



**16º Congresso Interinstitucional de Iniciação Científica - CIIC
2022**

**30 e 31 de agosto de 2022
ISBN: 978-65-88414-07-1**

TESTANDO A REDE NEURAL YOLOV5 PARA DETECÇÃO DE FRUTOS EM POMARES DE LARANJA

Alan Freitas **Ribeiro**¹; Kleber Xavier Sampaio de **Souza**²; João **Camargo Neto**³; Sônia **Ternes**⁴;
Inácio Henrique **Yano**⁵

Nº 22602

RESUMO – Uma das maiores fontes de renda do Brasil é a agropecuária, dentro deste setor, um dos mais importantes produtos é a laranja. O país é um dos maiores produtores deste fruto, com destaque para a região de São Paulo e do Triângulo mineiro, responsáveis por grande parte do cultivo. Este trabalho, tem como objetivo analisar a viabilidade do emprego de algoritmos computacionais por meio de redes neurais e visão computacional no processo de contagem de frutos através de imagens digitais, em especial, a utilização da rede neural YOLOv5. Obtivemos bons resultados na detecção dos objetos e na velocidade de execução da rede, revelando ser uma rede promissora para aplicações mobile.

Palavras-chaves: Redes neurais, visão computacional, aprendizado de máquina, aprendizado profundo, rede yolov5, detecção de objetos.

1 Autor, Bolsista CNPq (PIBIC): Graduação em Engenharia de Computação, Unicamp, Campinas-SP; alanxribeiro98@gmail.com.

2 Orientador: Pesquisador da Embrapa Informática Agropecuária, Campinas-SP; kleber.sampaio@embrapa.br.

3 Analista da Embrapa Informática Agropecuária, Campinas - SP.

4 Pesquisadora da Embrapa Informática Agropecuária, Campinas - SP.

5 Analista da Embrapa Informática Agropecuária, Campinas - SP.



**16º Congresso Interinstitucional de Iniciação Científica - CIIC
2022**

**30 e 31 de agosto de 2022
ISBN: 978-65-88414-07-1**

ABSTRACT – *One of the biggest sources of income in Brazil is agriculture, within this sector, one of the most important products is the orange. The country is one of the largest producers of this fruit, with emphasis on the region of São Paulo and the Triângulo Mineiro, responsible for most of the cultivation. This work aims to analyze the feasibility of using computational algorithms through neural networks and computer vision in the process of counting fruits through digital images, in particular, the use of the YOLOv5 neural network. We obtained good results in the detection of objects and in the execution speed of the network, showing it to be a promising neural network for mobile applications.*

Keywords: Neural networks, computer vision, machine learning, deep learning, yolov5 network, object detection.

1 INTRODUÇÃO

No Brasil, a agricultura é uma das áreas de maior importância para o desenvolvimento econômico do país, gerando altas receitas e empregos anualmente, sendo responsável por grande parte do produto interno bruto (PIB) nacional. Dentre vários gêneros de agricultura praticados no país, a citricultura é uma das mais importantes, no qual, o Brasil, é responsável por mais da metade da produção mundial de suco de laranja. Para a safra de 2022/23 é estimada uma produção de 316,95 milhões de caixas do fruto (FUNDECITRUS, 2022). Por ser uma área de suma importância, existe a constante necessidade de melhorar as técnicas de controle de produção já existentes, com o objetivo de melhorar a alocação de recursos e facilitar uma maior produção do fruto.

Para obter uma estimativa de produção atual e futura, é necessário fazer a contagem de frutos em laranjeiras selecionadas. Atualmente, é utilizado o método de derriça, em que as laranjas são retiradas das árvores e contadas manualmente, algo que é extremamente custoso em tempo e mão de obra, além de ser um processo destrutivo, pois remove os frutos das árvores escolhidas.

Para tentar solucionar este problema, já existem trabalhos anteriores que tentam empregar métodos computacionais para realizar a contagem automática destes frutos, tornando assim, o



processo mais eficiente e menos custoso. Dentre estes métodos, há um destaque na área de visão computacional, onde podemos utilizar redes neurais para auxiliar na contagem de laranjas. Alguns outros autores pesquisaram o uso de redes neurais para este fim, sendo elas arquiteturas de um estágio, como por exemplo a YOLOv3 (CAMARGO NETO et al., 2019) e a SSD (SOUSA et al., 2021), tanto quanto arquitetura de dois estágios, como a Faster-RCNN (CERQUEIRA et al., 2020). Neste trabalho, decidimos testar a aplicação da rede neural YOLOv5 (ULTRALYTICS, 2022), por ser uma rede que possui alguns modelos de arquitetura leves, que podem ser executados em dispositivos com menor poder computacional de forma efetiva sem perder muita precisão.

2 MATERIAL E MÉTODOS

2.1 A rede neural YOLOV5

YOLO, acrônimo de *You only look once*, é um algoritmo de detecção de objeto que modela o problema de detecção como um problema de regressão, para atingir tal feito, o algoritmo divide a imagem em um sistema de grade, onde em cada célula, são preditas N caixas delimitadoras com determinado nível de confiança contendo as probabilidades de cada classe. O modelo YOLO é um detector de um estágio, em que a predição de caixas delimitadoras está conectada a predição dos rótulos de classe, aprimorando assim, a velocidade de inferência da mesma.

A arquitetura da YOLOv5, possui três pontos principais. O primeiro é a construção das camadas *Backbone* (espinha dorsal), formada por uma rede neural convolucional, é responsável por extrair características das imagens de entrada. Na YOLOv5, as camadas *Backbone* implementam o *Cross Stage Partial Networks (CSP)* (WANG et al., 2019), utilizando uma versão modificada para a arquitetura da rede. O intuito do CSP é mitigar os efeitos do problema de duplicação do gradiente proveniente de outras redes convolucionais resultando em um menor número de parâmetros, diminuindo assim, as computações feitas pela rede e aumentando a acurácia da mesma. O segundo ponto são as camadas *Neck* (pescoço), cujo objetivo é combinar as características extraídas das camadas *Backbone* para passá-las para a fase de predição. A arquitetura da YOLOv5 utiliza a *Path Aggregation Network (PANet)*, que melhora o processo de segmentação da imagem, conseguindo preservar as informações espaciais (SHU et al., 2018). Por último, temos um conjunto de camadas *Head* (cabeça), que são camadas que recebem as características das camadas *Neck* para realizar as predições das caixas delimitadoras e das classes pertencentes de cada objeto.

2.2 Fonte e tratamento de dados



Para executar os treinamentos da YOLOv5, utilizamos dois conjuntos de dados, o primeiro, em que chamamos de Embrapa, consiste em 3064 imagens de laranjeiras, contendo as marcações das localizações de cada fruto em um arquivo de texto com mesmo nome da imagem. Todas as imagens possuem tamanho 416x416 pixels, e foram divididas em três partes: a) 2053 imagens para treino; b) 200 imagens para validação; e c) 811 imagens para teste. As imagens foram fornecidas pelo Fundecitrus, sendo retiradas de diversos dispositivos, como câmeras fotográficas e celulares, possuindo assim, uma grande diversidade de posições, condições de iluminação e variação de ambiente.

O segundo conjunto de dados, em que chamamos de EmbrapaNL, contém as 3064 imagens do conjunto anteriormente citado com um adicional de 430 novas imagens retiradas de um conjunto de dados feito por Kritik Seth, disponibilizado na plataforma *Kaggle* (SETH, 2022).

As imagens escolhidas foram de outros frutos e vegetais, como limão, tomate, uva, etc. O conjunto de dados original, continha imagens de desenhos que remetiam ao fruto original, porém estas foram descartadas, afim de escolher apenas imagens reais dos frutos e vegetais. As imagens foram adicionadas sem marcações, representando que não havia laranja na imagem, com intuito de melhorar o processo de extração de características do algoritmo da rede.

2.3 Treinamento

Para o treinamento, utilizamos a aplicação disponibilizada no *github* por *Ultralytics* (2022), feita em python e que utiliza o PyTorch como base. Testamos todos os modelos disponíveis da arquitetura P5: a) YOLOv5n, com 1,9 milhões de parâmetros; b) YOLOv5s, com 7,2 milhões; c) YOLOv5m, com 21,2 milhões; d) YOLOv5l, com 46,5 milhões; e e) YOLOv5x, com 86,7 milhões de parâmetros. A fim de comparação, resolvemos executar também o treinamento dos modelos da versão da YOLOv3 disponibilizada pela *Ultralytics*, além de reproduzir o trabalho feito por SOUSA et al., (2021) com a rede *Single Shot Detector* (SSD), testando a rede final neste novo conjunto de dados.

Os modelos da YOLOv3 e v5, foram treinados por 900 épocas, utilizando os valores padrão fornecidos na aplicação da *Ultralytics*, com exceção para o tamanho da imagem, onde executamos o treinamento com o tamanho de 416 ao invés de 640. Para a rede SSD, geramos um modelo treinado idêntico ao da SOUSA et. al., (2021), com a exceção de que salvamos apenas três modelos durante a execução: O que tinha o melhor *mean average precision* (mAP), o melhor F1 e o modelo gerado na última época.



3 RESULTADOS E DISCUSSÃO

3.1 Resultados da YOLOv5

Podemos classificar os objetos detectados nas imagens em três categorias: a) Verdadeiro positivo (VP), caso o objeto tenha sido detectado corretamente; b) Falso negativo (FN), caso o objeto exista mas o algoritmo não conseguiu encontrá-lo; e c) Falso positivo (FP), caso o objeto não exista mas o algoritmo o detectou de forma errônea. A partir daí, podemos medir o desempenho dos modelos gerados através das métricas como o mAP, a precisão ($VP/VP+FP$), revocação ($VP/VP+FN$) e o F1 ($2*precisão*revocação/(precisão+revocação)$). Para o mAP, utilizamos duas medições diferentes, o mAP_0.5, que representa o mAP calculado com o *Intersect over union* (IoU) em 0,5 e o mAP_0.5:0.95, aferido com o IoU variando de 0,5 até 0,95 em um passo de 0,05.

Em um primeiro momento, treinamos todos os modelos da YOLOv5 utilizando os pesos pré treinados da rede já disponibilizado pela *Ultralytics*, utilizando o conjunto de dados Embrapa, onde obtivemos os seguintes resultados na validação:

Tabela 1: Valores obtidos durante a validação do treinamento de cada modelo de rede.

Nome do modelo	Melhor mAP_0.5	Melhor mAP_0.5:0.95	Melhor precisão	Melhor revocação
YOLOv5n	0.9335	0.6575	0.9223	0.8514
YOLOv5s	0.9242	0.6603	0.9441	0.8369
YOLOv5m	0.9415	0.6552	0.9434	0.8473
YOLOv5l	0.9318	0.6500	0.9355	0.8514
YOLOv5x	0.9373	0.6493	0.9240	0.8566

E os seguintes resultados no conjunto de teste:

Tabela 2: Valores obtidos após a conclusão do treinamento de cada modelo no conjunto de teste.

Nome do modelo	Melhor mAP_0.5	Melhor mAP_0.5:0.95	Melhor precisão	Melhor revocação
YOLOv5n	0.951	0.667	0.942	0.878
YOLOv5s	0.950	0.675	0.945	0.875
YOLOv5m	0.955	0.671	0.932	0.894
YOLOv5l	0.956	0.669	0.934	0.893



YOLOv5x	0.958	0.671	0.937	0.906
---------	-------	-------	-------	-------

Pelos resultados obtidos, percebemos que, para este problema, a aplicação do modelo nano é bem vantajoso, pois ele oferece um resultado próximo dos modelos maiores, possuindo um custo de computação bem menor, com velocidade de inferência (ver **Tabela 4**) bem mais rápida do que os demais modelos.

3.2 Resultados do uso de imagens sem laranja

Após a testagem dos modelos mostrada no item anterior, testamos o comportamento do algoritmo quanto a adição de imagens com novos frutos e vegetais no conjunto de treino, afim de diminuir o enviesamento e forçar o algoritmo a extrair características mais específicas da laranja para uma melhor predição.

Executamos os testes utilizando o modelo YOLOv5m original como base, modificando apenas os conjuntos de dados utilizados e o uso de pesos pré treinados no conjunto de dados do *Common Objects in Context* (COCO) disponibilizado pela *Ultralytics*. Os modelos YOLOv5m e YOLOv5mNW foram treinados com o conjunto de dados Embrapa, enquanto os modelos YOLOv5mNL e YOLOv5mNLNW foram treinados com o conjunto de dados EmbrapaNL. Quanto a utilização dos pesos pré treinados, os modelos YOLOv5mNW e YOLOv5mNLNW foram inicializados com pesos aleatórios com o algoritmo de inicialização *default* da YOLOv5, já os modelos YOLOv5m e YOLOv5NL foram treinados com os pesos pré treinados disponibilizados pela *Ultralytics*. Obtivemos o seguinte resultado do conjunto de validação:

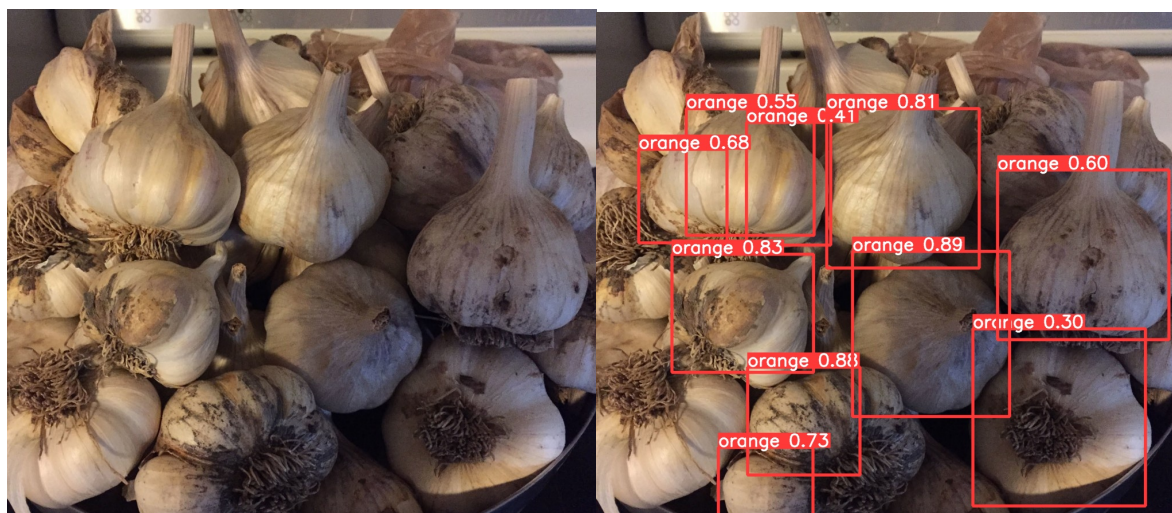
Tabela 3: Valores obtidos durante a validação do treinamento dos modelos referidos no item 3.2

Nome do modelo	Melhor mAP_0.5	Melhor mAP_0.5:0.95	Melhor precisão	Melhor revocação
YOLOv5m	0.9415	0.6552	0.9434	0.8473
YOLOv5mNL	0.9369	0.6560	0.9458	0.8435
YOLOv5mNW	0.9344	0.6530	0.9457	0.8409
YOLOv5mNLNW	0.9356	0.6610	0.9324	0.8631

Percebemos que ao introduzir imagens sem rótulos, conseguimos melhorar a detecção de falsos negativos nos modelos treinados, como podemos perceber nos valores da tabela 3, onde o modelo YOLOv5mNLNW apesar de ter uma pequena queda na precisão, algo que já esperávamos,



por termos inserido frutos muito parecidos com a laranja, como por exemplo o limão, obtivemos um ganho de aproximadamente 0.02 na revocação em comparação com os outros modelos.



Figuras 1 e 2. A esquerda, resultado da detecção do modelo YOLOv5mNLNW, a direita, resultado da detecção do modelo YOLOv5m.

Este resultado, aponta que pode haver uma vantagem para conjuntos de dados deste tipo, possuindo apenas uma classe (laranja), utilizar imagens com outras classes durante o treinamento para melhorar a extração de características próprias do fruto. Podemos perceber nas figuras 1 e 2, que o modelo exposto a outros tipos de frutos e vegetais conseguiu caracterizar melhor a laranja, de forma que o modelo que não foi exposto a estas imagens, identificou o alho como laranja apenas pelo formato arredondado. Nos testes, percebemos que o modelo treinado com outros tipos de frutos, detectou menos falsos positivos de folhas arredondadas e conseguiu detectar mais laranjas entre folhas.

3.3 Comparação com outras redes neurais

Para fazer a comparação, utilizamos o conjunto de testes do conjunto de dados Embrapa para obter os resultados mencionados na Tabela 4, os valores mAP_0.5 e F1, foram obtidos da testagem do conjunto de dados com 811 imagens, enquanto os *Frames per second* (FPS) foram obtidos diretamente do tempo de inferência, utilizando a detecção em uma imagem escolhida de dimensões 416x416 pixels. Os modelos utilizados, são os modelos anteriormente mencionados da YOLOv5, os modelos da YOLOv3 disponibilizados pela *Ultralytics* (2022), e o modelo da SSD, utilizado por SOUZA et al. (2021). Cabe destacar também, que a execução de cada rede foi feita em sua própria aplicação, e para realizar as inferências foi utilizada a gpu *Quadro RTX 4000*.



**16º Congresso Interinstitucional de Iniciação Científica - CIIC
2022
30 e 31 de agosto de 2022
ISBN: 978-65-88414-07-1**

Tabela 4: Valores obtidos durante os testes descritos na seção 3.3

Nome do modelo	Resolução	mAP_0.5	F1	Nº Parâmetros (em milhões)	FPS	Tamanho (em MB)
YOLOv5n	640x640	0.957	0.911	1,9	250	3,7
YOLOv5n	416x416	0.951	0.909	1,9	294	3,7
YOLOv5n	256x256	0.932	0.922	1,9	294	3,7
YOLOv5s	640x640	0.958	0.920	7,2	112	14
YOLOv5m	640x640	0.953	0.903	21,2	45	41
YOLOv5l	640x640	0.956	0.907	46,5	25	89
YOLOv5x	640x640	0.958	0.921	86,7	17	166
YOLOv3	640x640	0.951	0.914	61,5	21	118
YOLOv3SPP	640x640	0.947	0.915	62,5	21	120
YOLOv3Tiny	640x640	0.941	0.898	8,7	200	17
SSD	300x300	0.814	0.896	23,7	54	182

No geral, os modelos da YOLOv5 e YOLOv3 obtiveram resultados bem parecidos no conjunto de validação e teste, fato que pode ser explicado pela boa disposição de imagens do conjunto de dados, que refletiu bem vários tipos de cenas fotografadas, resultando em um bom treinamento dos modelos para detecção em situações adversas. Além disso, notamos durante os treinamentos, que os modelos rapidamente convergiam para bons resultados (em torno de 10 a 20 épocas já apresentavam mAP maior que 90).

Esta facilidade de extração de características para diferenciar folhas e frutos, pode ter sido o principal fator dos modelos mais robustos apresentarem desempenho similar aos modelos com menor quantidade de parâmetros, fato que dificilmente aconteceria em conjunto de dados com um maior número de classes e imagens ilustrando cenários de maior complexidade. Por este motivo, procuramos melhorar a extração de características da classe laranja, como discutido no item 3.2, para que ficassem as mais próximas do fruto possível.

Ao analisarmos a Tabela 4, percebemos que a YOLOv5n (nano) representa uma melhora significativa tanto com relação à YOLOv3, quanto relativa ao modelo SSD testado por SOUSA et al. (2021). O modelo nano possui tamanho de armazenamento 49x menor, velocidade de inferência 5x maior que a SSD, utilizando uma resolução maior (640x640 contra 300x300), conseguindo métricas melhores de mAP e F1. Com relação à YOLOv3, embora essas métricas tenham sido similares, a



YOLOv5 apresenta vantagem por possuir um tamanho de armazenamento menor e FPS maiores, possibilitando assim, uma melhor implantação, principalmente em plataformas móveis.



Figura 3 e 4. A esquerda, detecção do modelo YOLOv5n, a direita, detecção do modelo SSD.

Percebemos que a maior diferença entre as detecções dos modelos YOLO e da SSD foram nas imagens possuindo cachos de laranja. Como podemos ver nas figuras 3 e 4, os modelos da YOLO conseguem detectar melhor os frutos no cacho, porém, os frutos que estão atrás, geralmente recebem uma confiança menor na detecção da classe, em contrapartida, o modelo da SSD comumente detectava frutos a mais nas mesmas condições.

4 CONCLUSÃO

Neste trabalho, foi desenvolvida uma rede neural para detectar laranjas a partir de imagens fotografadas em condições comuns e sem necessitar de boas especificações técnicas, facilitando assim o uso de dispositivos com baixo poder computacional, com intuito de requerir um baixo investimento para a aplicação desta tecnologia no meio agrário. Os resultados obtidos foram bons, principalmente ao considerarmos o modelo YOLOv5n (nano), se comparado às outras arquiteturas testadas, apresentou bons resultados, uma velocidade de inferência muito superior aos demais modelos e tamanho de armazenamento pequeno, resultados muito promissores, principalmente se formos levar em consideração um ambiente de execução *mobile*.



5 AGRADECIMENTOS

Devemos agradecimentos ao PIBIC/CNPq, pela concessão da bolsa de Iniciação Científica ao aluno Alan Freitas Ribeiro, e à equipe do PES/Fundecitrus pela disponibilização das fotos de laranjeiras.

6 REFERÊNCIAS

CAMARGO NETO, J.; TERNES, S.; SOUZA, K. X. S. de; YANO, I. H.; QUEIROS, L. R. **Uso de redes neurais convolucionais para detecção de laranjas verdes**. CONGRESSO BRASILEIRO DE AGROINFORMÁTICA, 12., 2019, Indaiatuba. **Anais...** Disponível em: <<https://ainfo.cnptia.embrapa.br/digital/bitstream/item/216886/1/PC-Redes-neurais-SBIAGRO-2019.pdf>>. Acesso em: 10 jun. 2022.

CERQUEIRA, L. M.; SOUZA, K. X. S. de; TERNES, S.; CAMARGO NETO, J. **Usando a rede neural Faster-RCNN para identificar frutos verdes em pomares de laranja**. In: CONGRESSO INTERINSTITUCIONAL DE INICIAÇÃO CIENTÍFICA, 14., 2020. **Anais...** Campinas: Embrapa Informática Agropecuária, 2020. p. 1-9. Evento online. CIIC 2020. No 20605. Disponível em: <<https://www.alice.cnptia.embrapa.br/handle/doc/1127731>>. Acesso em: 04 jun. 2022.

FUNDECITRUS. **Estimativa da safra de laranja 2022/23 do cinturão citrícola de São Paulo e Triângulo/Sudoeste Mineiro**: sumário executivo. Araraquara, [2022]. 17 p. Disponível em: <https://www.fundecitrus.com.br/pdf/pes_relatorios/2022_05_26_Sumario_Executivo_da_Estimativa_da_Safra_de_Laranja_2022-2023.pdf>. Acesso em: 10 jun. 2022.

LIU, S.; QI, L.; QIN, H.; SHI, J.; JIA, J. **Path Aggregation Network for Instance Segmentation**. 2018. Disponível em: <<https://arxiv.org/pdf/1803.01534.pdf>>. Acesso em: 10 jun. de 2022.

SETH, K. **Fruits and Vegetables Image Recognition Dataset**. Disponível em: <<https://www.kaggle.com/datasets/kritikseth/fruit-and-vegetable-image-recognition>>. Acesso em: 10 mar. 2022.

SOUZA, M. A. de; SOUZA, K. X. S. de; CAMARGO NETO, J.; TERNES, S.; YANO, I. H. **Usando a rede neural SSD para identificar frutos verdes em pomares de laranja**. In: CONGRESSO INTERINSTITUCIONAL DE INICIAÇÃO CIENTÍFICA, 15., 2021. **Anais...** Campinas: Instituto de Zootecnia, 2021. p. 1-10. Evento online. CIIC 2021. No 21610. Disponível em: <<https://www.alice.cnptia.embrapa.br/handle/doc/1135159>>. Acesso em: 04 out. 2021.

ULTRALYTICS. **YOLOv3**. 2022. Disponível em: <<https://github.com/ultralytics/yolov3>>. Acesso em: 10 mar. 2022.



**16º Congresso Interinstitucional de Iniciação Científica - CIIC
2022
30 e 31 de agosto de 2022
ISBN: 978-65-88414-07-1**

ULTRALYTICS. **YOLOv5**. 2022. Disponível em: <<https://github.com/ultralytics/yolov5>>. Acesso em: 10 mar. 2022.

WU, Y.; WANG, C.; LIAO, H. M.; YEH, I.; CHEN, P.; HSIEH, J. **CSPNet**: A New Backbone that can Enhance Learning Capability of CNN. 2019. Disponível em: <<https://arxiv.org/pdf/1911.11929.pdf>>. Acesso em: 08 jun. de 2022.