



VII Congresso de Inovação e Tecnologia

IX Workshop de Iniciação Científica e Tecnológica

DESEMPENHO DO SSD EM DETECÇÃO DE OBJETOS: LIMITAÇÕES, CASOS DE FALHA E PROPOSTAS DE OTIMIZAÇÃO

Renan M. C. Gomes e Danilo F. Gomes, Análise Desenvolvimento de Sistemas
Fernando Muzzi, orientador, e Roberto Outa, coorientador



RESUMO

Este trabalho analisa o desempenho e as limitações do modelo de detecção de objetos Single Shot MultiBox Detector (SSD) em cenários complexos, como na detecção de objetos pequenos e sobrepostos, com foco em segurança e robótica. Apresentam-se otimizações, incluindo o refinamento das anchor boxes, integração de Feature Pyramid Networks (FPN), convoluções deformáveis e data augmentation. Essas melhorias visam aumentar a precisão e robustez do modelo em ambientes de alta densidade e variação de escala, mantendo a eficiência em tempo real.

INTRODUÇÃO

A detecção de objetos é fundamental para diversas aplicações, como segurança e robótica. O SSD, pela sua estrutura de única etapa, combina precisão e rapidez, tornando-o adequado para sistemas em tempo real. No entanto, apresenta dificuldades em cenários complexos, especialmente na detecção de objetos pequenos e sobrepostos, devido ao uso de anchor boxes fixas e à perda de detalhes nas camadas profundas.

CAUSAS DE DEFICIÊNCIA

Sobreposição de Objetos: As anchor boxes pré-definidas do SSD têm limitações para capturar objetos sobrepostos de forma eficaz, o que gera erros de classificação, especialmente em ambientes com muitos objetos próximos. Essa dificuldade ocorre porque o SSD não adapta dinamicamente os limites entre objetos próximos (Liu et al., 2016).

Detecção de Objetos Pequenos: O SSD apresenta problemas ao detectar objetos pequenos devido à redução de resolução dos feature maps nas camadas mais profundas, onde são capturadas menos informações detalhadas. Isso compromete a precisão, já que esses detalhes são cruciais para identificar objetos de menor escala com clareza (Xie et al., 2019).

Escalabilidade: Em ambientes de alta densidade de objetos, o SSD enfrenta limitações na detecção simultânea de objetos de escalas muito variadas, como grandes e pequenos. Essa dificuldade afeta seu desempenho em cenários complexos que exigem precisão em múltiplas escalas (Magalhães et al., 2021).

PROPOSTAS DE OTIMIZAÇÃO

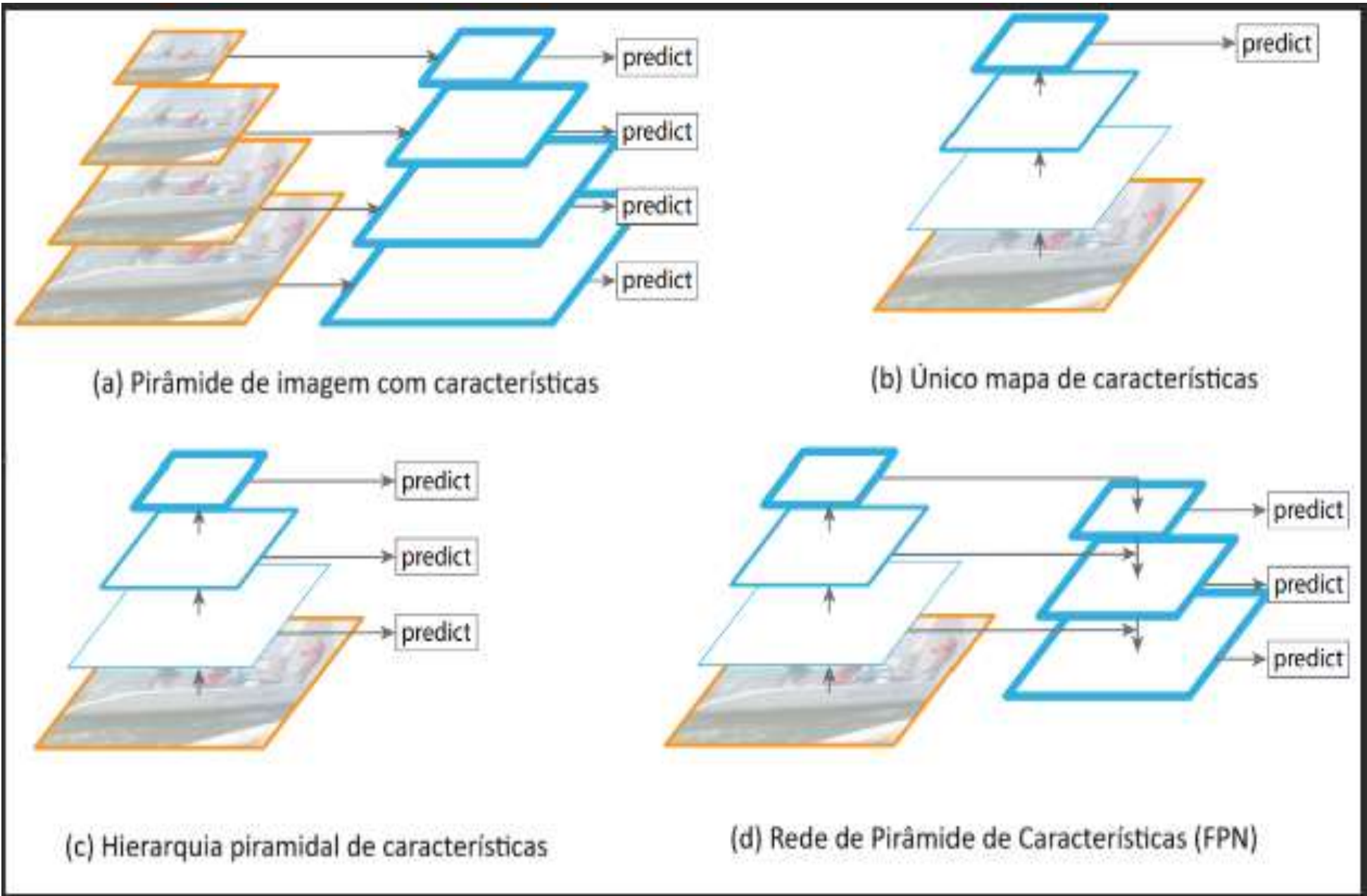
Aprimoramento das Anchor Boxes: Ajustar as dimensões das anchor boxes para refletir melhor os tamanhos dos objetos melhora a flexibilidade do modelo e aumenta a precisão na detecção de diferentes tamanhos de objetos (Liu et al., 2016).

Feature Pyramid Networks (FPN): A integração de FPNs permite ao SSD lidar com objetos de múltiplas escalas, combinando informações de camadas de diferentes profundidades. Essa técnica torna o modelo mais eficiente em captar detalhes e contextos variados (Lin et al., 2017).

Convoluções Deformáveis: Convoluções deformáveis permitem que o SSD ajuste dinamicamente seus campos receptivos, facilitando a detecção de objetos com formas complexas ou irregulares. Essa flexibilidade é essencial para melhorar a precisão em cenários de sobreposição (Dai et al., 2017).

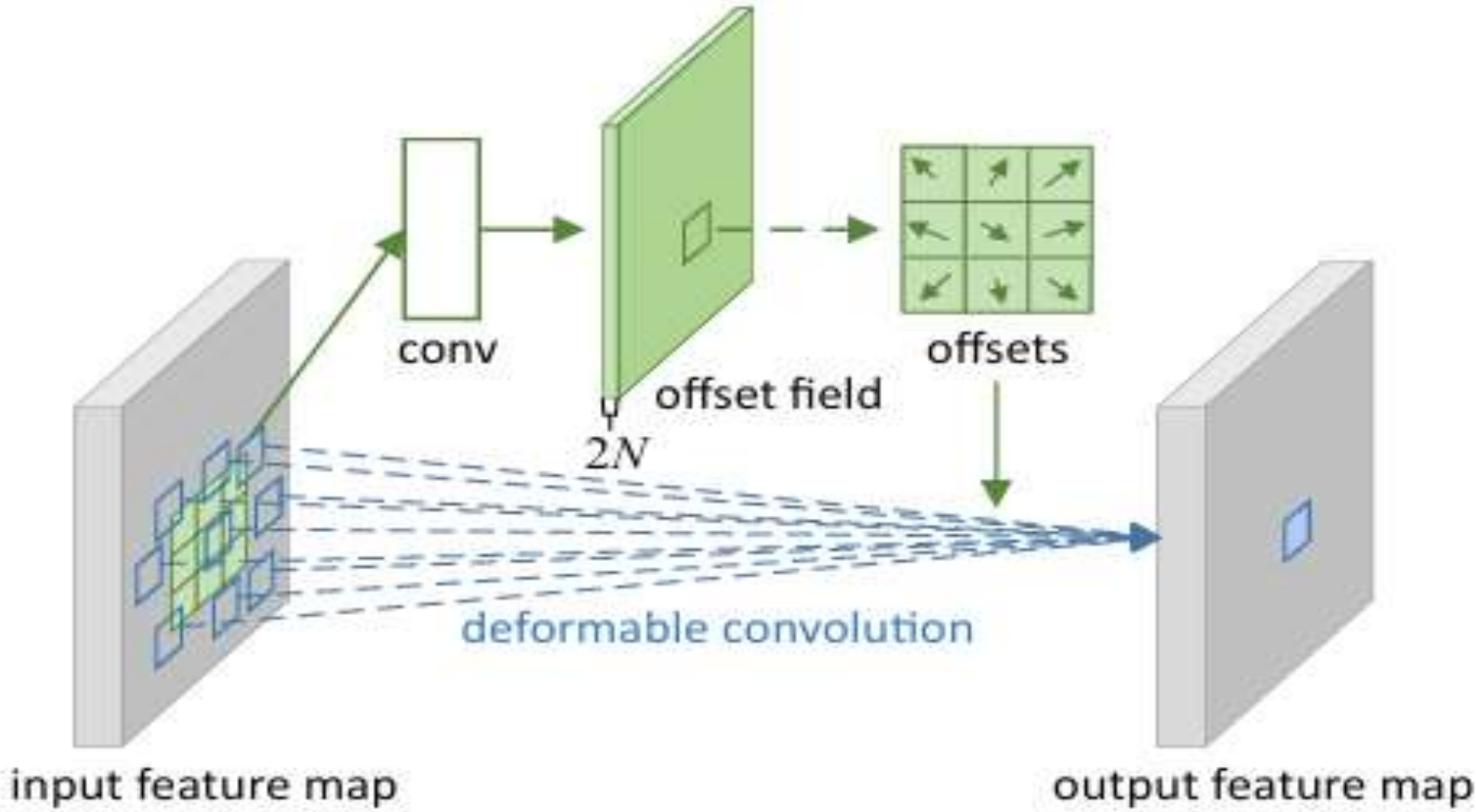
Data Augmentation: Técnicas avançadas, como GANs, aumentam a diversidade de dados no treinamento, ajudando o modelo a lidar melhor com variações de iluminação e posição. Esse aumento na diversidade ajuda o SSD a se adaptar a diferentes condições (Perez & Wang, 2017; Goodfellow et al., 2014).

Aprendizado Auto-Supervisionado (SSL): Estratégias de SSL permitem ao modelo SSD aprender representações robustas para detectar objetos em cenários com poucos dados rotulados, utilizando técnicas como DETReg e UP-DETR para maior eficácia (Chen et al., 2020; Bar et al., 2021; Dai et al., 2021).



Fonte: Chen et al., 2020.

Essa figura ilustra diferentes abordagens para a construção de pirâmides de características.



Fonte: Dai et al., 2017.

Temos nessa figura uma representação de convolução deformável

DISCUSSÃO

As otimizações propostas para o SSD, como o ajuste das anchor boxes e a integração das Feature Pyramid Networks (FPN), demonstram potencial para melhorar a detecção de objetos pequenos e sobrepostos. Convoluções deformáveis contribuem com ajuste dinâmico dos campos receptivos, enquanto a data augmentation e o aprendizado auto-supervisionado podem fortalecer o desempenho em cenários complexos. Contudo, essas adaptações trazem desafios, como o aumento da complexidade computacional, que pode limitar a aplicabilidade em dispositivos com recursos restritos.

CONCLUSÕES / CONSIDERAÇÕES FINAIS

O estudo mostrou que o SSD, com as adaptações propostas, apresenta potencial para melhorar a detecção em cenários de alta densidade e variabilidade de escala. Entretanto, essas melhorias precisam ser avaliadas em contextos práticos para verificar sua eficácia em condições adversas, mantendo a eficiência em tempo real. Como próximos passos, sugere-se explorar o aprendizado auto-supervisionado e a adaptação do SSD para dispositivos de baixo consumo, com possível aplicação em vigilância e robótica.

REFERÊNCIAS BIBLIOGRÁFICAS

ASLAM, A. et al. A Survey of Modern Deep Learning Based Object Detection Models. arXiv preprint arXiv:2104.11892, 2021. Disponível em: <https://arxiv.org/abs/2104.11892>. Acesso em: 9 dez. 2023.

BAR, A.; ARBEL, M.; HASSNER, T.; CHECHIK, G. DETReg: Unsupervised Pretraining with Region Priors for Object Detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021. p. 9658–9668. DOI: 10.1109/CVPR46437.2021.00956.

CHEN, X.; FAN, H.; GIRSHICK, R.; HE, K. Improved Baselines with Momentum Contrastive Learning. arXiv preprint arXiv:2003.04297, 2020. Disponível em: <https://arxiv.org/abs/2003.04297>.

DAI, Z.; CAI, B.; LIN, Y.; CHEN, J. UP-DETR: Unsupervised Pre-Training for Object Detection with Transformers. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021. p. 1601–1610. DOI: 10.1109/CVPR46437.2021.00163.

GOODFELLOW, I. et al. Generative Adversarial Nets. In: Advances in Neural Information Processing Systems, 2014. p. 2672–2680. Disponível em: <https://doi.org/10.48550/arXiv.1406.2661>.

LIN, T.-Y.; DOLLÁR, P.; GIRSHICK, R.; HE, K.; HARIHARAN, B.; BELONGIE, S. Feature Pyramid Networks for Object Detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017. DOI: 10.1109/CVPR.2017.106.

DAI, J. et al. Deformable Convolutional Networks. In: Proceedings of the IEEE International Conference on Computer Vision, 2017. DOI: 10.1109/ICCV.2017.89.

LIU, W. et al. SSD: Single Shot MultiBox Detector. In: European Conference on Computer Vision, 2016. DOI: 10.1007/978-3-319-46448-0_2.

MAGALHÃES, S. et al. Evaluating the Single-Shot MultiBox Detector and YOLO Deep Learning Models for the Detection of Tomatoes in a Greenhouse. Sensors, 2021. DOI: 10.3390/s21020423.

PEREZ, L.; WANG, J. The Effectiveness of Data Augmentation in Image Classification Using Deep Learning. arXiv preprint arXiv:1712.04621, 2017. Disponível em: <https://arxiv.org/abs/1712.04621>.