

Luis Antonio Spader Simon (18250065), Vinicius Cim (19250114)

Reconhecimento de produtos para auxílio a deficientes visuais em supermercados

Relatório final do projeto de reconhecimento de produtos para auxílio a deficientes visuais em supermercados EEL7802 - Projeto em Eletrônica II.

Professora: Dra. Daniela Ota Hisayasu Suzuki

Área de Concentração: Visão computacional, eletrônica digital.

1 Sumário

1 Sumário	2
2 Lista de Figuras	3
3 Introdução	4
4 Validação da ideia	5
4.1 Problemas relatados pelos deficientes visuais	5
5 Metodologia de desenvolvimento do projeto	7
5.1 Geração do conjunto de imagens	8
5.1.1 Fotos retiradas manualmente	8
5.1.2 Web Scraping	9
5.2 Data Augmentation	10
5.3 Algoritmo	10
5.3.1 Rede de classificação geral (Por categoria de produto)	10
5.3.2 Rede de classificação de marcas (Café)	11
5.3.3 Rede Yolo para detecção de produtos	11
6 Funcionamento do algoritmo	12
7 Problemas encontrados	13
7.1 Problema 1	13
7.2 Problema 2	13
7.3 Problema 3	14
8 Resultados	14
8.1 Vídeos:	15
9 Projeções futuras	16

2 Lista de Figuras

- [Figure 1. Fluxograma de desenvolvimento do projeto](#)
- [Figure 2. Fotos retiradas manualmente](#)
- [Figure 3. Fotos retiradas em casa](#)
- [Figure 4. Web Scraping de marcas de café](#)
- [Figure 5. Data Augmentation em imagens de produtos de café](#)
- [Figure 6. Fluxograma de funcionamento do algoritmo](#)
- [Figure 7. Resultados obtidos para o modelo de classificação geral.](#)
- [Figure 8. Resultados obtidos para o modelo de classificação entre marcas de café.](#)
- [Figure 9. Resultados obtidos para o modelo de detecção de objetos.](#)
- [Figure 10. Fluxograma do algoritmo a ser desenvolvido futuramente.](#)
- [Figure 11. Exemplo da rede de bounding box de detecção de objetos em prateleira.](#)

Table 1. Lista de Figuras

3 Introdução

Para o desenvolvimento deste projeto, primeiramente foi feita uma pesquisa com pessoas pertencentes ao público-alvo.

Nesta pesquisa, percebeu-se que os entrevistados possuíam uma dificuldade maior na identificação dos produtos na prateleira do que na locomoção pelo supermercado, pois foi relatado que com o hábito é possível decorar os caminhos e seções do mercado. Conforme relatado pelos entrevistados, uma solução completa envolveria um dispositivo que possibilitasse todo o processo de compras em um supermercado: locomoção dentro do ambiente, identificação dos produtos, leitura dos preços e datas de validade.

Porém o foco inicial deste projeto foi uma versão mais simples, com apenas uma das funções citadas acima: identificação dos produtos.

Esta função já traz maior independência dos deficientes visuais neste ambiente interno, pois como já citado, muitos memorizam os corredores e suas categorias de produtos.

Com base nisso, foi proposto neste trabalho a elaboração de um projeto de visão computacional para auxiliar nesta tarefa. A ideia inicial do projeto consiste na implementação de uma rede neural artificial para realizar classificação de produtos típicos de supermercado, onde após a classificação, um dispositivo deve informar qual é o produto de forma sonora.

Não foi possível desenvolver este projeto até o ponto de também informar o preço e validade do produto, mas isto certamente é um dos objetivos para implementação em versões futuras.

4 Validação da ideia

Relatórios anteriores demonstram como ideias existentes, parecem ser ainda muito primitivas nos quesitos: trazer autonomia ao deficiente visual, com praticidade e facilidade no uso. Aplicativos não chegam próximos de trazer tal independência ao usuário, como as implementações futuras citadas (não dependentes da ação do usuário de posicionar algum objeto e/ou câmera próximos um do outro).

A ideia aqui apresentada, funciona independente de qualquer rede de supermercado ou pequenos mercados locais. Com robustez à variações de iluminação e posição do produto em relação à câmera. Independente de investimentos e/ou modificações em supermercados. Resumindo, o usuário poderá levar o dispositivo a qualquer mercado e localizar os produtos desejados sem a necessidade de assistência.

4.1 Problemas relatados pelos deficientes visuais

Conforme apresentado em relatórios anteriores, pesquisas revelaram problemas em comum entre os entrevistados:

1. Os entrevistados relataram que o maior problema em um supermercado é a identificação de produtos, pois os mesmos acabam memorizando os corredores e suas categorias de produtos.
2. Utilizar um smartphone e ter que focar e/ou posicionar a câmera do mesmo para identificar o produto, se torna muito incômodo e demorado.
3. Dois dos entrevistados relataram utilizar de aplicativos para identificação de produtos mas se tornou muito incômodo ter de segurar o produto para identificá-lo e mirar o foco da câmera do smartphone, pois o produto pode não ser o desejado, mesmo após diversas tentativas e é muito mais rápido e cômodo ter a ajuda de um funcionário para fazer tal tarefa.
4. A demora dos aplicativos na identificação e sua sensibilidade à quaisquer mudanças na posição do aparelho se mostrou algo inviável para uso.

De acordo com as revelações anteriores, tornou-se claro a necessidade de desenvolvimento de um sistema que não dependa do usuário estar segurando algum

dispositivo em uma mão e o produto em outra. Esta abordagem anterior, foi o protótipo desenvolvido até então, em tempo hábil para esta disciplina.

Porém, a melhor solução seria um dispositivo automático que já identificasse os produtos desejados diretamente das prateleiras, sendo o mesmo localizado no carrinho de compras, evitando assim qualquer incômodo ao usuário. Esta última versão citada, é considerada pela dupla, para ser implementada nos próximos meses.

5 Metodologia de desenvolvimento do projeto

O fluxograma de desenvolvimento do projeto neste semestre é apresentado na Figura 1. Nele pode-se ver 3 etapas essenciais:

- Geração do conjunto de imagens:
 - Captura de imagens
 - Tratamento das imagens
 - Data Augmentation
- Algoritmo:
 - Treinamento do algoritmo
 - Seleção do modelo de algoritmo
 - Ajuste do algoritmo
- Reconhecimento de produtos (etapa de verificação dos resultados)

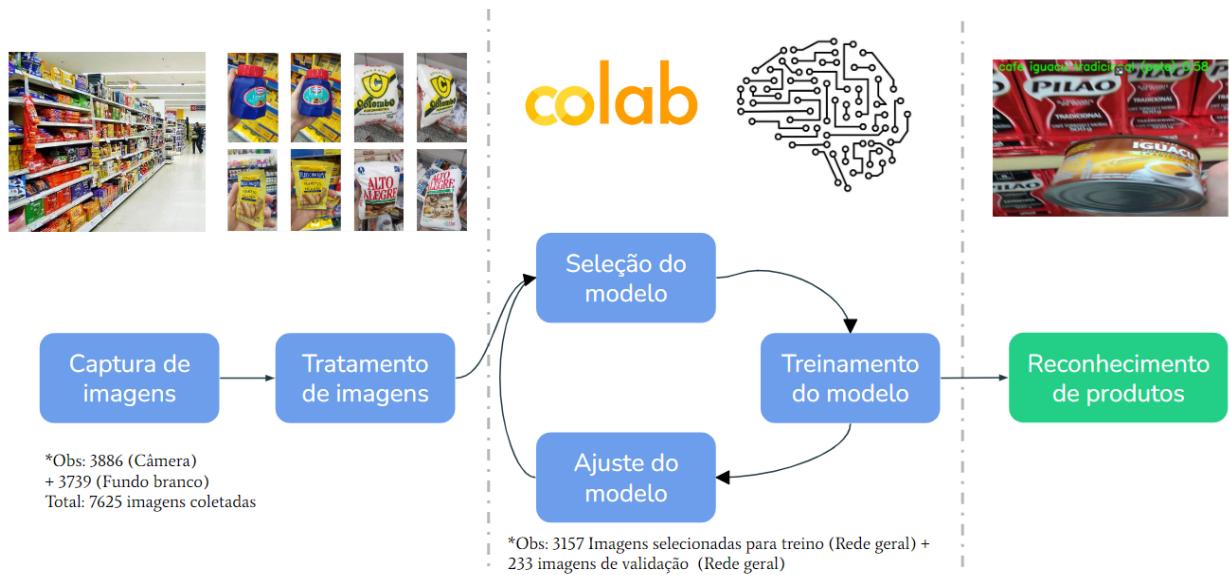


Figure 1. Fluxograma de desenvolvimento do projeto

5.1 Geração do conjunto de imagens

5.1.1 Fotos retiradas manualmente

A dupla se deslocou inúmeras vezes até supermercados ou mercados locais próximos de forma a tirar diversas fotos dos produtos selecionados, variando sempre cada tipo e marca do produto. Ao todo foram adquiridas até o momento, 3739 imagens desta forma. Também foram obtidas imagens de produtos que a equipe dispunha em casa, o que devido ao fundo diferenciado do que geralmente se encontra no mercado (marcas de outros produtos de fundo) contribui para um dataset mais diverso, o que evita o *overfitting* do modelo. Algumas das imagens adquiridas podem ser vistas na Figura 2 abaixo.

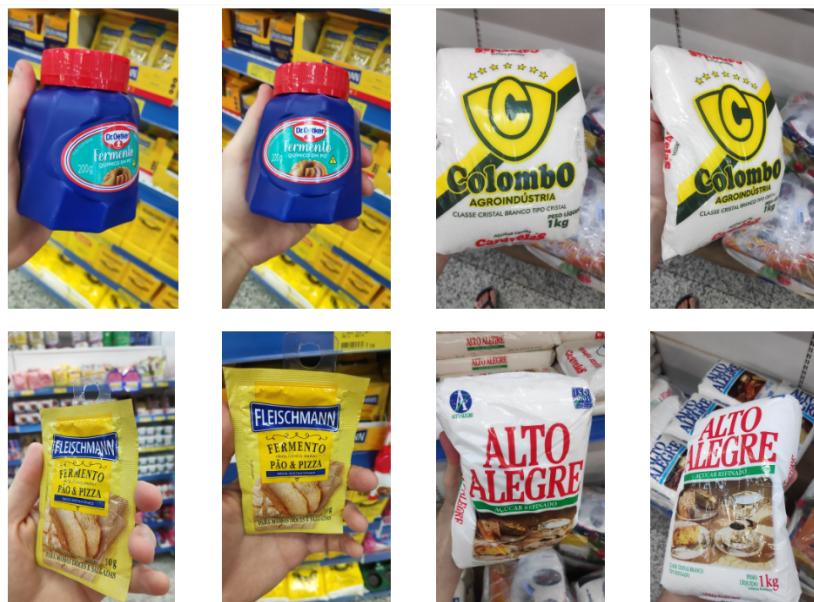


Figure 2. Fotos retiradas manualmente

Para adicionar maior variabilidade às imagens de treinamento do modelo, buscou-se alterar as posições do produto em relação à luz e ao fundo aparente nas fotos. Outra variação foi também a aquisição de imagens nas casas dos integrantes do projeto, conforme pode ser visto na Figura 3.

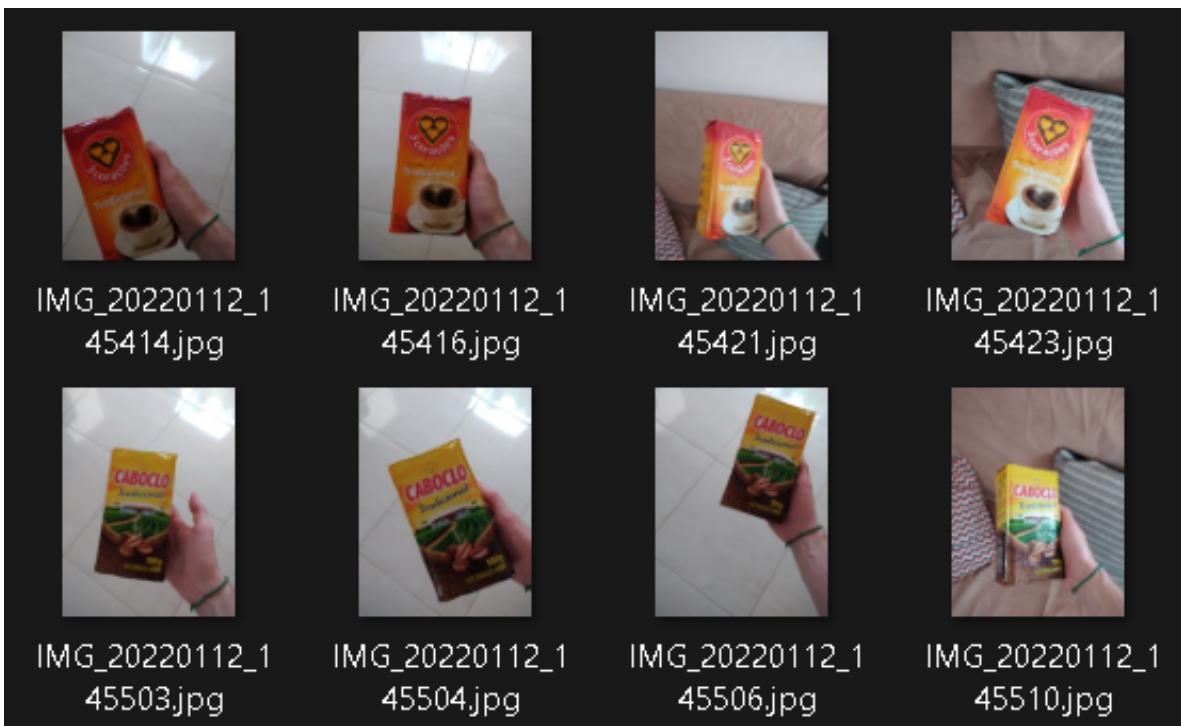


Figure 3. Fotos retiradas em casa

5.1.2 Web Scraping

A outra forma utilizada foi através de sites de supermercado e o próprio pesquisador Google, na categoria de imagens. Para tal foi utilizado um mecanismo conhecido como ‘Web Scraping’, que nada mais é do que utilizar técnicas, códigos de programação e/ou softwares para adquirir dados da rede de forma automatizada. Tal técnica permite uma alta automatização do processo, reduzindo muito o tempo de desenvolvimento do projeto.

Foi desenvolvido um algoritmo que automaticamente realiza o download de imagens de um determinado site, foram utilizados sites dos mercados: Angeloni, Superkock, Bistek, Supermercados Schmit, Pão de açúcar, Carrefour, Delivery Fort (Fort atacadista), Big e Hippo.

Outro site utilizado para o Web Scraping foi o Google Imagens. Para isto foi necessário a implementação de novos códigos de programação que se adequassem ao mesmo.

Na Figura 4, pode-se ver algumas destas imagens, neste caso para diferentes marcas de café moído.



Figure 4. Web Scraping de marcas de café

5.2 Data Augmentation

Técnica utilizada para gerar novas imagens, utilizando imagens já existentes, desta forma pode-se trazer maior variabilidade ao conjunto de imagens e por consequência, maior robustez ao algoritmo. Um exemplo das imagens geradas pode ser visto na Figura 5 abaixo, onde todas as imagens encontram-se rotacionadas e/ou redimensionadas, sendo estas, apenas algumas das modificações possíveis.



Figure 5. Data Augmentation em imagens de produtos de café

5.3 Algoritmo

5.3.1 Rede de classificação geral (Por categoria de produto)

A partir de 3157 imagens para o conjunto de treino foi realizada a operação de *data augmentation* (aumento de dados), conforme já demonstrado no tópico anterior. Para a validação (imagens nunca vistas pelo algoritmo, utilizadas para verificação do desempenho do mesmo), foram utilizadas 233 imagens nunca vistas pela rede, de forma a medir seu desempenho. Assim, utilizou-se de uma rede previamente treinada para a classificação de objetos, a Efficient NET B0, que foi utilizada para a realização de *transfer*

learning uma técnica onde o aprendizado de uma rede é utilizado para se treinar outra rede com outro objetivo, assim obteve-se o resultado visto na Figura 7, uma acurácia de 95,28% no conjunto de validação/testes.

5.3.2 Rede de classificação de marcas (Café)

De forma análoga ao desenvolvimento da rede geral, a rede de classificação de marcas para café utilizou-se da técnica de data augmentation e transfer learning com a Efficient NET B0. Foram utilizadas 1301 imagens correspondentes a 25 marcas de cafés encontradas em supermercados da região da grande Florianópolis e Vale do Itajaí, para a classificação de marcas viu-se a necessidade de um balanceamento no peso das classes, uma vez que o dataset é desbalanceado (Há um número não uniforme de imagens para cada classe, Ex: Café Melitta 100 imagens, Café pilão 10 imagens), este balanceamento foi realizado através da classe Class_weight do Sklearn, que permite que cada classe tenha um peso diferente no treinamento, assim a rede não fica tendenciosa para a classe com maior número de amostras. Dada a abordagem, obteve-se uma acurácia de 92,7% para o conjunto de validação, a acurácia para cada classe pode ser vista em detalhes na figura 8.

5.3.3 Rede Yolo para detecção de produtos

O objetivo do treinamento deste tipo de modelo é alcançar um protótipo que identifique os produtos diretamente das prateleiras do supermercado, sem a necessidade do usuário pegar o produto na mão ou posicionar a câmera com grandes exigências, conforme já citado ao final da seção 3.1.

Esta última rede desenvolvida é responsável por realizar a detecção do produto em foco na câmera, assim melhorando os resultados obtidos para as redes de classificação.

Utilizou-se a rede *YoloV5* com o conjunto de pesos *YoloV5n* que é a versão mais enxuta disponível, isso permite que tanto o treinamento quanto a inferência sejam mais rápidas e será importante na hora de implementar o projeto em um dispositivo móvel (futuramente). Os resultados obtidos podem ser vistos na Figura 9.

6 Funcionamento do algoritmo

Com as redes neurais artificiais treinadas, um algoritmo foi criado de forma a juntar todas as partes até então desenvolvidas.

Acompanhando o fluxograma da Figura 6 abaixo, temos primeiramente a detecção do objeto em foco na câmera, que consiste na detecção de objetos. Desta forma, recorta-se a imagem e se obtém um resultado de classificação mais preciso.

Após esta etapa, a imagem é fornecida para a rede de classificação geral que irá classificar a imagem em uma das 6 classes (Café, Arroz, Macarrão, Refrigerante, Leite e outros).

A qualquer momento pode-se dar um comando para que o algoritmo retorne um áudio informando a identificação do produto atual. Abaixo da rede de classificação geral temos as redes de classificação das marcas, das quais, neste projeto somente a rede de classificação de marcas de café foi implementada.

Esta rede então classifica o café em uma das 26 classes, sendo 25 marcas e uma classe para café não registrado, novamente, a qualquer momento o usuário, através de um comando pode solicitar um retorno audível.

Estas diferentes implementações foram realizadas de forma a provar o conceito e a viabilidade do projeto. Sabendo que, uma vez que a rede para cafés apresentar bons resultados, o processo de expansão para o modelo detectar e diferenciar entre as marcas de outras categorias de produto é natural e apenas repetição da metodologia já utilizada.

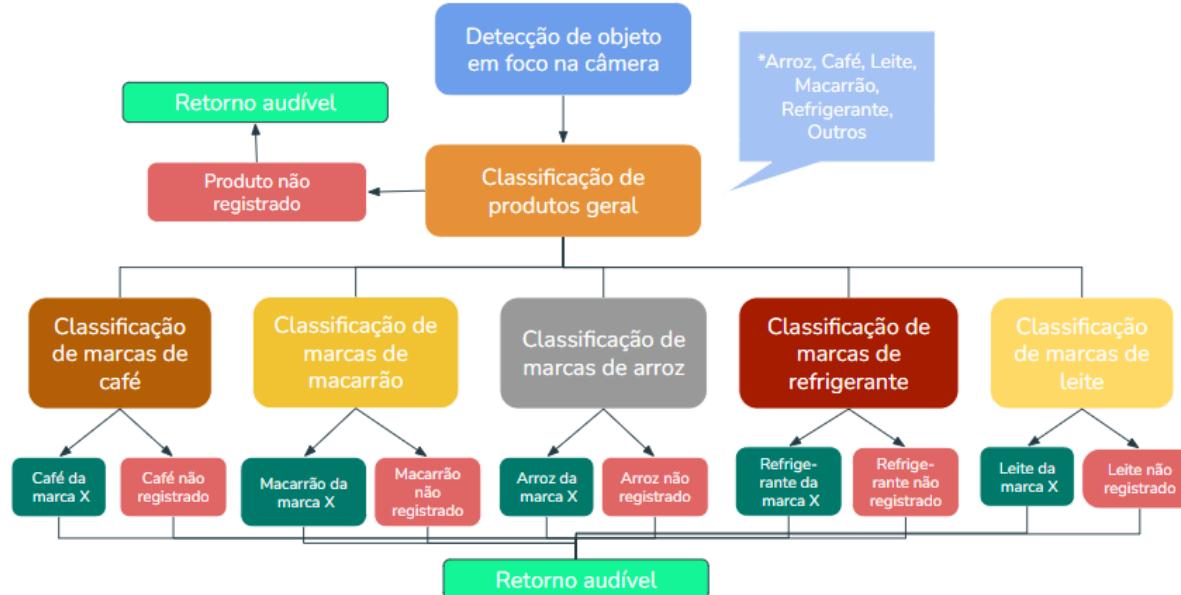


Figure 6. Fluxograma de funcionamento do algoritmo

7 Problemas encontrados

7.1 Problema 1

- O desenvolvimento do trabalho, necessitava de ajustes manuais o tempo todo como: editar as imagens, fotografar novas imagens em novos contextos ou novas posições do produto, retreinar o modelo com o novo dataset, etc.
 - Solução: embora não tenhamos solucionado todo o trabalho manual, pode-se dizer que a dupla automatizou diversos processos como:
 - Pré processamento do dataset, resize, data augmentation, obtenção dos diretórios de cada marca separadamente de forma automática;
 - Obtenção de imagens dos produtos com fundo branco de forma automatizada (web scraping).
 - Organização e automatização de etapas do código de treinamento do modelo (sem necessidade de grandes mudanças a cada nova decisão tomada);
 - Procura de hiperparâmetros do modelo, buscando a melhor configuração possível para a aplicação.
 - Muitas outras etapas internas ao código do modelo que não convém serem especificadas.

7.2 Problema 2

No início do treinamento, o desempenho do modelo no conjunto de validação (imagens nunca vistas) estava bem abaixo do desempenho no conjunto de imagens de treino (imagens já vistas).

- Solução:
 - Foram utilizadas diversas formas de regularização (transforma o modelo para deixar de “decorar” as imagens já vistas, tornando-o mais generalista, o que ajuda em imagens nunca vistas)
 - Foram editadas as imagens do conjunto de treinamento, de forma a retirar parte do fundo da imagem, focando mais no produto (desta forma o modelo aprende melhor que só o que está centralizado e mais próximo é o produto)

7.3 Problema 3

Algumas marcas de café não estavam em diversas lojas de supermercado. Como resultado obtivemos classes desbalanceadas (algumas marcas com muitas fotos e muitas marcas com poucas fotos).

- Solução:

- Foram incluídas no algoritmo de treinamento ‘pesos’ a cada classe que teriam a função de balancear por meio de pesos o dataset, resolvendo o problema já descrito. Em suma, marcas com muitas imagens possuem um peso menor. Marcas com poucas imagens possuem um peso maior.

8 Resultados

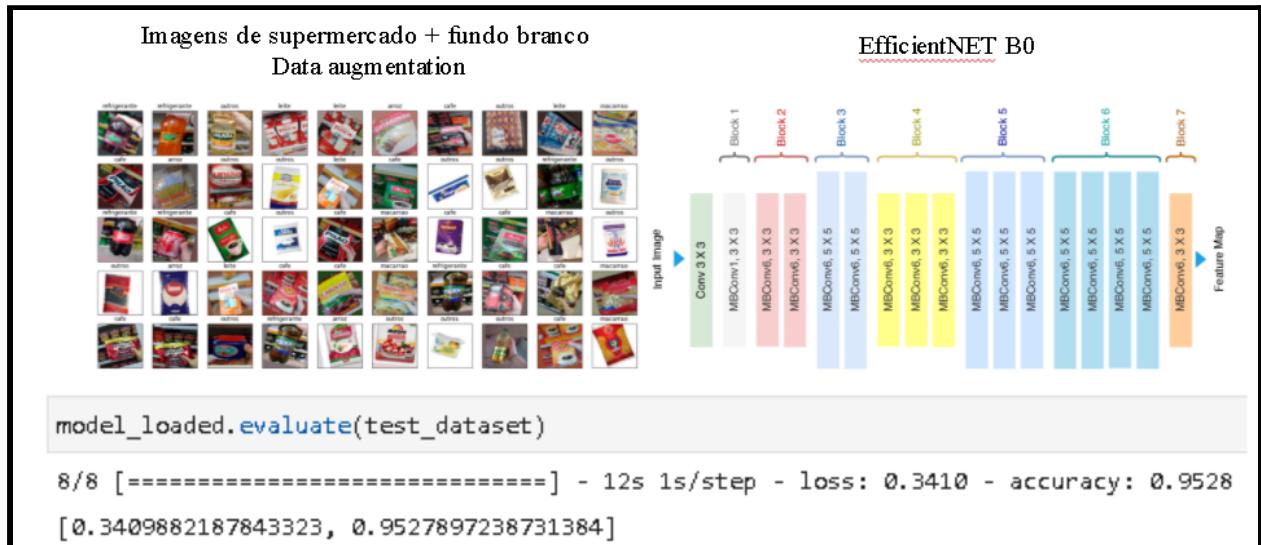


Figure 7. Resultados obtidos para o modelo de classificação geral.

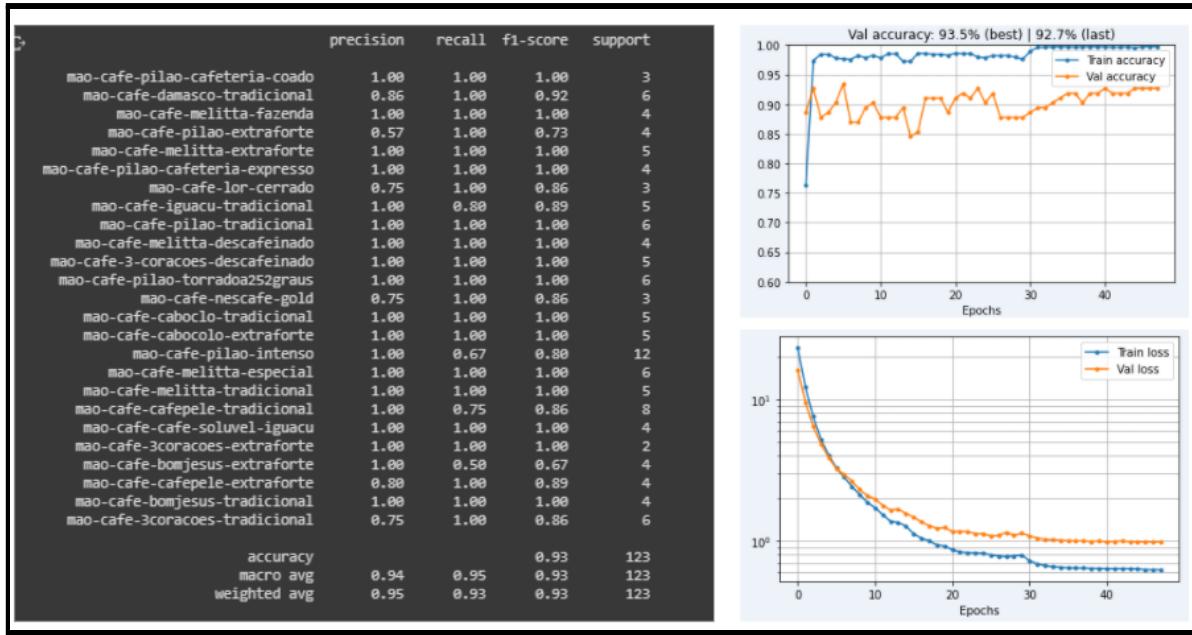


Figure 8. Resultados obtidos para o modelo de classificação entre marcas de café.

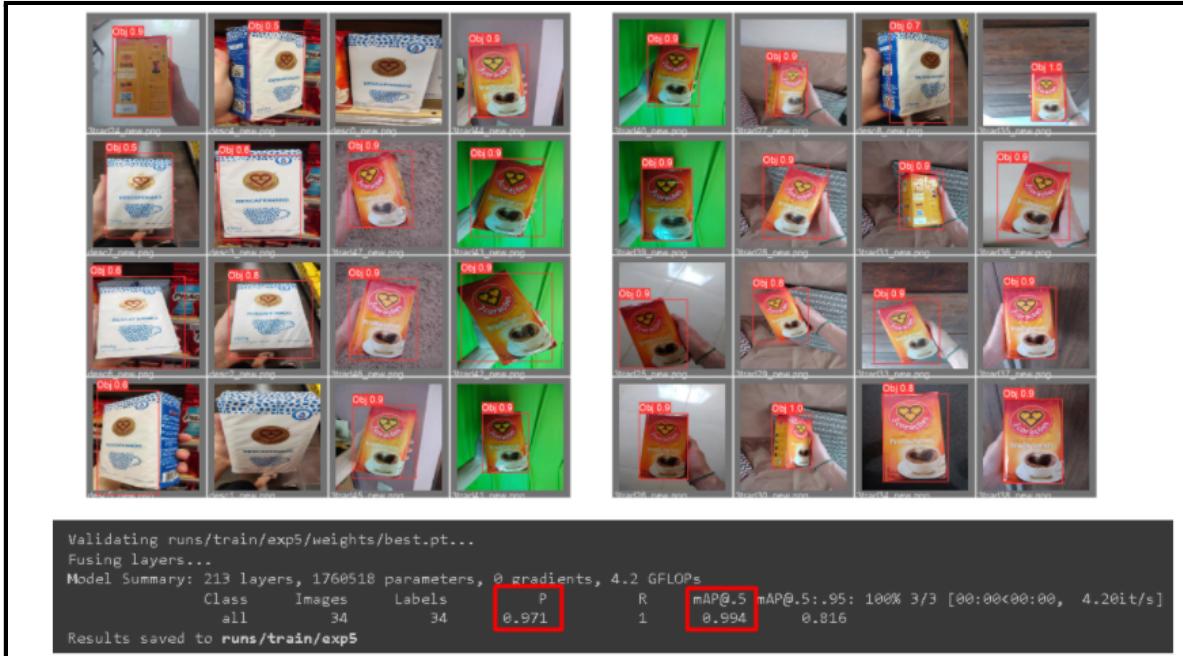


Figure 9. Resultados obtidos para o modelo de detecção de objetos.

8.1 Vídeos:

Classificação em supermercado: <https://www.youtube.com/watch?v=pLzRPXjsz6I>

Retorno audível: <https://www.youtube.com/watch?v=YgWOXNCdCj4>

9 Projeções futuras

Espera-se para um continuação do projeto após disciplina:

- Retorno audível da posição aproximada do produto na prateleira.
- Informação ao modelo, de uma ‘lista de compras’ que conterá os itens que o usuário deseja comprar no supermercado. A mesma será descrita pelo usuário através da fala.
- Retorno audível caso o produto desejado (contido na lista de compras) se encontre em alguma prateleira próxima.
- Retorno audível indicando qual é a seção do mercado (frios, bebidas, massas, enlatados... etc).
- Embarcar o modelo em um dispositivo móvel (smartphone através de um app, ou um sistema embarcado como Raspberry)

Acompanhando as figuras 10 e 11, espera-se desenvolver um algoritmo como especificado na figura 10, onde o usuário pode configurar previamente o aplicativo informando os produtos que deseja comprar no supermercado. Assim, ao ativar o aplicativo no supermercado duas redes irão trabalhar em paralelo, a rede de bounding box (que irá gerar imagens recortadas dos produtos diretamente da prateleira conforme a figura 11, desta forma, o usuário só precisará pegar produtos na mão após indicação sonora de sua localização aproximada) e a rede de identificação da seção do mercado (que irá auxiliar o usuário com a localização dentro do mercado).

Cada uma das imagens geradas pela rede de bounding box é então passada para o algoritmo de classificação desenvolvido até então (onde ocorre a classificação do produto). O resultado é comparado com a lista de produtos desejados pelo cliente.

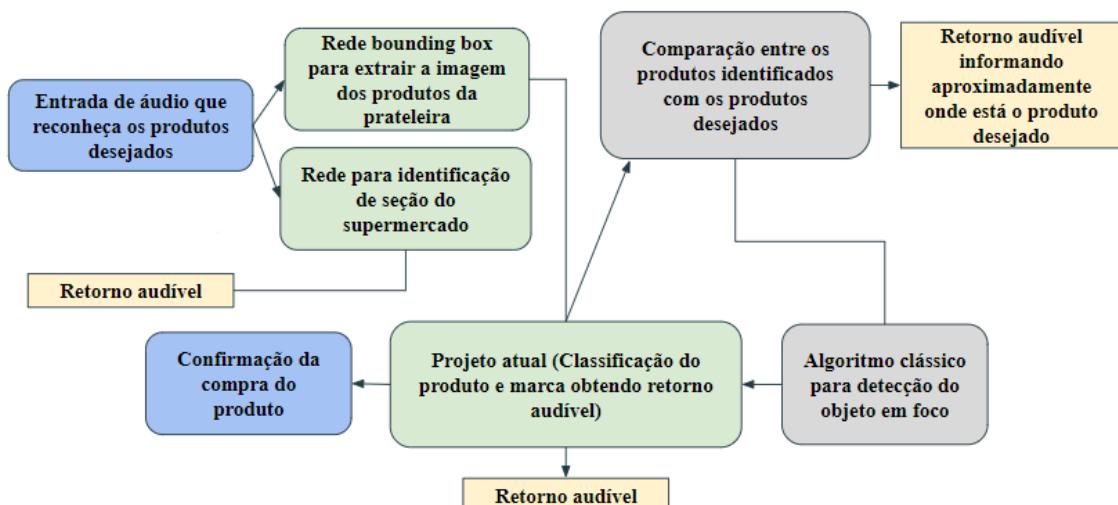


Figure 10. Fluxograma do algoritmo a ser desenvolvido futuramente.

Caso haja uma correspondência entre a lista de compras e o produto classificado, o algoritmo irá dar um retorno audível indicando para o usuário que determinado produto está em uma das prateleiras próximas, ainda informando uma localização aproximada da mesma e do produto.

Desta forma o usuário já sabe de que produto se trata e basta então colocá-lo em frente a câmera e acionar o retorno audível para obter uma confirmação de que o produto obtido pelo cliente é o desejado (mesma categoria e marca).

Ao escolher a marca desejada, o usuário informa ao dispositivo que então exclui o item da lista, o processo então torna a ocorrer assim identificando todos os itens da lista previamente informada. Com isso, espera-se que o usuário consiga identificar os produtos no supermercado de forma autônoma e relativamente rápida. Este processo deve então ser otimizado de forma a torná-lo o mais natural e simples possível para o usuário.



Figure 11. Exemplo da rede de bounding box de detecção de objetos em prateleira.