UNIVERSIDADE FEDERAL DE SANTA CATARINA

Programa de Pós-graduação em Ciência da Computação
Departamento de Informática e Estatística - INE

Contagem de Salgadinhos Sortidos em Caixas Mistas

Vicente Knobel Borges • Luiz Gustavo Coutinho Xavier

Descrição do Problema

O problema escolhido para atividade consiste na identificação de salgadinhos em caixas sortidas, adotando como exemplo encomendas de salgados produzidas por uma padaria local. Apesar de considerar a produção de um comércio específico, o problema estende-se a indústria de alimentos de modo geral, ao permitir a separação de encomendas de forma mais ágil assegurando controle de qualidade, especialmente em períodos de maior demanda de produção.

Caixas sortidas são usualmente encomendadas com 4 ou 5 classes de salgados distintos, e possuem um tamanho padrão para conter um máximo de até 100 salgados.

> **Abordagens testadas:** Pipeline de visão computacional clássica com segmentação por cor e operações morfológicas, e solução com deep learning usando RF-DETR com **mAP@50:95 de 90,51%**.

Visão Computacional Clássica

Para o escopo do problema nesta etapa, considera-se a existência das seguintes classes:

- Croquete de carne
- Coxinha de frango
- Empada folhada
- Sanduíche colorido
- Sanduíche aberto (canapé)
- Canudinho

É importante destacar a similaridade entre classes de salgadinhos no que diz respeito a sua cor e formato. Croquetes de carne, apesar de majoritariamente marrons, possuem pigmentos pastéis da mesma cor das coxinhas. Coxinhas e empadas também possuem grande similaridade de formato e cor.

Dataset

21 imagens capturadas em alta resolução (4032 × 3024) por câmera de celular.

Solução com Visão Clássica

Foi implementado um pipeline simples de processamento de imagem utilizando algoritmos clássicos com o objetivo de identificar uma classe de salgadinhos por vez:

1. **Segmentação por cor:** Máscara binária para valores HSV usando `cv2.inRange`
2. **Fechamento morfológico:** Dilatação seguida de erosão para preencher buracos
3. **Borramento bilateral:** `cv2.bilateralFilter` para suavização
4. **Detecção de bordas:** Algoritmo de Canny
5. **Contagem:** Detecção de elipses nos objetos remanescentes

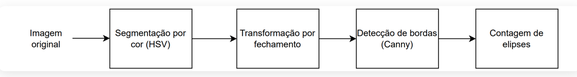
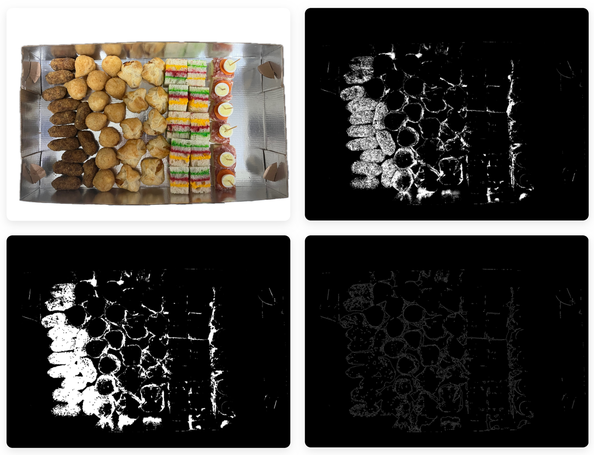


Diagrama do pipeline de visão clássica

Resultados da Abordagem Clássica

Não foi possível isolar os salgadinhos desejados com acurácia satisfatória. Na tentativa de identificar croquetes, resquícios de outros elementos permaneciam após a segmentação. Melhores resultados obtidos com HSV [8, 130, 10] a [24, 180, 140].



Deep Learning

Metodologia

O desenvolvimento seguiu cinco etapas:

1. Busca de imagens no Google
2. Geração de novel view synthesis com Nano Banana Pro
3. Labeling automático com SAM3
4. Refinamento no Label Studio
5. Treinamento com RF-DETR

Geração do Dataset

14 classes estabelecidas: Bolinha de queijo, Canapé, Canudo, Coxinha, Croquete, Empadinha, Enroladinho, Esfiha, Folhado, Pastelzinho, Pão de queijo, Quibe, Risoles e Sanduíche. 30 imagens por classe, priorizando fotos com múltiplos exemplares em ângulo top-down.

Imagens de coxinhas selecionadas do Google

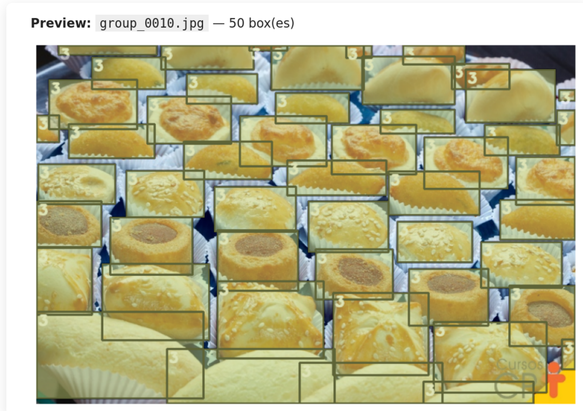
Novel View Synthesis

Modelo Nano Banana Pro da Google para gerar quatro variações: vista lateral, top-down no chão, top-down com iluminação noturna, e top-down na disposição original.

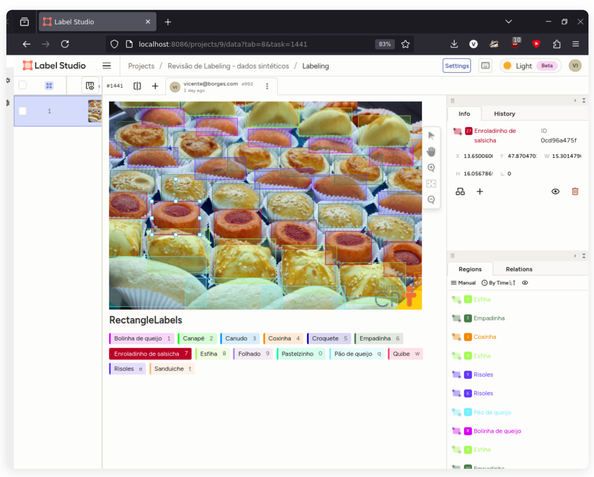
Imagens geradas pelo Nano Banana Pro

Labeling com SAM3

SAM3 para segmentação automática com prompts como "food" ou "snacks". Refinamento no Label Studio corrigindo classificações incorretas.



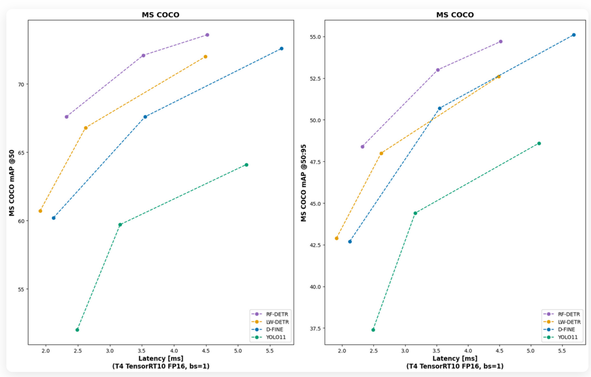
Bounding boxes geradas pelo SAM3



Interface de revisão no Label Studio

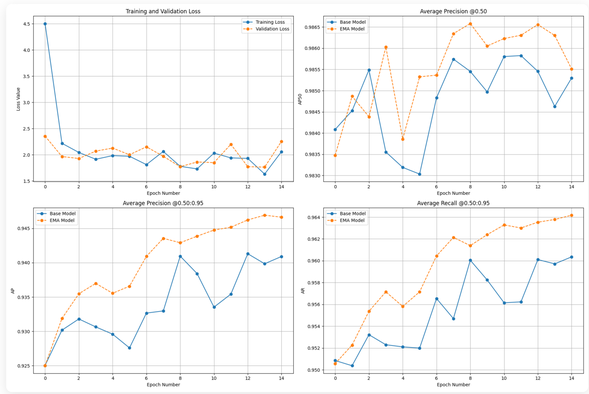
Modelo RF-DETR

Selecionado por ser performante em sistemas com poucos recursos, de fácil implementação em Python, e open-source (Apache-2.0). Escolhido o **RF-DETR Small**.

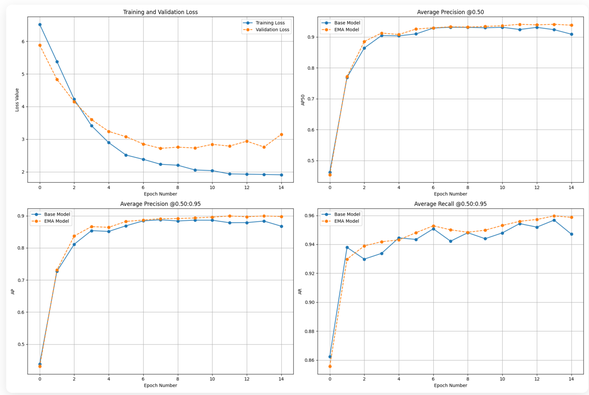


Comparação de métricas mAP no COCO (Fonte: Roboflow)

Resultados de Treinamento



Curvas de treinamento - modelo monolabel



Curvas de treinamento - modelo multilabel

Modelo Monolabel

mAP@50:95 (Teste)	**95,29%**
mAP@50 (Teste)	**98,64%**
Precision	**97,45%**
Recall	**97,00%**

Métricas por Classe - Multilabel (Teste)

Classe	mAP @50:95 (Val)	mAP @50:95 (Test)	Precision (Val)	Precision (Test)
Bolinha de queijo	75,52%	86,90%	54,46%	85,54%
Canapé	94,18%	93,93%	92,86%	95,10%
Canudo	89,91%	95,30%	94,78%	99,35%
Coxinha	74,24%	89,85%	34,05%	90,83%
Croquete	94,39%	90,84%	94,59%	89,07%
Empadinha	93,93%	96,29%	97,93%	100,00%
Enroladinho	91,87%	87,96%	97,80%	98,05%
Esfiha	94,57%	98,85%	99,22%	99,53%
Folhado	97,69%	92,49%	100,00%	91,04%
Pastelzinho	89,68%	91,42%	95,33%	96,03%
Pão de queijo	96,18%	95,09%	98,45%	100,00%
Quibe	91,01%	62,63%	88,11%	51,04%
Risoles	99,41%	98,11%	100,00%	98,11%
Sanduíche	74,31%	87,43%	78,11%	88,66%
Média	**89,78%**	**90,51%**	**87,55%**	**91,60%**

Discussão

O modelo monolabel apresentou convergência rápida e estável, com métricas superiores a 94%. O modelo multilabel mostrou comportamento mais complexo.

> **Alto desempenho:** Risoles (99,41%), Folhado (97,69%), Pão de queijo (96,18%), Esfiha (94,57%)

> **Desafios:** Quibe apresentou degradação significativa no teste (62,63%). As oscilações refletem dificuldade em generalizar entre 14 categorias visualmente similares.