamento de objetos e segmentação de objetos[2].

Com o rápido desenvolvimento das técnicas de *deep learning*, a detecção de objetos chegou a novos patamares[3], sendo aplicada em diversas áreas, como carros autônomos [4] e diagnósticos médicos[5]. Tradicionalmente, os datasets utilizados nos modelos de detecção de imagem modernos exigem grandes quantidades de dados rotulados para treinar suas redes neurais de forma eficaz. No entanto, para a criação destes datasets, exige a rotulagem manual que é um processo dispendioso e demorado[6][4].

Neste contexto, surgem os modelos de detecção de imagem *zero-shot*, que representam uma abordagem inovadora para superar as limitações impostas pela necessidade de dados rotulados. Esses modelos são capazes de identificar e classificar objetos em imagens sem a necessidade de terem sido previamente treinados com exemplos específicos dessas categorias[7]. Através da utilização de técnicas avançadas, como o aprendizado por transferência[8][9] e a incorporação de conhecimento semântico[10][11], os modelos *zero-shot* oferecem uma solução promissora para expandir a aplicabilidade da visão computacional em cenários onde a disponibilidade de dados é restrita.

3 Problema de Pesquisa

O aprendizado *zero-shot* é uma abordagem promissora para a detecção de objetos em imagens, mas ainda existem desafios significativos a serem superados. A eficácia

dos modelos *zero-shot* depende fortemente da qualidade dos vetores de características extraídos das imagens e da capacidade do modelo de generalizar para novas categorias de objetos. Além disso, a interpretabilidade dos modelos *zero-shot* é uma questão crítica, uma vez que a capacidade de explicar as decisões tomadas pelo modelo é essencial para a confiança e aceitação dos usuários[7].

Como o aprendizado *zero-shot* busca identificar objetos que não foram previamente observados durante o treinamento[12][13], é visto uma rápida ascenção no número de métodos deste tipo propostos todos os anos. Como resultado, a comparação entre os diferentes modelos existentes torna-se uma tarefa desafiadora, uma vez que os métodos propostos variam significativamente em termos de arquitetura, complexidade e desempenho[14].

Sendo assim, como podemos avaliar e comparar de forma eficaz os modelos *zero-shot* para detecção de objetos em imagens? Quais são as principais métricas e benchmarks utilizados para avaliar o desempenho desses modelos? Quais são os desafios e oportunidades para a pesquisa futura nesta área? Estas são algumas das questões que motivam esta proposta de pesquisa.

4 Justificativa

Parte significante do avanço das técnicas de detecção de imagem, é proveniente de competições feitas em cima de datasets públicos desafiadores, como o COCO[15] e Pascal VOC[16], tendo como principal métrica a mediana da precisão média (mAP).

Como apontado por Bolya et al.[17], há um grande problema em optimizar seu modelo para a mAP, que ao prioriza-la, podemos inadvertidamente deixar de lado a importância relativa de cada tipo de erro que podem variar entre aplicações, assim, entender como as fontes de erro afetam o mAP geral é crucial para desenvolver novos modelos e escolher o modelo correto para uma determinada aplicação.

Para a análise dos modelos, proponho a utilização da ferramenta TIDE, uma ferramenta de interpretação de detecção de objetos que: (i) sumariza os tipos de erro de forma compacta para fácil comparação; (ii) isola completamente a contribuição de cada tipo de erro, de forma que não haja variáveis conflitantes que afetem conclusões; (iii) não requer um tipo específico de anotações, assim, permitindo a comparação entre datasets; (iv) permite uma análise detalhada se desejado, para que a fonte de erro possa ser isolada. [17].

5 Objetivos

5.1 Objetivo Geral:

Analisar e comparar modelos *zero-shot* para detecção de objetos em imagens, avaliando a eficácia, interpretabilidade e generalização dos modelos em diferentes cenários

utilizado a ferramenta TIDE.

5.2 Objetivos Específicos:

1. Analizar a eficácia de diferentes modelos *zero-shot* para detecção de imagem.
2. Realizar uma análise comparativa entre modelos *zero-shot*.

6 Etapas de Pesquisa

Etapas:

1. Realizar revisão bibliográfica
2. Definir diferentes *datasets* para testes em diferentes cenários.
3. Aplicar diferentes modelos *zero-shot* nos diferentes cenários.
4. Analisar os resultados e comparar os modelos.
5. Escrita.
6. Defesa.

7 Cronograma

Atividades	2024-2025							
	Set	Out	Nov	Dez	Jan	Fev	Mar	Abr
Revisão bibliográfica								
Definição de bases de dados								
Aplicação de modelos								
Análise e comparação de resultados								
Escrita								
Defesa								

Referências

[1] Yali Amit, Pedro Felzenszwalb, and Ross Girshick. *Object Detection*, pages 875–883. Springer International Publishing, Cham, 2021.

[2] Zhengxia Zou, Keyan Chen, Zhenwei Shi, Yuhong Guo, and Jieping Ye. Object detection in 20 years: A survey. *Proceedings of the IEEE*, 111(3):257–276, 2023.

[3] Yann LeCun, Y. Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521:436–44, 05 2015.

[4] Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. The cityscapes dataset for semantic urban scene understanding, 2016.

[5] Hao Dong, Guang Yang, Fangde Liu, Yuanhan Mo, and Yike Guo. Automatic brain tumor detection and segmentation using u-net based fully convolutional networks, 2017.

[6] Shreya Shankar, Yoni Halpern, Eric Breck, James Atwood, Jimbo Wilson, and D. Sculley. No classification without representation: Assessing geodiversity issues in open data sets for the developing world, 2017.

[7] Ankan Bansal, Karan Sikka, Gaurav Sharma, Rama Chellappa, and Ajay Divakaran. Zero-shot object detection. In *Proceedings of the European Conference on Computer Vision (ECCV)*, September 2018.

[8] Guo-Jun Qi, Charu Aggarwal, Yong Rui, Qi Tian, Shiyu Chang, and Thomas Huang. Towards cross-category knowledge propagation for learning visual concepts. In *CVPR 2011*, pages 897–904, 2011.

[9] Yanwei Fu, Tao Xiang, Yu-Gang Jiang, Xiangyang Xue, Leonid Sigal, and Shaogang Gong. Recent advances in zero-shot recognition, 2017.

[10] Armand Joulin, Edouard Grave, Piotr Bojanowski, and Tomas Mikolov. Bag of tricks for efficient text classification, 2016.

[11] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space, 2013.

[12] Christoph H. Lampert, Hannes Nickisch, and Stefan Harmeling. Attribute-based classification for zero-shot visual object categorization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(3):453–465, 2014.

[13] Marcus Rohrbach, Michael Stark, and Bernt Schiele. Evaluating knowledge transfer and zero-shot learning in a large-scale setting. In *CVPR 2011*, pages 1641–1648, 2011.

[14] Yongqin Xian, Christoph H. Lampert, Bernt Schiele, and Zeynep Akata. Zero-shot learning – a comprehensive evaluation of the good, the bad and the ugly, 2020.

[15] Tsung-Yi Lin, Michael Maire, Serge Belongie, Lubomir Bourdev, Ross Girshick, James Hays, Pietro Perona, Deva Ramanan, C. Lawrence Zitnick, and Piotr Dollár. Microsoft coco: Common objects in context, 2015.

[16] Mark Everingham, Luc Van Gool, Christopher Williams, John Winn, and Andrew Zisserman. The pascal visual object classes (voc) challenge. *International Journal of Computer Vision*, 88:303–338, 06 2010.

[17] Daniel Bolya, Sean Foley, James Hays, and Judy Hoffman. Tide: A general toolbox for identifying object detection errors. In *ECCV*, 2020.