

DESAFIOS E POTENCIALIDADES DA VISÃO COMPUTACIONAL PARA APLICAÇÕES DO MUNDO REAL: UMA ANÁLISE DETALHADA DO SSD E SEU IMPACTO

Endereço institucional: Estrada Mário Covas Junior, km 1 – Vila Guararapes, Lins - SP, 16403-025

Resumo: *Este estudo tem como principal objetivo aprofundar na compreensão do Single Shot MultiBox Detector (SSD) e das teorias subjacentes. Posteriormente, aplicaremos o SSD a imagens obtidas pelos autores. Nossa motivação é derivada da necessidade de uma compreensão sólida de visão computacional para desenvolver um algoritmo demonstrativo de reconhecimento de padrões em diversas imagens. O resultado consiste na análise e compreensão aprofundada do Single Shot MultiBox Detector (SSD) e suas teorias subjacentes, bem como na aplicação prática do SSD a imagens obtidas pelos autores. Desta forma, o estudo proporcionou uma apreciação qualitativa e quantitativa dos resultados, permitindo uma melhor compreensão do desempenho do método. Esta pesquisa também mostra insights importantes sobre a aplicação de técnicas de visão computacional em reconhecimento de padrões em imagens variadas.*

Palavras-chaves: *Single Shot MultiBox Detector (SSD); Visão Computacional; Reconhecimento de Padrões; Análise de Imagens; Filtros Adaptativos.*

Abstract: *The main objective of this study is to deepen the understanding of the Single Shot MultiBox Detector (SSD) and the underlying theories. Subsequently, we will apply the SSD to images obtained by the authors. Our motivation is derived from the need for a solid understanding of computer vision to develop a demonstrative algorithm for pattern recognition in diverse images. The result consists of the in-depth analysis and understanding of the Single Shot MultiBox Detector (SSD) and its underlying theories, as well as the practical application of the SSD to images obtained by the authors. Thus, the study provided a qualitative and quantitative appreciation of the results, allowing a better understanding of the performance of the method. This research also shows important insights into the application of computer vision techniques in pattern recognition in varied images.*

Keywords: *Single Shot MultiBox Detector (SSD); Computer Vision; Pattern Recognition; Image Analysis; Adaptive filters.*

1 – INTRODUÇÃO

Nos últimos anos, houve uma crescente necessidade de sistemas de visão computacional para solucionar problemas do mundo real. No entanto, a maioria dos modelos e métodos atuais não consegue lidar com desafios além de tarefas básicas. Para atender às demandas de aplicações do mundo real, como navegação, reconhecimento de alvos e outros, é essencial desenvolver técnicas de visão robustas e flexíveis capazes de funcionar bem sob diversas condições desafiadoras (SEBE et al., 2005).

O objetivo deste trabalho é analisar e entender o conceito do *Single Shot MultiBox Detector* – SSD e das respectivas teorias, para posteriormente aplicá-la com imagens adquiridas pelos autores, para a análise e compreensão dos resultados qualitativos e quantitativos. A motivação deste trabalho advém da necessidade de entender o conceito da visão computacional para a construção de um algoritmo que demonstre o conceito e aplicação em reconhecimento de padrão para diferentes imagens.

Este trabalho de pesquisa é dividido em duas etapas. Na primeira etapa, usaremos o estudo de Liu et al. (2016) para explorar o SSD, um método que detecta objetos em imagens usando uma única rede neural profunda. O SSD utiliza uma abordagem que discretiza o espaço de saída das caixas delimitadoras em caixas padrão com diferentes proporções de aspecto e escalas. Durante a inferência, a rede gera pontuações para a presença de objetos em cada caixa padrão e ajusta as caixas para melhor corresponder à forma do objeto. O SSD combina previsões de múltiplos mapas de características com diferentes resoluções para lidar com objetos de tamanhos variados. Uma característica crucial do SSD é a eliminação da geração de propostas de objetos, simplificando o treinamento e a integração em sistemas que requerem detecção de objetos.

Na segunda fase algumas figuras serão aplicadas a uma implementação em Python do algoritmo de Liu et al. (2016), desta forma, serão considerados como o resultado desta aplicação, o entendimento e análise qualitativa e quantitativa do método proposto. Basicamente, o resultado desta análise baseia-se na junção do entendimento dos conceitos teóricos e da simulação computacional para a obtenção de um único entendimento na construção futura de um algoritmo similar.

Pesquisadores como Li et al. (2020) já utilizaram o conhecimento de Liu et al. (2016) para propor métodos, como o TridentSSD, focado em detecção de objetos em diferentes escalas. Este trabalho demonstra a viabilidade de treinar ramos separados para objetos de diferentes escalas, com resultados inferiores ao TridentSSD em experimentos nos conjuntos de dados PASCAL VOC2007 e VOC2012.

2 – CONTEXTUALIZAÇÃO TEÓRICA (CONCEITO DE VISÃO COMPUTACIONAL E SEUS PERIFÉRICOS)

2.1 – Visão Computacional

A visão computacional é um campo de estudo que se concentra em auxiliar os computadores a enxergarem, usando dados de imagens para inferir informações sobre o mundo (PRINCE, 2012). É uma área multidisciplinar, um subcampo da inteligência artificial e do aprendizado de máquina, envolvendo métodos especializados e algoritmos gerais (BROWNLEE, 2019). Os principais problemas que a visão computacional tem de

lidar são o reconhecimento óptico de caracteres, classificação de imagens, detecção de objetos, detecção de rosto e reconhecimento facial. Essas tarefas estão reunidas sob o termo "reconhecimento de objetos," que engloba a identificação, localização e extração de conteúdo específico em fotografias digitais (BROWNLEE, 2019).

2.2 - Redes Neurais Artificiais (*Artificial Neural Networks* - ANN)

Redes neurais, inspiradas na biologia, possibilitam que computadores aprendam com dados observacionais (NIELSEN, 2019; FURTADO, 2019). Essas redes artificiais consistem em camadas de neurônios interconectados (FURTADO, 2019). Na sequência de processamento, a camada de entrada atua como uma ponte, permitindo a passagem de dados brutos ou características específicas para as camadas subsequentes (HAYKIN, 2001). Nas camadas ocultas, ocorre o processamento sequencial dos dados (MIN et al., 2017). A camada de saída gera valores de probabilidade para identificar a classe da entrada, baseando-se no conhecimento adquirido nas camadas anteriores (SHANTINI et al., 2017).

Na visão computacional, identificamos três principais tipos de redes neurais: as *Convolutional Neural Networks* (CNN), focadas no processamento de dados de imagens; as *Recurrent Neural Networks* (RNNs), como as LSTMs, projetadas para sequências de dados e informações temporais, adequadas para vídeos; e as redes *feedforward*, voltadas para tarefas de classificação e regressão. As redes *feedforward* podem ser categorizadas em *Perceptron* e *Multi-Layer Perceptron* (MLP), dependendo do número de camadas ocultas (YAMASHITA et al., 2018; AGARWAL, 2018).

2.3 - Redes Neurais Convolucionais (CNN)

As *Convolutional Neural Networks* (CNNs), diz Yamashita et al. (2018), são um tipo de modelo de aprendizado profundo especialmente projetado para processar dados com uma estrutura de grade, como imagens. Esses modelos foram inspirados na organização do córtex visual de animais e têm a capacidade de aprender hierarquias de características espaciais, que vão desde características de baixo nível até características de alto nível.

Na detecção de objetos, os principais tipos de redes neurais podem ser divididos de acordo com o número de etapas que eles requerem: **Detecção de Objetos em Uma Etapa: *You Only Look Once* (YOLO)**: Divide a imagem em células e prevê caixas delimitadoras e classes diretamente para cada célula em uma única passagem pela rede. ***Single Shot MultiBox Detector* (SSD)**: utiliza várias caixas delimitadoras em diferentes escalas para a detecção de objetos. **Detecção de Objetos em Duas (ou mais) Etapas: *Region Based Convolutional Neural Networks* (R-CNN)**: Ele utiliza a primeira etapa para extrair regiões de interesse e a segunda etapa para classificar essas regiões. **Fast R-CNN**: Uma melhoria do R-CNN. **Faster R-CNN**: Introduziu o conceito de uma Rede de Proposta de Região (RPN) para tornar a geração de regiões mais eficiente. **Mask R-CNN**: Uma extensão do Faster R-CNN que adiciona uma terceira etapa para segmentar objetos em detalhes, não apenas detectá-los (ULLAH, et al., 2023).

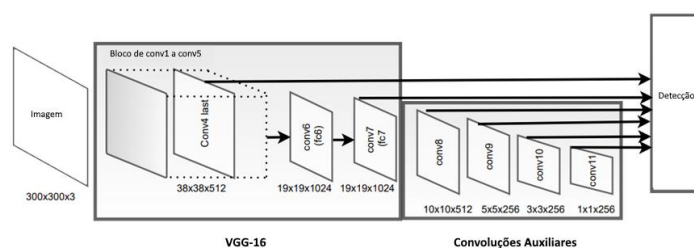
2.4 - SSD – Arquitetura

O SSD (*Single Shot MultiBox Detector*) é uma arquitetura de rede neural convolucional projetada para detecção de objetos. A abordagem do SSD permite detecções em uma única passagem pela rede, seguida por uma etapa de supressão não máxima (NMS) para obter as detecções finais (LIU et al., 2016).

A arquitetura do SSD, esquematizada na figura 2, consiste em duas partes principais: extração de características e detecção de objetos. Geralmente, é utilizado um modelo de classificação de alto desempenho, como a VGG16, para extrair características de alto nível da imagem de entrada, adicionando mapas de características (*feature maps*) adicionais com resoluções decrescentes para possibilitar a detecção em várias escalas. Camadas de convolução auxiliares são incorporadas no SSD para reduzir progressivamente o tamanho dos *feature maps* (LIU et al., 2016; MAGALHÃES et al., 2021).

O SSD utiliza âncoras padrão (*priors*) para limitar as variações nas formas das caixas delimitadoras (MAGALHÃES et al., 2021). Além disso, os preditores convolucionais geram previsões de detecção por meio de camadas de convolução paralelas para regressão e classificação (XIE et al., 2019), possibilitando que o SSD faça previsões sobre deslocamentos em relação às formas das caixas padrão e os escores por classe para detectar objetos em várias escalas e razões de aspecto usando múltiplos mapas de características (LIU et al., 2016).

Figura 2: Esquematização da Arquitetura do SSD



Fonte: Magalhães et al., 2021

Nesse esquema temos a entrada, uma imagem 300x300, a *backbone* VGG16, as convoluções auxiliares e a detecção (MAGALHÃES et al., 2021).

2.4 - SSD - Treinamento

O treinamento do SSD se diferencia de modelos convencionais de detecção de objetos devido à necessidade de associar informações *Ground Truth*, permitindo a retropropagação no processo ponta a ponta. A estratégia de correspondência é necessária para rotular as caixas delimitadoras padrão durante o treinamento, e o objetivo do treinamento envolve a função de perda total, que combina perda de confiança e perda de localização. Escalas e razões de aspecto são selecionadas durante o treinamento para permitir que diferentes camadas capturem detalhes e contextos globais. Para equilibrar exemplos positivos e negativos, o SSD utiliza a técnica de mineração de negativos difíceis. Além disso, o *Data Augmentation* é aplicado para melhorar o desempenho e a generalização do modelo (LIU et al., 2016; XIE et al., 2016; SHORTEN e KHOSHGOFTAAR, 2019).

3 – METODOLOGIA EXPERIMENTAL

A metodologia experimental será dividida em duas partes, sendo que, a primeira, mostra o discorrimento do princípio e metodologia praticada no *paper* de Liu et al. (2016). Na segunda parte, serão inseridas figuras deste próprio autor, para que o algoritmo com a respectiva metodologia, seja processado e demonstrado o resultado esperado do SSD.

Portanto, a primeira fase é um discorrimento e relato reduzido do procedimento de Liu et al. (2016), que pode ser visto no texto abaixo descrito.

A VGG16 foi utilizada como backbone do SSD, pré-treinada no dataset ILSVRC CLS-LOC. Algumas camadas da rede foram modificadas para melhorar o desempenho, e o modelo foi refinado usando o algoritmo Stochastic Gradient Descent (SGD) para evitar overfitting.

O modelo SSD é comparado com os métodos Fast R-CNN e Faster R-CNN, usando a VGG16 como *backbone*, treinado no *dataset* Pascal VOC 2007. A Tabela 1 demonstra que o modelo SSD300 supera o Fast R-CNN e o Faster R-CNN, especialmente com imagens maiores e treinamento adicional. Além disso, as Figuras 3 mostra que o SSD é capaz de detectar várias categorias de objetos com alta qualidade e enfrenta desafios na detecção de objetos menores e em categorias semelhantes.

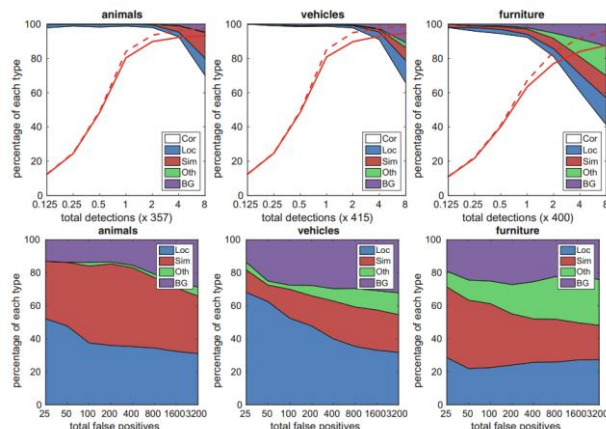
Tabela 1 – Resultados de Detecção com o Pascal VOC 2007 *test*

Method	data	mAP	aero	bike	bird	boat	bottle	bus	car	cat	chair	cow	table	dog	horse	mbike	person	plant	sheep	sofa	train	tv
Fast	07	66.9	74.5	78.3	69.2	53.2	36.6	77.3	78.2	82.0	40.7	72.7	67.9	79.6	79.2	73.0	69.0	30.1	65.4	70.2	75.8	65.8
Fast	07+12	70.0	77.0	78.1	69.3	59.4	38.3	81.6	78.6	86.7	42.8	78.8	68.9	84.7	82.0	76.6	69.9	31.8	70.1	74.8	80.4	70.4
Faster	07	69.9	70.0	80.6	70.1	57.3	49.9	78.2	80.4	82.0	52.2	75.3	67.2	80.3	79.8	75.0	76.3	39.1	68.3	67.3	81.1	67.6
Faster	07+12	73.2	76.5	79.0	70.9	65.5	52.1	83.1	84.7	86.4	52.0	81.9	65.7	84.8	84.6	77.5	76.7	38.8	73.6	73.9	83.0	72.6
Faster	07+12+COCO	78.8	84.3	82.0	77.7	68.9	65.7	88.1	88.4	88.9	63.6	86.3	70.8	85.9	87.6	80.1	82.3	53.6	80.4	75.8	86.6	78.9
SSD300	07	68.0	73.4	77.5	64.1	59.0	38.9	75.2	80.8	78.5	46.0	67.8	69.2	76.6	82.1	77.0	72.5	41.2	64.2	69.1	78.0	68.5
SSD300	07+12	74.1	74.6	80.2	72.2	66.2	47.1	82.9	83.4	86.1	54.4	78.5	73.9	84.4	84.5	82.4	76.1	48.6	74.3	75.0	84.3	74.0
SSD300	07+12+COCO	79.6	80.9	86.3	79.0	76.2	57.6	87.3	88.2	88.6	60.5	85.4	76.7	87.5	89.2	84.5	81.4	55.0	81.9	81.5	85.9	78.9
SSD512	07	71.6	75.1	81.4	69.8	60.8	46.3	82.6	84.7	84.1	48.5	75.0	67.4	82.3	83.9	79.4	76.6	44.9	69.9	69.1	78.1	71.8
SSD512	07+12	76.8	82.4	84.7	78.4	73.8	53.2	86.2	87.5	86.0	57.8	83.1	70.2	84.9	85.2	83.9	79.7	50.3	77.9	73.9	82.5	75.3
SSD512	07+12+COCO	81.5	86.9	87.5	82.0	75.5	66.4	88.2	88.7	89.3	65.2	88.3	74.4	87.1	88.9	85.9	84.5	57.6	84.6	80.7	87.1	81.7

Fonte: Liu et al., 2016.

O Fast e o Faster R-CNN usam imagens de entrada, cuja dimensão mínima é de 600 pixels. Os dois modelos SSD têm exatamente as mesmas configurações, exceto que possuem tamanhos de entrada diferentes (300×300 e 512×512). Os dados são definidos como: "07": treinamento e validação do VOC2007, "07+12": união do treinamento e validação do VOC2007 e VOC2012, "07+12+COCO": primeiro treinamento no COCO trainval35k e, em seguida, ajuste fino em "07+12".

Figura 3 - Visualização do desempenho do SSD 512 em animais, veículos e móveis no teste do VOC2007



Fonte: Liu et al., 2016.

Os gráficos na parte superior da imagem mostram a fração cumulativa de detecções que estão corretas (Cor) ou falsos positivos devido à má localização (Loc), confusão com categorias similares (Sim), com outras (Oth), ou com o fundo (BG). Já os da parte inferior mostram a distribuição dos tipos de falsos positivos mais bem classificados.

A análise detalhada do modelo SSD envolveu experimentos controlados, mantendo configurações consistentes e um tamanho de entrada de 300x300, destacando o impacto do aumento de dados no desempenho, bem como a importância de usar várias camadas de saída com diferentes resoluções para aprimorar a detecção de objetos.

Os modelos SSD300 e SSD512 foram treinados com o VOC2012 trainval e VOC2007 trainval/test. O SSD300 superou o Fast/Faster R-CNN em precisão e alcançou 4,5% de melhoria ao aumentar o tamanho das imagens para 512×512 . Comparado ao YOLO, o SSD é mais preciso, atingindo 80,0% de mAP no SSD512 após refinamento a partir de modelos COCO. Os resultados estão apresentados na tabela 2.

Tabela 2. Resultados de detecção no teste PASCAL VOC2012

Method	data	mAP	aero	bike	bird	boat	bottle	bus	car	cat	chair	cow	table	dog	horse	mbike	person	plant	sheep	sofa	train	tv
Fast[6]	07++12	68.4	82.3	78.4	70.8	52.3	38.7	77.8	71.6	89.3	44.2	73.0	55.0	87.5	80.5	80.8	72.0	35.1	68.3	65.7	80.4	64.2
Faster[2]	07++12	70.4	84.9	79.8	74.3	53.9	49.8	77.5	75.9	88.5	45.6	77.1	55.3	86.9	81.7	80.9	79.6	40.1	72.6	60.9	81.2	61.5
Faster[2]	07++12+COCO	75.9	87.4	83.6	76.8	62.9	59.6	81.9	82.0	91.3	54.9	82.6	59.0	89.0	85.5	84.7	84.1	52.2	78.9	65.5	85.4	70.2
YOLO[5]	07++12	57.9	77.0	67.2	57.7	38.3	22.7	68.3	55.9	81.4	36.2	60.8	48.5	77.2	72.3	71.3	63.5	28.9	52.2	54.8	73.9	50.8
SSD300	07++12	72.4	85.6	80.1	70.5	57.6	46.2	79.4	76.1	89.2	53.0	77.0	60.8	87.0	83.1	82.3	79.4	45.9	75.9	69.5	81.9	67.5
SSD300	07++12+COCO	77.5	90.2	83.3	76.3	63.0	53.6	83.8	82.8	92.0	59.7	82.7	63.5	89.3	87.6	85.9	84.3	52.6	82.5	74.1	88.4	74.2
SSD512	07++12	74.9	87.4	82.3	75.8	59.0	52.6	81.7	81.5	90.0	55.4	79.0	59.8	88.4	84.3	84.7	83.3	50.2	78.0	66.3	86.3	72.0
SSD512	07++12+COCO	80.0	90.7	86.8	80.5	67.8	60.8	86.3	85.5	93.5	63.2	85.7	64.4	90.9	89.0	88.9	86.8	57.2	85.1	72.8	88.4	75.9

Fonte: Liu et al., 2016.

O Fast/Faster R-CNN usa imagens com no mínimo 600x600 de dimensão, enquanto o YOLO recebe imagens 448×448 . O “07++12”: representa a união do VOC2007 trainval e test com o VOC2012 trainval. Já o “07++12+COCO”: Indica que o treinamento foi feito com o COCO trainval35k, então teve seu ajuste fino feito com o 07++12.

Para supressão não máxima eficiente, usamos um limiar de confiança de 0,01 e aplicamos o NMS com uma sobreposição de Jaccard de 0,45 por classe, mantendo as 200

melhores detecções por imagem. Isso é feito em cerca de 1,7 ms por imagem no modelo SSD300.

Na Tabela 3, o SSD300 e o SSD512 superam o Faster R-CNN em velocidade e precisão. O SSD300 é o primeiro método em tempo real com mAP acima de 70%. A otimização da rede base (VGG16) poderia tornar o SSD512 adequado para aplicações em tempo real.

Tabela 3. Resultados no teste do Pascal VOC2007.

Método	mAP	FPS	Tamanho de lote	Nº de caixas
Faster R-CNN (VGG16)	73.2	7	1	300
Faster R-CNN (ZF)	62.1	17	1	300
YOLO	63.4	45	1	98
Fast YOLO	52.7	155	1	98
SSD300	74.3	46	1	8732
SSD512	76.8	19	1	24564
SSD300	74.3	59	8	8732
SSD512	76.8	22	8	24564

Fonte: Liu et al., 2016.

O SSD300 é dos únicos métodos de detecção em tempo real que podem atingir uma precisão acima de 70% no mAP. Ao usar uma imagem de entrada maior, o SSD512 supera todos os métodos em precisão, mantendo uma velocidade próxima ao tempo real.

O uso de um tamanho de lote de teste de 8 melhora ainda mais a velocidade.

Na detecção de objetos em imagens, há duas abordagens principais: janelas deslizantes e classificação de propostas de regiões. Antes das redes neurais, métodos como DPM e *Selective Search* eram comuns. O R-CNN uniu propostas do *Selective Search* com CNNs, popularizando as abordagens por propostas de regiões. Outros métodos focam em melhorar as propostas usando redes neurais. Alguns preveem diretamente caixas delimitadoras e confianças para várias categorias de objetos, como *OverFeat* e YOLO. O SSD não usa propostas, empregando caixas padrão para maior flexibilidade.

Este estudo introduz o SSD, um detector de objetos eficiente e rápido com múltiplas saídas de caixas delimitadoras convolucionais em várias escalas. Demonstramos que o uso de caixas padrão bem escolhidas melhora o desempenho. Os modelos SSD superam outros detectores em termos de precisão e velocidade, especialmente o SSD512, que supera significativamente o Faster R-CNN. O SSD300 opera em tempo real a 59 FPS, superando alternativas como o YOLO em termos de precisão. Além disso, sugerimos que o modelo SSD pode ser um componente valioso em sistemas maiores e incentivamos pesquisas futuras sobre sua integração com redes neurais recorrentes para detecção e rastreamento de objetos em vídeo em tempo real.

Na segunda parte deste trabalho, utilizaremos o modelo SSD300, disponível no repositório que podemos acessar por esse link: <https://github.com/sgrvinod/a-PyTorch-Tutorial-to-Object-Detection>, treinado com o conjunto de dados Pascal VOC 2007 e

2012, para a detecção de objetos. As imagens-alvo incluirão uma moto e uma bicicleta. A inclusão da moto e da bicicleta visa verificar a precisão da identificação, identificando possíveis desafios ou confusões.

Figura 4 – Uma Motocicleta – Imagem sem e com Identificação do Objeto



Fonte: Os autores.

A figura 5 mostra a esquerda uma motocicleta, cujo processamento de SSD ainda não ocorreu. Nesta condição a imagem escolhida é analisada e identificada por um SSD treinado usando o *dataset* Pascal VOC 2007 e 2012. A análise e identificação é um processo intrínseco ao algoritmo SSD300, cujo resultado de identificação pode ser mostrado na parte direita da figura 5. A detecção e classificação foram corretas. O retângulo define todas as partes essenciais da motocicleta.

Figura 5 – Uma Bicicleta – Imagem com e sem Identificação do Objeto



Fonte: Os autores.

A figura 6 a esquerda o objeto antes da detecção e a direita após a detecção, nas mesmas condições da figura 5. A bicicleta foi identificada e rotulada corretamente.

4 – DISCUSSÃO GERAL

Este estudo analisou a aplicação do Single Shot MultiBox Detector (SSD) na visão computacional do mundo real. Concluímos que o SSD é altamente eficaz, superando concorrentes em precisão e velocidade, tornando-se valioso para aplicações práticas, como carros autônomos e vigilância. Além disso, ele inova ao simplificar o treinamento e a integração. Mas o SSD pode enfrentar desafios com objetos pequenos e categorias

semelhantes. Essas limitações merecem futuras investigações para aprimorar a visão computacional.

5 – CONCLUSÃO

O estudo enfoca o SSD como uma ferramenta promissora na visão computacional e detecção de objetos. Ele destaca a importância das CNNs e do SSD no processamento de informações visuais e demonstra sua eficácia por meio de resultados práticos, indicando um futuro promissor na aplicação de detecção de objetos em cenários do mundo real. Em resumo, o SSD representa uma abordagem eficaz e ágil na visão computacional, com potencial para aplicações práticas.

6 – PROJETOS FUTUROS

Nossos projetos futuros na área de detecção de objetos abrangem diversas possibilidades, incluindo: Desenvolvimento de Algoritmo SSD Personalizado e Otimização da Arquitetura e Criação de *Datasets* Específicos.

Essas iniciativas têm o objetivo de aprimorar a detecção de objetos, tornando-a mais precisa e adaptável a diversas aplicações em setores como agricultura e manufatura. Para isso, buscaremos colaborações com especialistas, avaliaremos o impacto de nossos esforços e estabeleceremos metas mensuráveis.

7- AGRADECIMENTOS

Agradeço ao Centro Paula Souza (CPS) pela oportunidade de desenvolver um trabalho de iniciação científica na unidade da Fatec Lins, no laboratório de desenvolvimento de sistemas.

REFERÊNCIAS

BROWNLEE, J. **SSD: Deep Learning for Computer Vision Image Classification, Object Detection, and Face Recognition in Python**, v.1.8, 563p. 2020. Disponível: https://www.google.com.br/books/edition/Deep_Learning_for_Computer_Vision/DOamDwAAQBAJ?hl=ptBR&gbpv=1&dq=Jason+Brownlee,+Deep+Learning+for+Computer+Vision,+2019&pg=PA56&printsec=frontcover.

FURTADO, M., I., V., **SSD: Redes Neurais Artificiais: Uma Abordagem Para Sala de Aula**, Atena Editora – Ponta Grossa/PR, 2019. p99. DOI:10.22533/at.ed.262191504

HAYKIN, S., **SSD: Redes Neurais: Princípios e prática**. Tradução: Paulo M. Engels. 2. ed. Porto Alegre: Artmed, 2007. 898 p. ISBN 0-13-273350-1

LI, X., LI, X., GUAN, S., JIN, J., & GE, Y.. Trident SSD: A Trident Single-Shot Multibox Object Detector with Deconvolution. **Journal of Physics: Conference Series**, v.1631, 1-11p. 2020. DOI:10.1088/1742-6596/1631/1/012182

LIU, W., ANGUELOV, D., ERHAN, D., SZEGEDY, C., REED, S., FU, C.Y., BERG, A.C., SSD: Single Shot MultiBox Detector, **Lecture Notes in Computer Science: Cornell University – Arxiv.org**, v.5, 21-37p., 21–37p. 2016. DOI:10.1007/978-3-319-46448-0_2

MAGALHÃES, S., A., CASTRO, L., MOREIRA, G., SANTOS, F., N., CUNHA, M., DIAS, J., MOREIRA, A., P., SSD: **Evaluating the Single-Shot MultiBox Detector and YOLO Deep Learning Models for the Detection of Tomatoes in a Greenhouse**, MDPI 2021.

MIN, S., LEE, B., YOON, S., **Deep learning in bioinformatics. Briefings in Bioinformatics**, Oxford University Press. 851-869, 2017. DOI:<https://doi.org/10.1093/bib/bbw068>.

NIELSEN, M., SSD: **Neural Networks and Deep Learning**, 224p. 2019. Disponível: <http://neuralnetworksanddeeplearning.com/>.

PENG, H., HUANG, B., SHAO, Y., LI, Z., ZHANG, C., CHEN, Y., XIONG, J., SSD: **General Improved SSD Model for Picking Object Recognition of Multiple Fruits in Natural Environment**, em **Transactions of the Chinese Society of Agricultural Engineering**, 2018.

PRINCE, J., D., S., SSD: **Computer Vision: Models, Learning, and Inference**. University College London, 579p. 2012. DOI: 9781107011793

SEBE, N., COHEN, I., GARG, A., HUANG., T.S., **Machine Learning in Computer Vision**, Springer, Netherlands, 2005. 249p.

SHANTINI, A., MANOGARAN, G., VADIVU, G., SSD: **Deep Convolutional Neural Network for The Prognosis of Diabetic Retinopathy**, Springer - Singapura, 2023.

SHORTEN, C., KHOSHGOFTAAR, T., M., SSD: **A survey on Image Data Augmentation for Deep Learning**, SpringerOpen, 48p. 2019. DOI:<https://doi.org/10.1186/s40537-019-0197-0>

ULLAH, N., MOHANTA, P., P., SSD: - **Recent Advances in Video Captioning with Object Detection**. Em MIR, R. N., SHARMA, V. K., ROUT, R. K., & UMER, S. (Eds.), **Advancement of Deep Learning and its Applications in Object Detection and Recognition**, River Publishers, 2023, 286p.

XIE, H., ZHANG, Y., WU, Z., SSD: **An Improved Fabric Defect Detection Method Based on SSD**, em **2nd Artificial Intelligence on Fashion and Textile International Conference**, v.8, 2019.

YAMASHITA, R., NISHIO, M., DO, R. K. G., TOGASHI, K., SSD: **Convolutional neural networks: an overview and application in radiology**, Springer, 19p. 2018. DOI:<https://doi.org/10.1007/s13244-018-0639-9>