

Universidade Federal do ABC
Centro de Engenharia, Modelagem e Ciências Sociais Aplicadas
Trabalho de Graduação em Engenharia de Informação

Estudo e aplicação de redes neurais profundas na solução de detecção e reconhecimento de texto em cenas

Matheus Milani

**Maio de 2022
Santo André - SP**

Matheus Milani

Estudo e aplicação de redes neurais profundas na solução de detecção e reconhecimento de texto em cenas

Trabalho de Graduação apresentado para conclusão da Graduação em Engenharia de Informação, como parte dos requisitos necessários para a obtenção do Título de Bacharel em Engenharia de Informação.

Universidade Federal do ABC
Centro de Engenharia, Modelagem e Ciências Sociais Aplicadas
Trabalho de Graduação em Engenharia de Informação

Orientador: Murilo Bellezoni Loiola

Santo André - SP

Maio de 2022

Resumo

Segundo a ABNT, o resumo deve ressaltar o objetivo, o método, os resultados e as conclusões do documento. A ordem e a extensão destes itens dependem do tipo de resumo (informativo ou indicativo) e do tratamento que cada item recebe no documento original. O resumo deve ser precedido da referência do documento, com exceção do resumo inserido no próprio documento. Umas 10 linhas (...) As palavras-chave devem figurar logo abaixo do resumo, antecedidas da expressão Palavras-chave:, separadas entre si por ponto e finalizadas também por ponto.

Palavras-chaves: latex. abntex. editoração de texto.

Abstract

This is the english abstract.

Keywords: latex. abntex. text editoration.

Lista de ilustrações

Lista de tabelas

Lista de abreviaturas e siglas

ABNT	Associação Brasileira de Normas Técnicas
abnTeX	Normas para TeX

Lista de símbolos

Γ	Letra grega Gama
Λ	Lambda
ζ	Letra grega minúscula zeta
\in	Pertence

Sumário

1	INTRODUÇÃO	1
1.1	Scene Text Recognition	2
1.2	Objetivo	2
2	FUNDAMENTAÇÃO TEÓRICA	5
2.1	Considerações Finais	5
3	METODOLOGIA	7
3.1	Considerações Finais	7
4	RESULTADOS E DISCUSSÃO	9
4.1	Base de Dados	9
4.2	Considerações Finais	9
5	CONCLUSÃO	11
	REFERÊNCIAS	13

1 Introdução

Termos como inteligência artificial e aprendizado de máquina já são praticamente constantes do cotidiano humano, mesmo que muitas vezes não visível, já estão presentes em muitas aplicações que muitas vezes sequer imaginamos, desde a assistente virtual dos dispositivos móveis e eletrônicos, até sistemas capazes de embasar diagnósticos médicos, como por exemplo uma aplicação capaz de auxiliar em diagnósticos de COVID-19 durante a pandemia do vírus SARS-COV2 ^[1].

Um escopo de aplicação de técnicas de aprendizado de máquina que tem evoluído bastante com o crescimento da popularidade e acessibilidade dos conceitos que envolvem *machine learning* são aplicações voltadas para reconhecimento de texto.

A escrita com certeza foi uma das grandes habilidades que a humanidade desenvolveu que mudou como relações e sociedades funcionavam, sendo adotada para transmissão e armazenagem de dados e informação, meio de comunicação e expressão.

Com os avanços da tecnologia e em especial, dos computadores, um grande desafio emergiu: Como fazer com que computadores entendam o que está escrito em documentos físicos? Uma das primeiras patentes para soluções de OCR (abreviação de *Optical Character Recognition*) data de 1929 ^[2], mas isso não nega o fato que a capacidade de transportar texto do meio físico para o meio digital de forma eficiente é um desafio interessante e que motiva pesquisas até hoje.

Em linhas gerais, o reconhecimento óptico de caracteres é uma ampla tarefa de reconhecimento de padrões e, para que máquinas consigam identificar os padrões presentes na instâncias de texto, elas precisam conhecer ao menos algumas características dos caracteres e do texto que serão reconhecidos. Para exemplificar, a Reading Machine de Tauschek ^[2] era um aparelho mecânico que utilizava um disco de comparação, que continha o gabarito de cada um dos caracteres do alfabeto suportado. Esse foi o meio de "ensinar" a máquina a reconhecer um dado caracter.

Muitos anos depois, soluções ainda tem a missão de "treinar" computadores a identificar as características de caracteres e de textos como um todo e a principal ferramenta utilizada nos dias atuais são métodos sob o domínio de aprendizado de máquina, justamente pela capacidade de predição desses algoritmos dado um processo de treinamento. Ao longo deste trabalho de graduação outros conceitos que circundam o tópico de aprendizado de máquina serão introduzidos com um pouco mais de profundidade.

1.1 Scene Text Recognition

Um sub-conjunto de casos do espaço de aplicações de reconhecimento optico de caracteres ganhou bastante tração no ultima decada, impulsionada pelo alto poder computacional dos dispositivos modernos, acelerados por unidades gráficas, e a acessibilidade ao desenvolvimento de soluções de aprendizado de máquina. Esse sub-conjunto é conhecido como STR (abreviação de *Scene Text Recognition*, em inglês). Uma analogia para o STR seria aplicar soluções de OCR diretamente de fotos capturadas por uma câmera de um dispositivo móvel.

Como o nome sugere, STR classifica no sub-conjunto onde o problema a ser resolvido é a detecção e o reconhecimento do texto em imagens cotidianas, em cenas. A diferença entre um problema de STR comparado ao caso mais comum de OCR é, em termos simples, a aparência do texto e como ele será observado. Em uma imagem de cena, como por exemplo uma imagem da faixa de um supermercado, podemos ter textos em diferentes tamanhos, com diversas fontes, cores e orientações. Adicionalmente, por estarem muitas vezes sob influência do ambiente onde estão inseridos, outros fatores influenciam a observação desse texto, como iluminação, oclusão, danos devido ao clima, etc.

As soluções para problemas de STR, para lidarem com o nível de generalização necessário para reconhecer texto nos mais diversos casos, são largamente baseadas nos conceitos de deep learning, que demonstram ser capazes de ir um passo à frente no quesito reconhecimento de padrões em comparação às técnicas clássicas de processamento de imagem e conseguirem ser aplicadas com eficiência sobre imagens, sendo a aplicação das redes neurais convolucionais, comumente abreviadas para CNN (*Convolutional Neural Networks*, em inglês), um divisor de águas na evolução dessas soluções. [3, 4].

Assim, dada a importância do deep-learning, e em especial das CNNs nesse contexto de detecção e reconhecimento de texto, a próxima seção irá apresentar alguns conceitos básicos associados a essas estruturas.

1.2 Objetivo

Nulla malesuada risus ut urna. Aenean pretium velit sit amet metus. Duis iaculis. In hac habitasse platea dictumst. Nullam molestie turpis eget nisl. Duis a massa id pede dapibus ultricies. Sed eu leo. In at mauris sit amet tortor bibendum varius. Phasellus justo risus, posuere in, sagittis ac, varius vel, tortor. Quisque id enim. Phasellus consequat, libero pretium nonummy fringilla, tortor lacus vestibulum nunc, ut rhoncus ligula neque id justo. Nullam accumsan euismod nunc. Proin vitae ipsum ac metus dictum tempus. Nam ut wisi. Quisque tortor felis, interdum ac, sodales a, semper a, sem. Curabitur in velit sit amet dui tristique sodales. Vivamus mauris pede, lacinia eget, pellentesque quis,

scelerisque eu, est. Aliquam risus. Quisque bibendum pede eu dolor.

2 Fundamentação Teórica

Maecenas accumsan dapibus sapien. Duis pretium iaculis arcu. Curabitur ut lacus. Aliquam vulputate. Suspendisse ut purus sed sem tempor rhoncus. Ut quam dui, fringilla at, dictum eget, ultricies quis, quam. Etiam sem est, pharetra non, vulputate in, pretium at, ipsum. Nunc semper sagittis orci. Sed scelerisque suscipit diam. Ut volutpat, dolor at ullamcorper tristique, eros purus mollis quam, sit amet ornare ante nunc et enim.

Phasellus fringilla, metus id feugiat consectetur, lacus wisi ultrices tellus, quis lobortis nibh lorem quis tortor. Donec egestas ornare nulla. Mauris mi tellus, porta faucibus, dictum vel, nonummy in, est. Aliquam erat volutpat. In tellus magna, porttitor lacinia, molestie vitae, pellentesque eu, justo. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Sed orci nibh, scelerisque sit amet, suscipit sed, placerat vel, diam. Vestibulum nonummy vulputate orci. Donec et velit ac arcu interdum semper. Morbi pede orci, cursus ac, elementum non, vehicula ut, lacus. Cras volutpat. Nam vel wisi quis libero venenatis placerat. Aenean sed odio. Quisque posuere purus ac orci. Vivamus odio. Vivamus varius, nulla sit amet semper viverra, odio mauris consequat lacus, at vestibulum neque arcu eu tortor. Donec iaculis tincidunt tellus. Aliquam erat volutpat. Curabitur magna lorem, dignissim volutpat, viverra et, adipiscing nec, dolor. Praesent lacus mauris, dapibus vitae, sollicitudin sit amet, nonummy eget, ligula.

Cras egestas ipsum a nisl. Vivamus varius dolor ut dolor. Fusce vel enim. Pellentesque accumsan ligula et eros. Cras id lacus non tortor facilisis facilisis. Etiam nisl elit, cursus sed, fringilla in, congue nec, urna. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Integer at turpis. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Duis fringilla, ligula sed porta fringilla, ligula wisi commodo felis, ut adipiscing felis dui in enim. Suspendisse malesuada ultrices ante. Pellentesque scelerisque augue sit amet urna. Nulla volutpat aliquet tortor. Cras aliquam, tellus at aliquet pellentesque, justo sapien commodo leo, id rhoncus sapien quam at erat. Nulla commodo, wisi eget sollicitudin pretium, orci orci aliquam orci, ut cursus turpis justo et lacus. Nulla vel tortor. Quisque erat elit, viverra sit amet, sagittis eget, porta sit amet, lacus.

2.1 Considerações Finais

Donec et nisl id sapien blandit mattis. Aenean dictum odio sit amet risus. Morbi purus. Nulla a est sit amet purus venenatis iaculis. Vivamus viverra purus vel magna. Donec in justo sed odio malesuada dapibus. Nunc ultrices aliquam nunc. Vivamus facilisis pellentesque velit. Nulla nunc velit, vulputate dapibus, vulputate id, mattis ac, justo. Nam

mattis elit dapibus purus. Quisque enim risus, congue non, elementum ut, mattis quis, sem. Quisque elit.

3 Metodologia

Maecenas accumsan dapibus sapien. Duis pretium iaculis arcu. Curabitur ut lacus. Aliquam vulputate. Suspendisse ut purus sed sem tempor rhoncus. Ut quam dui, fringilla at, dictum eget, ultricies quis, quam. Etiam sem est, pharetra non, vulputate in, pretium at, ipsum. Nunc semper sagittis orci. Sed scelerisque suscipit diam. Ut volutpat, dolor at ullamcorper tristique, eros purus mollis quam, sit amet ornare ante nunc et enim.

Phasellus fringilla, metus id feugiat consectetur, lacus wisi ultrices tellus, quis lobortis nibh lorem quis tortor. Donec egestas ornare nulla. Mauris mi tellus, porta faucibus, dictum vel, nonummy in, est. Aliquam erat volutpat. In tellus magna, porttitor lacinia, molestie vitae, pellentesque eu, justo. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Sed orci nibh, scelerisque sit amet, suscipit sed, placerat vel, diam. Vestibulum nonummy vulputate orci. Donec et velit ac arcu interdum semper. Morbi pede orci, cursus ac, elementum non, vehicula ut, lacus. Cras volutpat. Nam vel wisi quis libero venenatis placerat. Aenean sed odio. Quisque posuere purus ac orci. Vivamus odio. Vivamus varius, nulla sit amet semper viverra, odio mauris consequat lacus, at vestibulum neque arcu eu tortor. Donec iaculis tincidunt tellus. Aliquam erat volutpat. Curabitur magna lorem, dignissim volutpat, viverra et, adipiscing nec, dolor. Praesent lacus mauris, dapibus vitae, sollicitudin sit amet, nonummy eget, ligula.

Cras egestas ipsum a nisl. Vivamus varius dolor ut dolor. Fusce vel enim. Pellentesque accumsan ligula et eros. Cras id lacus non tortor facilisis facilisis. Etiam nisl elit, cursus sed, fringilla in, congue nec, urna. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Integer at turpis. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Duis fringilla, ligula sed porta fringilla, ligula wisi commodo felis, ut adipiscing felis dui in enim. Suspendisse malesuada ultrices ante. Pellentesque scelerisque augue sit amet urna. Nulla volutpat aliquet tortor. Cras aliquam, tellus at aliquet pellentesque, justo sapien commodo leo, id rhoncus sapien quam at erat. Nulla commodo, wisi eget sollicitudin pretium, orci orci aliquam orci, ut cursus turpis justo et lacus. Nulla vel tortor. Quisque erat elit, viverra sit amet, sagittis eget, porta sit amet, lacus.

3.1 Considerações Finais

Donec et nisl id sapien blandit mattis. Aenean dictum odio sit amet risus. Morbi purus. Nulla a est sit amet purus venenatis iaculis. Vivamus viverra purus vel magna. Donec in justo sed odio malesuada dapibus. Nunc ultrices aliquam nunc. Vivamus facilisis pellentesque velit. Nulla nunc velit, vulputate dapibus, vulputate id, mattis ac, justo. Nam

mattis elit dapibus purus. Quisque enim risus, congue non, elementum ut, mattis quis, sem. Quisque elit.

4 Resultados e Discussão

Integer vel enim sed turpis adipiscing bibendum. Vestibulum pede dolor, laoreet nec, posuere in, nonummy in, sem. Donec imperdiet sapien placerat erat. Donec viverra. Aliquam eros. Nunc consequat massa id leo. Sed ullamcorper, lorem in sodales dapibus, risus metus sagittis lorem, non porttitor purus odio nec odio. Sed tincidunt posuere elit. Quisque eu enim. Donec libero risus, feugiat ac, dapibus eget, posuere a, felis. Quisque vel lectus ut metus tincidunt eleifend. Duis ut pede. Duis velit erat, venenatis vitae, vulputate a, pharetra sit amet, est. Etiam fringilla faucibus augue.

4.1 Base de Dados

Praesent facilisis, augue a adipiscing venenatis, libero risus molestie odio, pulvinar consectetur felis erat ac mauris. Nam vestibulum rhoncus quam. Sed velit urna, pharetra eu, eleifend eu, viverra at, wisi. Maecenas ultrices nibh at turpis. Aenean quam. Nulla ipsum. Aliquam posuere luctus erat. Curabitur magna felis, lacinia et, tristique id, ultrices ut, mauris. Suspendisse feugiat. Cras eleifend wisi vitae tortor. Phasellus leo purus, mattis sit amet, auctor in, rutrum in, magna. In hac habitasse platea dictumst. Phasellus imperdiet metus in sem. Vestibulum ac enim non sem ultricies sagittis. Sed vel diam.

4.2 Considerações Finais

Aenean velit sem, viverra eu, tempus id, rutrum id, mi. Nullam nec nibh. Proin ullamcorper, dolor in cursus tristique, eros augue tempor nibh, at gravida diam wisi at purus. Donec mattis ullamcorper tellus. Phasellus vel nulla. Praesent interdum, eros in sodales sollicitudin, nunc nulla pulvinar justo, a euismod eros sem nec nibh. Nullam sagittis dapibus lectus. Nullam eget ipsum eu tortor lobortis sodales. Etiam purus leo, pretium nec, feugiat non, ullamcorper vel, nibh. Sed vel elit et quam accumsan facilisis. Nunc leo. Suspendisse faucibus lacus.

5 Conclusão

Proin non sem. Donec nec erat. Proin libero. Aliquam viverra arcu. Donec vitae purus. Donec felis mi, semper id, scelerisque porta, sollicitudin sed, turpis. Nulla in urna. Integer varius wisi non elit. Etiam nec sem. Mauris consequat, risus nec congue condimentum, ligula ligula suscipit urna, vitae porta odio erat quis sapien. Proin luctus leo id erat. Etiam massa metus, accumsan pellentesque, sagittis sit amet, venenatis nec, mauris. Praesent urna eros, ornare nec, vulputate eget, cursus sed, justo. Phasellus nec lorem. Nullam ligula ligula, mollis sit amet, faucibus vel, eleifend ac, dui. Aliquam erat volutpat.

Referências

- 1 ZHAO, W.; JIANG, W.; QIU, X. Deep learning for covid-19 detection based on ct images. *Scientific Reports*, v. 11, 2021. Citado na página [1](#).
- 2 Gustav Tauschek. *Reading machine*. US2026330A, 27 maio 1929. Citado na página [1](#).
- 3 RAISI, Z. et al. Text detection and recognition in the wild: A review. *ArXiv*, abs/2006.04305, 2020. Citado na página [2](#).
- 4 LONG, S.; HE, X.; YAO, C. Scene text detection and recognition: The deep learning era. *International Journal of Computer Vision*, v. 129, p. 161–184, 2020. Citado na página [2](#).