

APLICAÇÃO DE REDES NEURAIIS CONVOLUCIONAIS PARA DETECÇÃO E CONTAGEM DE PRODUTOS EM GÔNDOLAS DE SUPERMERCADOS UTILIZANDO YOLOV11 E RETINANET

Mariele Sarmiento Cavalca – marieles207@gmail.com

Matheus Brendon Francisco – matheus_brendon@unifei.edu.br

Carlos Henrique Pereira Mello – carlos.mello@unifei.edu.br

O estudo propõe o uso de redes neurais convolucionais (CNNs) para detecção e contagem automatizada de produtos em gôndolas de supermercados, visando reduzir rupturas e otimizar a gestão de estoques. A metodologia adota o framework CRISP-DM, estruturado em compreensão, preparação, modelagem e avaliação. Serão testadas as arquiteturas YOLOv11 e RetinaNet em imagens reais capturadas sob diferentes condições de luminosidade e disposição. Espera-se que a YOLOv11, devido à sua estrutura *end-to-end* aprimorada, apresente melhor desempenho em tempo real ($mAP@0.5 \geq 0,90$), enquanto a RetinaNet, apoiada na *Focal Loss*, demonstre maior sensibilidade na detecção de objetos pequenos ($recall \geq 0,88$). Acredita-se que o sistema permitirá identificar e contar produtos com erro médio absoluto inferior a ± 1 item por imagem, configurando-se como solução acessível e eficaz para o varejo inteligente.

Palavras-chave: visão computacional; redes neurais convolucionais; YOLOv11; RetinaNet; gestão de gôndolas.

1. INTRODUÇÃO

O setor supermercadista brasileiro movimentou, em 2024, cerca de R\$ 1,067 trilhão, equivalente a 9,12% do PIB nacional (ABRAS, 2025). Esse peso econômico reforça a necessidade de processos eficientes no gerenciamento de gôndolas, que afetam diretamente a disponibilidade de produtos e a experiência de compra. A ruptura de gôndola, caracterizada pela falta de produtos disponíveis, gera perdas anuais de US\$ 1,7 trilhão no varejo mundial (IHL Group, 2024). No Brasil, a indisponibilidade média varia entre 12% e 14% (Neogrid, 2025), comprometendo faturamento e fidelidade.

Apesar dos avanços logísticos, o monitoramento ainda depende de auditorias manuais, lentas e suscetíveis a erros (Sovis, 2024). A Inteligência Artificial (IA) surge como ferramenta estratégica ao permitir inspeção visual automatizada de gôndolas, garantindo atualização contínua e execução de planogramas (Fortune Business Insights, 2024).

Casos como o da rede Asda, que utiliza câmeras inteligentes para identificar produtos em falta, demonstram a viabilidade da abordagem (The Sun, 2025). Entre as técnicas de IA, as redes neurais convolucionais (CNNs) destacam-se por extrair padrões visuais automaticamente, superando métodos tradicionais (Goodfellow; Bengio; Courville, 2016). Arquiteturas como YOLO e RetinaNet combinam rapidez e precisão, sendo promissoras para a detecção e contagem de produtos (Redmon et al., 2016; Lin et al., 2017).

O estudo objetiva desenvolver e comparar os modelos YOLOv11 e RetinaNet aplicados à identificação e contagem de produtos, contribuindo para a automação do varejo e a transformação digital na gestão supermercadista.

2. DESENVOLVIMENTO

2.1 Referencial teórico

2.1.1 Redes convolucionais (CNNs)

As CNNs são compostas por camadas de convolução, *pooling* e camadas totalmente conectadas. A convolução aplica filtros que capturam bordas e texturas, enquanto o *pooling* reduz a dimensionalidade preservando informações relevantes (Scherer et al., 2010). Essa

estrutura permite reconhecer padrões complexos, tornando-se amplamente usada em visão computacional.

A YOLO (*You Only Look Once*) reformulou a detecção de objetos como um único problema de regressão, prevendo classes e caixas delimitadoras em tempo real (Redmon et al., 2016). Versões recentes, como a YOLOv11, incorporam *transfer learning* e backbones otimizados, aumentando a precisão e reduzindo o tempo de inferência.

Já a RetinaNet, proposta por Lin et al. (2017), introduziu a *Focal Loss*, que reduz o impacto de exemplos fáceis e enfatiza os difíceis, tornando-a eficaz em cenários com desbalanceamento e objetos pequenos.

Embora apresente maior custo computacional, destaca-se pela precisão em prateleiras com produtos sobrepostos. As arquiteturas são avaliadas por métricas como *precision*, *recall*, *F1-score* e *mean Average Precision* (mAP), que mensuram acurácia e abrangência das predições (Everingham et al., 2015).

2.1.2 Arquiteturas das CNNs

Em detecção de objetos em ambientes complexos, como gôndolas, CNNs tradicionais voltadas apenas à classificação são insuficientes. Arquiteturas especializadas, como YOLO e RetinaNet, localizam e identificam múltiplos itens simultaneamente (Redmon et al., 2016; Lin et al., 2017).

A YOLO converte a detecção em regressão única, prevendo *bounding boxes* e probabilidades em um único passo, ideal para análises em tempo real. As versões recentes incorporam *data augmentation* e regularização, aumentando a precisão mesmo em imagens com múltiplos objetos (Jocher et al., 2023).

A RetinaNet foi projetada para lidar com desbalanceamento de classes e objetos pequenos, utilizando *Focal Loss* para priorizar exemplos complexos. Apesar de maior tempo de inferência, apresenta excelente desempenho em detecções detalhadas. Ambas utilizam backbones como ResNet e EfficientNet para extrair características hierárquicas (He et al., 2016; Tan & Le, 2019).

O uso de *transfer learning* reduz a necessidade de grandes bases, acelerando o treinamento (Yosinski et al., 2014). Técnicas de *data augmentation* como rotações e ajustes

de brilho aumentam a robustez, permitindo reconhecer produtos sob diferentes condições (Shorten & Khoshgoftaar, 2019). Como limitações, as arquiteturas exigem GPU e podem ter dificuldade em generalizar produtos não representados no treinamento. YOLO tende a errar em objetos muito pequenos, enquanto RetinaNet é mais lenta (Redmon et al., 2016; Lin et al., 2017). O ajuste de hiperparâmetros também é um desafio prático.

2.1.2 Gestão de gôndola

A gestão de gôndolas é estratégica, pois conecta disponibilidade de produtos, experiência do cliente e desempenho financeiro. Mais do que organização, envolve planejamento de layout, categorias e políticas de reposição.

Com a Indústria 4.0, soluções de IoT e *analytics* ampliaram a precisão nas previsões e reduziram perdas (Carpitella; Izquierdo, 2025). Ainda assim, a ruptura de gôndola persiste, levando consumidores à substituição ou desistência da compra (Gruen; Corsten; Bharadwaj, 2002).

Auditorias manuais continuam predominantes, caras e suscetíveis a erro (Sovis, 2024). Alternativas com visão computacional, como as câmeras inteligentes implantadas pela Asda em 2025, demonstram o potencial de automação (The Sun, 2025). Pesquisas recentes reforçam avanços e lacunas. Destacam-se detectores YOLO aplicados diretamente em gôndolas e abordagens multimodais que integram visão computacional e OCR (Tonioni; Serra; Di Stefano, 2018; Oliveira; Costa; Mesquita, 2024; Muñoz; Sanchez; Kemper, 2024; Pettersson; Riveiro; Löfström, 2024).

No Brasil, Oliveira, Costa e Mesquita (2024) demonstraram a viabilidade do YOLOv8 em mini gôndolas com hardware acessível, enquanto Muñoz, Sanchez e Kemper (2024) obtiveram acurácia superior a 98% na detecção de bebidas. Tais estudos evidenciam o avanço das soluções, mas também lacunas na aplicação em escala, abrindo espaço para investigações comparativas em contextos reais.

Quadro 1 – Síntese de estudos relacionados sobre reconhecimento e contagem de produtos em gôndolas.

Autor / Ano	Objetivo principal	Abordagem técnica	Métricas de contagem reportadas	Lacunas identificadas
Tonioni, Serra & Di Stefano (2018)	Reconhecimento em larga escala de itens em gôndolas.	CNN para detecção + descritores globais + <i>K-NN</i> para reconhecimento; re-ranking de resultados.	Não reporta métricas de contagem explícitas; foca em acurácia de reconhecimento.	Alto custo de treinamento; dependência de imagens de referência curadas; ausência de métricas quantitativas de contagem.
Oliveira, Costa & Mesquita (2024)	Propor sistema de VC para apoio ao inventário em <i>smart markets</i> .	YOLOv8 aplicado em mini gôndola; hardware acessível (GPU/Colab).	Não apresenta métricas quantitativas de contagem; apenas demonstra viabilidade.	Ausência de resultados numéricos robustos (mAP, F1, MAE); validação experimental restrita.
Muñoz, Sanchez & Kemper (2024)	Detectar níveis de estoque de bebidas em tempo real.	CNN para validação + YOLOv2 (<i>transfer learning</i>); câmeras fixas + <i>single-board computers</i> ; app móvel.	Acurácia > 98% na contagem de garrafas; métricas adicionais: Acurácia de 99,72%, Precisão de 99,60%, Sensibilidade de 100%, Especificidade de 99,02%.	Escopo restrito a bebidas; não utiliza métricas de erro de contagem (MAE, RMSE).
Pettersson, Riveiro & Löfström (2024)	Reconhecimento <i>fine-grained</i> de produtos semelhantes.	CNNs + OCR (Transformer); <i>dataset FineGrainOCR</i> (256 classes); uso de Locount dataset com instâncias por posição de prateleira.	Contagem presente no <i>dataset</i> Locount (<i>instance count</i>), mas usada como suporte à classificação.	Contagem não é métrica central; complexidade maior pela multimodalidade; <i>dataset</i> restrito a categorias específicas.

Fonte: Elaboração própria com base em Tonioni, Serra & Di Stefano (2018), Oliveira, Costa & Mesquita (2024), Muñoz, Sanchez & Kemper (2024) e Pettersson, Riveiro & Löfström (2024).

2.2 Metodologia

O desenvolvimento segue o modelo CRISP-DM (*Cross Industry Standard Process for Data Mining*) (Chapman et al., 2000), adaptado à visão computacional, dividido em seis fases:

1. Compreensão do negócio: definição do problema da ruptura de gôndolas como desafio de detecção e contagem de produtos.

2. Compreensão dos dados: coleta de imagens reais capturadas em diferentes condições de luminosidade e ângulo.
3. Preparação dos dados: limpeza, anotação e divisão em treino, validação e teste, com *data augmentation* (giros, espelhamentos e ajustes de brilho).
4. Modelagem: treinamento supervisionado das arquiteturas YOLOv11 e RetinaNet, utilizando Python e GPU no Google Colab, com ajuste de hiperparâmetros.
5. Avaliação: comparação de desempenho por meio de métricas *precision*, *recall*, *F1-score*, *IoU* e *mAP@0.5*, além de erro médio absoluto (MAE) para a contagem.
6. Implantação: criação de interface experimental em Colab que realiza detecção, contagem produtos.

4. CONSIDERAÇÕES FINAIS

O desenvolvimento deste estudo deverá confirmar o potencial das redes neurais convolucionais na automação da gestão de gôndolas. Espera-se que a YOLOv11 e a RetinaNet demonstrem complementaridade, unindo velocidade, precisão e sensibilidade, e que o sistema resultante se consolide como um modelo de auditoria inteligente, replicável e economicamente viável. O trabalho também deverá contribuir metodologicamente ao aplicar o CRISP-DM em um contexto de visão computacional, fortalecendo o uso da inteligência artificial como ferramenta estratégica da Indústria 4.0 no setor supermercadista.

REFERÊNCIAS

- ABRAS – ASSOCIAÇÃO BRASILEIRA DE SUPERMERCADOS. **Ranking ABRAS 2025**. São Paulo: ABRAS, 2025.
- CAMPO, K.; GIJSBRECHTS, E.; NISOL, P. **Towards understanding consumer response to stock-outs**. *Journal of Retailing*, v. 76, n. 2, p. 219–242, 2000.
- CARPITELLA, S.; IZQUIERDO, M. **Smart retail and the role of IoT in shelf management**. *International Journal of Production Research*, v. 63, n. 4, p. 745–762, 2025.
- CHAPMAN, P. et al. **CRISP-DM 1.0: Step-by-step data mining guide**. SPSS Inc., 2000.
- EVERINGHAM, M. et al. **The PASCAL Visual Object Classes Challenge: A retrospective**. *International Journal of Computer Vision*, v. 111, n. 1, p. 98–136, 2015.
- FORTUNE BUSINESS INSIGHTS. **Artificial Intelligence in Retail Market Size, Share and Forecast (2023–2032)**. Londres: Fortune Business Insights, 2024.
- GHOSH, A.; CHATTERJEE, S. **Deep learning architectures and applications**. *AI Review*, v. 45, n. 3, p. 556–572, 2024.
- GONDAL, M. W. et al. **Computer vision-based retail shelf analysis: Techniques and challenges**. *Applied Sciences*, v. 11, n. 21, p. 10234, 2021.
- GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep Learning**. Cambridge: MIT Press, 2016.
- GRUEN, T.; CORSTEN, D.; BHARADWAJ, S. **Retail out-of-stock: A worldwide examination of causes, rates, and consumer responses**. *Grocery Manufacturers of America*, Washington, 2002.
- HE, K. et al. **Deep residual learning for image recognition**. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016. p. 770–778.
- HL GROUP. **Out-of-stock report: global losses and trends**. Nashville: IHL Group, 2024.
- IONOS. **How convolutional neural networks work**. Disponível em: <https://www.ionos.com/>. Acesso em: 10 set. 2025.
- JOCHER, G. et al. **YOLOv8: Ultralytics YOLO Implementation**. *GitHub repository*, 2023. Disponível em: <https://github.com/ultralytics/ultralytics>.
- LIN, T. Y. et al. **Focal Loss for Dense Object Detection**. In: *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*. 2017. p. 2980–2988.
- NEOGRID. **Índice de Ruptura do Varejo Brasileiro – Relatório 2025**. São Paulo: Neogrid, 2025.
- PETTERSSON, N.; RIVEIRO, M.; LÖFSTRÖM, T. **Multimodal retail product recognition using CNNs and OCR integration**. *Computers & Industrial Engineering*, v. 187, p. 109–118, 2024.
- REDMON, J. et al. **You Only Look Once: Unified, real-time object detection**. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016. p. 779–788.

SHORTEN, C.; KHOSHGOFTAAR, T. M. **A survey on image data augmentation for deep learning.** *Journal of Big Data*, v. 6, n. 1, p. 60, 2019.

SOVIS, L. **Manual vs. automated audits in grocery retail: efficiency and errors.** *Retail Operations Journal*, v. 14, n. 3, p. 200–214, 2024.

TAULLI, T. **Artificial Intelligence Basics: A non-technical introduction.** Berkeley: Apress, 2020.

THE SUN. **Asda installs AI cameras to monitor shelf stock and expiry dates.** *The Sun*, Londres, 2025.
Disponível em: <https://www.thesun.co.uk>.

TONIONI, A.; SERRA, E.; DI STEFANO, L. **A Deep Learning Pipeline for Product Recognition on Store Shelves.** *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2018.