

# Proposta para o projeto final

Predição do volume do corpo de um objeto através de imagens

Ariosvaldo Zacarias de Barros Filho

## 1 Descrição da tarefa

O objetivo do projeto consiste na predição do volume de balões com água através de imagens, utilizando abordagens de visão computacional como detecção e segmentação e comparando os seus respectivos resultados, para a medição da área do objeto e correlaciona-la com o seu respectivo volume.

Através de imagens e anotações de detecção e segmentação na internet, será realizado a divisão entre dados de treino e validação, e realizado o treinamento de um modelo capaz de realizar a segmentação de uma determinada classe de uma imagem.

Assim que o modelo estiver treinado, será utilizado um conjunto próprio de imagens de balões com água com os dados a respeito dos seus volumes, essas imagens possui também um objeto de referência onde as dimensões são fixas e conhecidas.

Conhecendo a área do objeto de referência e sua razão pela sua quantidade de pixels que foi fornecida através da segmentação, será possível determinar a área do balão com água e com isso gerar uma série de amostras correlacionando a área com o volume e consequentemente seu peso, visto que a densidade da água é  $997kg/m^3$ .

Com isso será aplicado algoritmos de regressão para estimar a curva que representa a variação da área do balão com seu peso.

A utilização de um balão com água se deve ao fato de tal objeto apresentar uma maior facilidade de aquisição e obtenção do volume. Além de que a metodologia é utilizada para qualquer objeto que tenha um conjunto de imagens e seus respectivos dados de volume ou peso.

## 2 Motivação

A coleta do peso de um animal de criação é fundamental para o criador. Através desse dado é possível identificar animais desnutridos ou que estão prontos para a venda.

Com isso o peso do animal possui uma forte correlação com o seu bem-estar, promovendo abordagens que mitiguem o prejuízo e otimizem os lucros. Desta forma é fundamental obter o seu peso de forma rápida e fácil.

Por outro lado, realizar a pesagem do animal pode ser uma tarefa demorada e difícil além de proporcionar uma situação estressante para o animal, considerando que o mesmo seria forçado para o local onde está situada a balança de medição [1].

No cenário onde há uma grande criação, torna-se inviável e custosa a resolução desta tarefa pelos métodos clássicos de pesagem do animal [2].

Um caminho alternativo para a abordagem desse problema é a utilização de imagens para estimar o peso do animal com alta precisão [3].

A utilização de um modelo computacional, aliado a equipamentos simples e baratos, como câmeras de qualquer dispositivo, e um modelo de aprendizado profundo capaz de realizar a detecção ou segmentação de uma classe com base em seus dados de treino, é possível obter o peso com precisão e em tempo real de cada animal [4].

Também é possível estimar o peso e o volume de maçãs para identificar quais estão prontas para serem colhidas e selecionar as que apresentam uma maior qualidade [5].

### 3 Descrição dos dados

Para a realização deste projeto será necessário a obtenção das imagens referentes a classe desejada para a predição do volume, onde neste caso são imagens de balões, além de imagens de alguma classe que será utilizada como objeto comparativo, de modo que suas dimensões sejam conhecidas e fixas, neste caso será utilizado um celular.

Também poderia ser utilizado um fundo quadriculado, onde as dimensões dos quadrados já são conhecidas [5], observar a Figura 1. A utilização do celular se deve ao fato de ser um objeto fácil de possuir, além de que sua base de dados de treinamento é alta.

Além das imagens, é necessário possuir os arquivos de anotações das imagens que contém os dados referente as classes, como as coordenadas dos eixos x e y da caixa delimitadora para detecção, e as coordenadas dos eixos x e y dos pontos que representam os polígonos que contornam o objeto para a segmentação.

A criação dessas anotações em geral são realizadas de forma manual através de ferramentas de código livre, onde é necessário criar uma série de pontos na extremidade do objeto para a formação de um polígono que representa a área do objeto.

Considerando que será treinado um modelo baseado em rede neural, é necessário uma quantidade elevada de dados para se obter um desempenho satisfatório



Figura 1: Detecção e segmentação de maçãs.

da detecção e segmentação, e com isso seria um tarefa árdua a criação de cada uma das anotações.

Desta forma, tanto as imagens quanto as anotações de ambas as classes, foram obtidas através do Open Images Dataset, que é um repositório livre de dados voltados a problemas de visão computacional.

Nas Figuras 2 e 3, é possível observar a distribuição dos dados de treino e validação entre as classes. A quantidade segmentações é maior do que a quantidade de imagens, visto que cada imagem possui pelo menos um objeto segmentado.

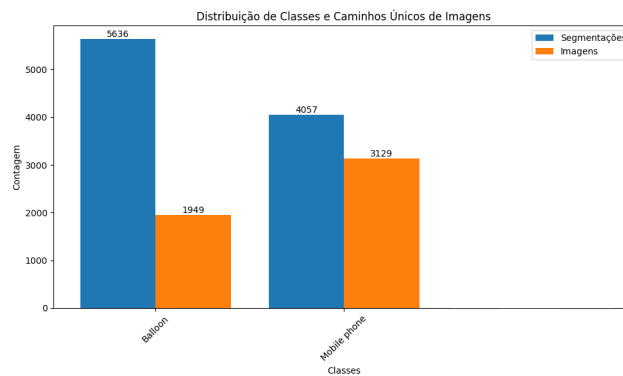


Figura 2: Dados de treino

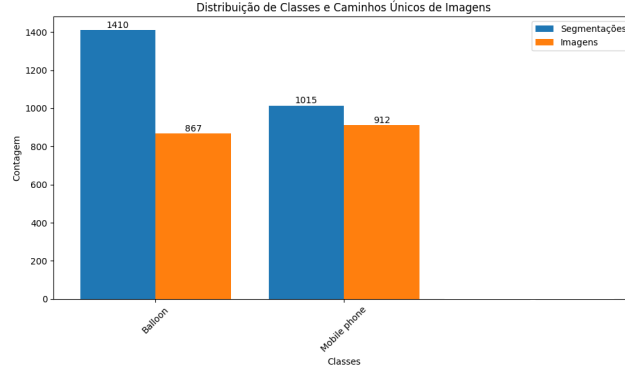


Figura 3: Dados de validação.

## 4 Métricas

É possível utilizar as métricas de classificação como Precision, Recall, F1-Score e o Mean Average Precision (mAP), observar as equações 1, 2, 3 e 4 respectivamente.

$$\text{Precision} = \frac{\text{True Positives (TP)}}{\text{True Positives (TP)} + \text{False Positives (FP)}} \quad (1)$$

$$\text{Recall} = \frac{\text{True Positives (TP)}}{\text{True Positives (TP)} + \text{False Negatives (FN)}} \quad (2)$$

$$\text{F1-Score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (3)$$

$$\text{mAP} = \frac{1}{N} \sum_{i=1}^N AP_i \quad (4)$$

Tais métricas são aplicadas na avaliação do desempenho do modelo em conseguir identificar e classificar corretamente as classes presentes em uma imagem.

Contudo para problemas de detecção, também é interessante observar o Intersection over Union (IoU), observar Figura 4, onde basicamente é analisado a caixa delimitadora (bounding box) real e predita, e com isso é calculado a razão da área de interseção pela área de união, de modo que é possível interpretar esse resultado como o quão exato a caixa delimitadora prevista está da real, onde o valor 1 representa a precisão total, como é possível ver na Figura 5.

Também é possível aplicar essa lógica para a segmentação, onde neste caso será usado a máscara de segmentação prevista e real para calcular a razão dos pixels corretos em comum entre a máscara real e predita em relação aos pixels que estão na união das máscaras de segmentação.

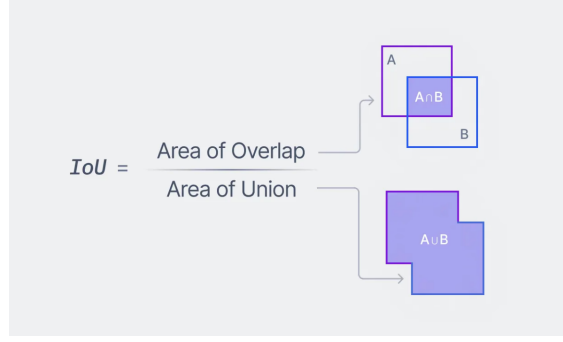


Figura 4: Cálculo do Intersection over Union



Figura 5: Interpretação dos valores de Intersection over Union

## 5 Trabalhos relacionados

Em *Determination of the Live Weight of Farm Animals with Deep Learning and Semantic Segmentation Techniques*, de 2023 [1], utilizou-se o modelo Deeplab v3+ para o treinamento de um conjunto de dados de 85 imagens de gado bovino, mas precisamente vacas leiteiras, onde a margem de erro foi de  $\pm 20kg$ , e foi destacado a importância na utilização de um conjunto mais elevado de dados com maior diversidade.

No projeto *Estimation of Volume and Weight of Apple by Using 2D Contactless Computer Vision Measuring Method*, de 2022 [5], é utilizado o modelo Mask R-CNN para a predição de volumes e pesos de maçãs, utilizando um fundo com tabuleiro quadriculado para a comparação dos pixels e foi obtido um erro médio percentual de 11,97% para o volume e 11,49% para o peso. Foi utilizado 4 tipos de maçãs para 51 imagens de treino e 16 de validação.

Em *Intelligent weight prediction of cows based on semantic segmentation and back propagation neural network*, de 2024 [3], utilizou-se um conjunto de 550 imagens de vista superior para vacas entre 4 e 23 meses de idade, aplicando O ResNet-101-D para a segmentação semântica, e depois a comparação entre diferentes métodos para correlacionar a área da segmentação com o peso do animal, usando Rede Neural BP, SVM, Árvore de Decisão, Regressão Linear Múltipla e Regressão Gaussiana, de modo que a Rede Neural BP apresentou o melhor resultado com um  $R^2$  de 0,99 MAE de 13,11 libras e RMSE de 22,73 libras.

No artigo *Vision analysis and prediction for estimation of pig weight in slaughter pens* [2], de 2023, foi aplicado a técnica de Background updating, atualização da imagem de fundo atual usando uma medida de similaridade de textura (TSM), que foi utilizado em 18 porcos entre 25,4 kg e 87,3 kg durante 10 semanas, onde o sistema alcançou uma precisão média de 97,76% na estimativa do peso médio dos porcos.

## 6 Objetivos

Busca-se nesse projeto a aplicação de técnicas abordadas nos artigos de referência, utilizando outros objetos para a predição de volume, além de explorar o desempenho do sistema através da variação da quantidade de dados de treinamento e épocas de treinamento e observar a diferença na predição dos pesos com a detecção e segmentação.

Tais objetivos são fundamentais, visto que o treinamento de um modelo capaz de detectar ou segmentar um objeto de uma imagem exige uma quantidade elevada de dados e exige tempo de processamento, além do custo computacional por geralmente exigir a utilização de uma GPU robusta.

Também é preciso ressaltar que pelo fato da segmentação exigir uma série de coordenadas para a representação do objeto, ao passo que a detecção exige somente duas coordenadas, que é o ponto superior esquerdo da caixa delimitadora e o ponto inferior direito, isso faz com que modelos de detecção sejam mais rápidos para serem treinados além de exigir um menor custo computacional.

Por isso é importante comparar e apresentar esse conhecimento através de comparações para que se conheça os melhores parâmetros de treinamento e assim evitar desperdício de tempo para se obter muitos dados e treinar o modelo de forma demasiada sem que isso agregue de forma significativa a qualidade da predição do volume do objeto.

## 7 Metodologia

Através da ferramenta "OIDv4\_ToolKit" disponibilizada de forma livre através de um repositório no GitHub, será baixada somente as imagens das classes desejadas do Open Images Dataset, que nesse caso é o balão que é o objeto que deseja-se obter o volume e o celular que será usado para comparar a sua área que já é conhecida com a do balão.

Essas imagens serão utilizadas para obter os seus respectivos IDs, e com isso, através da própria plataforma do Open Images Dataset será baixado todas as imagens junto com as suas respectivas máscaras de segmentação, que nesse caso são imagens de mesma dimensão onde o objeto segmentado está totalmente em branco, ao passo

que o fundo da imagem está todo em preto.



Figura 6: Imagens com suas respectivas máscaras de segmentação.

Com os IDs das classes selecionadas anteriormente, será realizada uma filtragem para apenas as máscaras de segmentação que possuem os IDs correspondentes.

Em seguida, através da biblioteca OpenCV do Python, é possível criar uma lista de coordenadas para cada ponto da borda da imagem segmentada, e desta forma será estruturado um arquivo Json conhecido como *Format Annotation JSON* que é um dicionários com todos os dados da imagem e das classes nela contidas, tanto para treino quanto para validação.

E assim, todo esse conjunto será treinado através da Mask-RCNN (*Mask Regional-Convolutional Neural Network*), que é um modelo de aprendizado profundo projetado para segmentação de instâncias em imagens.

Ele é uma extensão do Faster R-CNN, que realiza detecção de objetos, mas com a capacidade adicional de gerar uma máscara binária para cada objeto detectado.

Como foi discutido na seção anterior, será realizada uma série de treinamentos utilizando tamanhos e épocas variadas e observar o desempenho do modelo através das métricas de avaliação nas imagens de teste abordadas na seção 4 deste documento, no conjunto de teste, onde já se possui os dados a respeito do volume dos balões.

Desta forma, apesar de ser analisado o desempenho do modelo em relação a sua detecção e segmentação, o foco é observar o desempenho final na predição do volume do balão.

## 8 Transparência

O trabalho atual possui relação com um projeto desenvolvido pelo próprio autor deste trabalho no programa Future Alaatus do Fundo Catarina em parceria com a Cargill que foi iniciado em outubro de 2024. O projeto é desenvolvido de

forma individual sob a orientação de Gilson Regadas Filho, Cientista de Dados da Cargill. Tal projeto não envolve nenhum dado sensível da empresa, sendo executado através de dados livres de repositórios da internet, como as imagens do Open Images Dataset e os pesos do modelo provenientes do COCO (Common Objects in Context).

## Referências

- [1] Erdal Guvenoglu. Determination of the live weight of farm animals with deep learning and semantic segmentation techniques. *Applied Sciences*, 13(12):6944, 2023.
- [2] Gang Jun Tu and Erik Jørgensen. Vision analysis and prediction for estimation of pig weight in slaughter pens. *Expert Systems with Applications*, 220:119684, 2023.
- [3] Beibei Xu, Yifan Mao, Wensheng Wang, and Guipeng Chen. Intelligent weight prediction of cows based on semantic segmentation and back propagation neural network. *Frontiers in Artificial Intelligence*, 7:1299169, 2024.
- [4] Chuanqi Xie, Yuji Cang, Xizhong Lou, Hua Xiao, Xing Xu, Xiangjun Li, and Weidong Zhou. A novel approach based on a modified mask r-cnn for the weight prediction of live pigs. *Artificial Intelligence in Agriculture*, 12:19–28, 2024.
- [5] Afiq Ikhwan Mohd Fauzi, Mohd Zamri Bin Ibrahim, and Muhammad Salihin Bin Saealal. Estimation of volume and weight of apple by using 2d contactless computer vision measuring method. In *2022 International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME)*, pages 1–6. IEEE, 2022.