

Reconhecimento de Objetos em Prateleiras de Supermercados

Tiago L. P. D. Fonseca¹, Prof. Dr. Edson J. R. Justino²

¹Escola politécnica

²Programa de Pós-Graduação em Informática

Pontifícia Universidade Católica do Paraná, PUCPR
R. Imaculada Conceição, 1155 – 80215-901 – Curitiba – PR

tiagoapolo@gmail.com, edson.justino@pucpr.br

Abstract. *This paper presents an approach for identifying and locating products on supermarket shelves with a focus on distinguishing different flavors of the same product, whose problem remains a challenge due to the dynamism of the context in which the product is inserted and the frequent change in packaging. Whose tests resulted in an accuracy of 83.6% and precision of 87.0%.*

Resumo. *Este artigo apresenta uma abordagem para a identificação e localização de produtos em prateleiras de supermercados com foco em distinguir diferentes sabores de um mesmo produto, cujo problema continua sendo um desafio devido ao dinamismo do contexto ao qual o produto está inserido e a frequente mudança nas embalagens. Cujos resultados dos testes apresentaram uma acurácia de 83,6% e uma precisão de 87,0%.*

1. Introdução

A identificação automática de objetos aplicada no reconhecimento de produtos, mais especificamente em supermercados, tem recebido crescente atenção devido as oportunidades de melhoria na experiência de compra (e.g., compras sem fila e sem atendimento, sistemas de tecnologia assistiva para deficientes visuais e auditivos, aplicativos de compras) e aumento na eficiência da gestão da loja (e.g., monitoramento de prateleira, sistemas antifurto, validação de planograma).

Segundo [MERLER et al., 2007] em seu trabalho a identificação de produtos em prateleiras é um problema clássico de reconhecimento de objetos, porém composto de desafios peculiares. Pois apresenta um cenário complexo devido à alta dinamicidade na representação dos produtos, com mudanças relevantes em tamanho, posição, condições de iluminação e oclusão. Por um lado, produtos diferentes podem parecer muito similares em razão das pequenas variações físicas (e.g., diferenças no texto do produto ou na coloração da embalagem), isso usualmente acontece quando um produto apresenta diferentes sabores. Por outro lado, um produto pode apresentar edições comemorativas ou promocionais com alterações na sua embalagem (Figura 1). Vale a pena citar que a classificação e identificação em tempo real de macro categorias (e.g., chocolate, café, biscoito) é relativamente mais simples que reconhecer produtos específicos ou microcategorias (e.g., Coca-Cola Zero, Coca-Cola Original, Campo Largo chá mate).

A aplicação de um sistema de visão computacional capaz de reconhecer microcategorias em prateleiras de supermercados habilitará o próximo passo de inovação aos supermercados e varejistas, que estão investindo em tecnologias que facilitem a vida dos clientes (e.g., compras sem filas) e tragam vantagens competitivas

perante aos concorrentes. Portanto neste trabalho será criado um sistema capaz de reconhecer em tempo real microcategorias de produtos em prateleiras de mercado a partir de imagens capturadas por vídeo.



Figura 1. Produtos diferentes com variabilidade pequena e produtos iguais com embalagens diferentes.

2. Fundamentação Teórica

O reconhecimento de objetos de uma forma genérica é um dos principais problemas da visão computacional, amplamente estudado pela comunidade científica e com diversas soluções propostas de extratores e identificadores de características, cuja função é identificar e localizar objetos nas imagens. Utilizando algoritmos como HOG e SVM [DALAL et al., 2005] para resolução de problemas reais. Contudo, a disponibilidade de grandes bases de imagens e recursos computacionais a preços acessíveis [DENG et al., 2009] incentivaram e viabilizaram o uso de aprendizado profundo (e.g., SSD, YOLO, Faster R-CNN) e definiram como nova técnica mais usada para detecção de objetos [LIU et al., 2016].

2.1. Reconhecimento de Produtos em Prateleiras

O problema de reconhecimento objetos com ênfase em produtos de prateleiras foi endereçado na pesquisa de [GEORGE et al., 2014], cuja contribuição foi o estudo do reconhecimento de classes de objetos em prateleiras utilizando uma base de imagens (GroceryProducts) coletada pelos autores e rotuladas em classes com o objetivo de reconhecer macro categorias de produtos.

2.2. Reconhecimento em Tempo Real com Redes Neurais

Entre os algoritmos de reconhecimento de objetos em tempo real, o Single Shot Detector apresenta o uma boa performance com 74% de precisão média a 59 quadros por segundo na extração de características de imagens e reconhecimento de objetos por [LIU et al., 2016]. Cuja arquitetura utiliza é uma imagem de tamanho fixo, 300x300 (Figura 2) como entrada. A restrição de tamanho fixo é principalmente para obter um treinamento eficiente. Enquanto a saída é um mapa de características ou chamado também de mapa de previsões com 8732 previsões possíveis, porém o SSD filtra a maioria das previsões. Além de usar a rede VGG-16 [SIMONYAN et al., 2015] como base, mas descarta as camadas totalmente conectadas.

O algoritmo em questão apresenta características que se adequam ao trabalho aqui proposto, alta precisão em tempo real, adequação a representação do objeto em

escalas diferentes e o uso de transferência de conhecimento permitindo a adequação do algoritmo ao contexto de reconhecimento desejado com um número reduzido de imagens de treino.

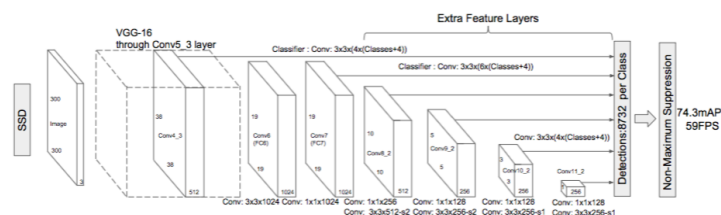


Figura 2. Arquitetura do SSD usando entrada de 300x300x3

2.3. Base de imagens

A aquisição de imagens para construção da base de imagens é uma etapa fundamental para o treinamento do modelo e a medição dos resultados. O método de captura de imagens no contexto de produtos em prateleiras utilizando captura de vídeo usando a configuração de 29 quadros por segundo e posteriormente foi selecionando uma foto do produto a cada 5 quadros do vídeo [MERLER et al., 2007], contém as variações de iluminação, cor, pose e fundo necessárias para criar um maior dinamismo do modelo e maior precisão. Portanto, sendo um método de aquisição de imagens candidato para este trabalho.

2.4. Reconhecimento de microcategorias

Finalmente, [VAROL et al., 2014], uma abordagem próxima a deste trabalho, apresenta uma estratégia para reconhecimento de microcategorias de cigarros em prateleiras. Utiliza dos algoritmos HOG, SVM e SIFT, com o propósito de obter características da forma do objeto sem depender da escala de representação e usando o contexto dos produtos próximos aos de interesse como uma característica, ou seja, produtos diferentes dos desejados, mas contidos na mesma prateleira são levados em consideração no processo de identificação das microcategorias dos pacotes de cigarros. Contudo, a utilização do contexto como uma característica de identificação inviabiliza a generalização da solução, pois cada supermercado pode ter uma disposição diferente dos pacotes de cigarro, alterando o contexto.

3. Metodologia

Este trabalho utiliza o método de investigação experimental, pois envolve a manipulação da extração de características identificadoras dos objetos nas imagens no grupo definido de cinco classes de produtos, Coca-Cola lata 350ml, Coca-Cola lata 310ml, Coca-Cola sem açúcar lata 350ml, Coca-Cola garrafa 600ml e o Campo Largo Chá Mate, com a finalidade de atingir uma performance superior a 50% na identificação de microcategorias. Sendo o trabalho dividido em quatro etapas: a aquisição da base de imagens, pré-processamento, processamento das imagens e análise dos resultados.

3.1. Aquisição da base de imagens

Para a fase de aquisição das fotos usadas na criação da base de imagens serão usadas duas abordagens, escaneamento de objetos em 360 graus e vídeos dos produtos nas prateleiras dos mercados, com a finalidade de obter um número expressivo de imagens para cada uma das cinco classes definidas.

3.1.1. Escaneamento 360 graus do objeto

Foi utilizado um scanner de objetos (Figura 3), disponibilizado pelo laboratório de engenharia elétrica da PUC-PR, cuja funcionalidade é tirar sequências de fotos ao redor

do objeto, permitindo configurar a quantidade de fotos por segundo e a angulação da câmera em relação ao objeto. O que permitirá obter uma quantidade massiva de fotos de cada uma das cinco classes, para uma melhor representação do objeto, permitindo a rede neural lidar com maior sucesso casos de oclusão e identificação por aparição parcial do objeto.

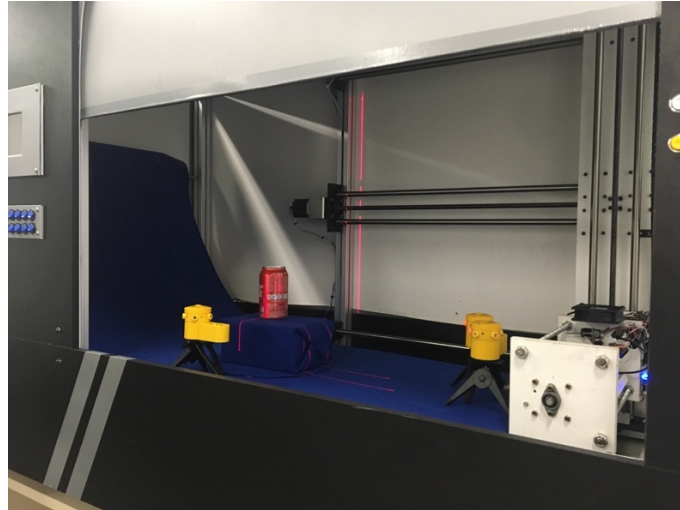


Figura 3. Scanner 360 graus de objetos

3.1.2. Fotos dos produtos no mercado

Usando a câmera de um celular serão filmados 5 vídeos no total, um para cada classe desejada no trabalho e presente nas prateleiras do supermercado, usando a configuração de 1920x1080 pixels e 30 quadros por segundo e selecionando uma foto a cada 5 quadros. Assim, capturando cenários onde existam iluminações distintas e produtos contidos quanto não contidos nas classes especificadas anteriormente.

3.2. Pré-processamento

O pré-processamento é composto de duas partes, o tratamento das fotos coletadas para base de imagens, onde as fotos serão tratadas usando algoritmo CLAHE, que é uma técnica de processamento de imagens usada para melhorar o contraste usando um método adaptativo, e a rotulagem delas para informar a localização da classe de interesse a rede neural.

3.3. Processamento

O processamento é dividido em duas etapas, o treinamento do modelo, tendo como entrada um arquivo XML com as rotulagens e a base de imagens, responsável por reconhecer as cinco classes definidas e a previsão em tempo real, responsável por avaliar as imagens capturadas pelo smartphone e prever as classes dos objetos apresentados.

3.4. Análise dos resultados

Para análise dos resultados será usada o valor da média de AP ou mean average precision, que consiste na media entre as medias obtidas da precisão média, e da matriz de confusão, onde são medidas as frequências de classificação para cada classe do modelo, ao final de cada treino afim de calcular a precisão e a acurácia obtidas do modelo e entender se houve progresso ou regresso ao aumentar de imagens da base ou o número de iterações durante o treino.

4. Resultados

Na figura 4 é apresentada a média AP versus a quantidade de iterações realizadas no treino do modelo sendo possível observar que o treino com 7500 iterações e 100 imagens por classe obteve o máximo de performance ao obter um escore de 0,92 de média AP e os treinos com iterações maiores mostraram valores de escore menores. Portanto, o modelo escolhido para realizar os testes de classificação e cujos valores obtidos dos testes estão presentes na tabela 1, que foi usada como base para o cálculo da precisão e da acurácia e cujos valores obtidos respectivamente foram de 87,0% e 83,6%.

Analisando a tabela 1, observa-se que apesar da forma semelhante entre as classes Coca 350ml Sem açúcar e Coca 350ml Original, o modelo treinado foi capaz de obter uma acurácia maior que 80% e para as classes com maiores distinções entre suas formas obteve uma acurácia acima de 90%.

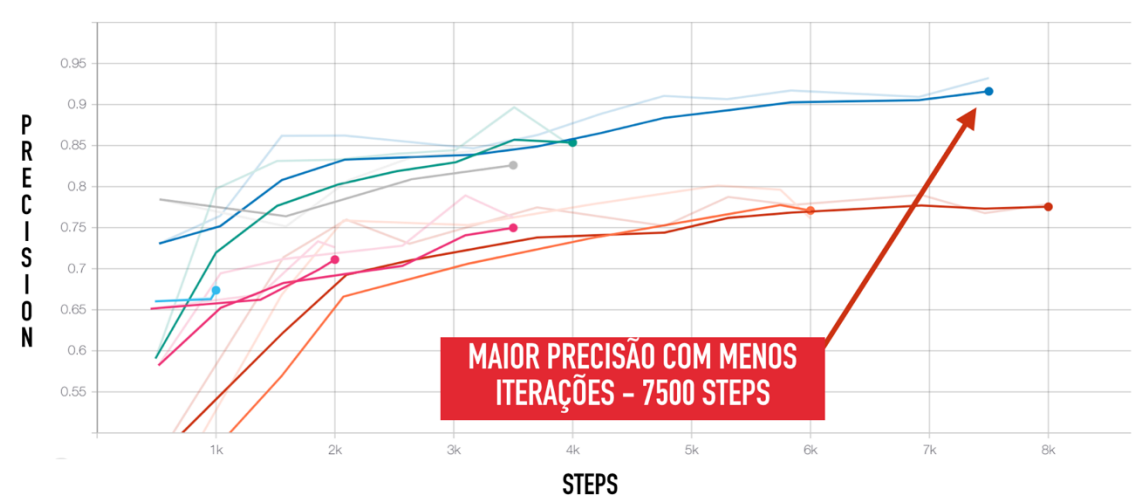


Figura 4. Análise comparativa da função média AP @.75

Tabela 1. Matriz de confusão

| | Coca 350ml Sem açúcar | Coca 350ml Original | Coca 310ml Original | Campo Largo | Coca-Cola 600ml |
|--------------------------|--------------------------|------------------------|------------------------|----------------|--------------------|
| Coca 350ml Sem açúcar | 85 | 15 | 0 | 0 | 0 |
| Coca 350ml Original | 20 | 75 | 5 | 0 | 0 |
| Coca 310ml Original | 0 | 20 | 80 | 0 | 0 |
| Campo Largo | 0 | 0 | 0 | 100 | 0 |
| Coca-Cola 600ml | 0 | 0 | 0 | 10 | 90 |

6. Conclusões e trabalhos futuros

Os resultados obtidos demonstraram uma performance promissora no reconhecimento de microcategorias apesar da quantidade reduzida de amostras de imagens por classes e validaram os métodos de aquisição da base de imagens.

Para trabalhos futuros, propõe-se aumentar a base de imagens com propósito de aumentar a acurácia e a generalização de contexto do modelo, objetivando uma solução robusta para aplicação comercial.

7. Referências

- DALAL, N.; TRIGGS, B. Histograms of oriented gradients for human detection. In: Proceedings international conference on computer vision and pattern recognition, San Diego, CA, 2005, p. 886–893.
- DENG, J.; DONG, W.; SOCHER, R.; LI, L.-J.; LI, K.; FEI -FEI, L. Imagenet: A large-scale hierarchical image database. In: IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, 2009.
- GEORGE, M.; MIRCIC, D.; SOROS, G.; FLOERKEMEIER, C.; MATTERN, F. Fine-Grained Product Class Recognition for Assisted Shopping. In: The IEEE International Conference on Computer Vision (ICCV), Zurich, Switzerland, 2015, p. 154-162
- WEI & ANGUELOV, DRAGOMIR & ERHAN, DUMITRU & SZEGEDY, CHRISTIAN & REED, SCOTT & FU, CHENG-YANG & C. BERG, ALEXANDER. SSD: Single Shot MultiBox Detector. In: 14th European Conference (ECCV), Amsterdam, NL, 2016.
- MERLER, M.; GALLEGUILLOS, C.; BELONGIE, S. Recognizing groceries in situ using in vitro training data. In: Proceedings of the IEEE conference on computer vision and pattern recognition, Minneapolis, MN. Anais..., 2007, p. 1–8.
- SIMONYAN, K.; ZISSERMAN, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. In: International Conference on Learning Representations, San Diego, CA, 2015.
- VAROL, G.; KUZU, R. S. Toward retail product recognition on grocery shelves. In: Sixth International Conference on Graphic and Image Processing (ICGIP), Beijing, China, 2014.