

# Segmentação Automática de Imagens de Carcaças Bovinas

André Henrique F. Costa<sup>1</sup>, Antônio Almedia S. Neto<sup>1</sup>,  
José Augusto M. Nacif<sup>1</sup>, Lucas Bragança da Silva<sup>1</sup>, Ricardo Ferreira<sup>1</sup>

<sup>1</sup>Instituto de Ciências Exatas e Tecnológicas – Universidade Federal de Viçosa (UFV)

{andre.franco, antonio.a.neto, jnacif, lucas.braganca, ricardo}@ufv.br

**Resumo.** *O percentual de gordura exposta de uma carcaça bovina é um dado estatístico para quantificar o quanto de gordura cobre uma carcaça. Esse percentual é extremamente relevante para avaliar a qualidade de uma carcaça e precificá-la. A obtenção dessa estatística é realizada através da análise presencial da carcaça feita por um especialista. Essa análise consiste na relação entre a carcaça como um todo e as regiões de colágeno e músculo, denominadas regiões sem cobertura de gordura. O objetivo deste trabalho é utilizar redes neurais convolucionais para auxiliar na obtenção do percentual de gordura exposta. Para isso, um especialista rotulou um conjunto de imagens que segmenta a carcaça em três classes, uma para delimitar o contorno da carcaça, uma para regiões de colágeno e uma para regiões de músculo. Esse conjunto de dados foi utilizado para treinar uma rede neural convolucional de segmentação de imagens. Os resultados obtidos foram validados através da métrica mean Average Precision (mAP), que avalia a similaridade das caixas delimitadoras (bounding boxes) predita e real. Utilizando essa métrica e considerando o limiar de 0,5, obtivemos um resultado de 42,7% para a segmentação em todas as classes e 98,2% para a segmentação da classe que delimita o contorno da carcaça.*

## 1. Introdução

Redes neurais têm sido amplamente utilizadas em múltiplos contextos e é possível encontrar na literatura diversos estudos que já comprovaram sua eficácia em problemas do mundo real [Coşkun et al. 2017, Grzesiak et al. 2006, Čandek Potokar et al. 2015]. Apesar da vasta gama de arquiteturas para redes neurais, as mais frequentemente utilizadas são as Redes Neurais Profundas (RNP), Redes Neurais Recorrentes (RNR) e Redes Neurais Convolucionais (RNC) [Guo et al. 2018].

Aplicações que utilizam RNCs para lidar com grandes conjuntos de imagens, problemas de visão computacional e processamento de linguagem natural apresentaram resultados com alto grau de precisão [Albawi et al. 2017]. Para os problemas de detecção, classificação e segmentação de imagens, as RNCs reduzem significativamente a carga de técnicas de pré-processamento, alcançando desempenho próximo ao de humanos quando se tem um conjunto de dados de treinamento robusto, com quantidade suficiente de imagens e adequado ao contexto.

A inferência do percentual de gordura exposta de uma carcaça bovina é uma tarefa complexa e ambígua até mesmo para profissionais experientes da área, pois a diferenciação de tecidos gordurosos para outros tecidos bovinos não é uma tarefa trivial. Atualmente, essa informação é obtida de maneira rudimentar, com o profissional analisando a carcaça pessoalmente. O objetivo deste trabalho é implementar um modelo

de aprendizado de máquina para o auxílio na obtenção do percentual de gordura exposta de maneira automatizada, tornando esse processo mais simples e eficiente.

Para alcançar o objetivo, foi criado um conjunto de dados com imagens de carcaças obtidas em um frigorífico local e propôs-se a utilização de uma RNC para a automatização da tarefa. Com a ajuda de um especialista, esse conjunto de dados foi rotulado, explicitando o contorno da carcaça, regiões de músculo exposto e de colágeno exposto. A gordura exposta da carcaça é aquilo que não foi considerado como músculo ou colágeno, ou seja, basta subtrair os pixels de colágeno e músculo da quantidade total de pixels da carcaça para estimar o percentual de gordura.

Este trabalho está organizado da seguinte forma: os trabalhos relacionados são abordados na Seção 2; as principais tarefas que possibilitaram a segmentação automática de imagens de carcaças bovinas e os principais conceitos envolvidos são discutidos na Seção 3; Na Seção 4 são apresentados os resultados obtidos pela rede; Por fim, na Seção 5, as conclusões são apresentadas e o que pode ser feito para melhorar os resultados é discutido.

## **2. Trabalhos Relacionados**

No decorrer desta seção serão apresentados diversos trabalhos da literatura que abordam técnicas e estratégias utilizadas para processar, segmentar e classificar imagens. Também serão discutidos alguns trabalhos que realizam análises em carcaças bovinas, que são o principal objeto de estudo deste trabalho.

Tratando-se de problemas do mundo real, redes neurais são recorrentemente utilizadas em aplicações que os dados são coletados em ambientes que podem interferir na coleta, ou seja, fora de ambientes controlados. Uma rede neural convolucional de oito camadas foi proposta por [Coşkun et al. 2017] para reconhecimento da face de pessoas. A arquitetura personalizada com cinco camadas convolucionais e três camadas de agrupamento apresentou bons resultados na tarefa de reconhecimento facial. Já as redes neurais artificiais são comparadas com modelos de regressão por [Grzesiak et al. 2006] para prever a produção diária de leite de vacas leiteiras. O uso das redes neurais artificiais trouxe um ganho em acurácia e conseqüentemente um erro menor. Também identificou-se um maior coeficiente de correlação entre a predição e os dados originais ao se utilizar redes neurais artificiais.

No intuito de analisar carcaças bovinas, informações como peso e idade do boi na data do abate foram submetidas a uma rede neural artificial de aprendizado supervisionado por [Čandek Potokar et al. 2015] para tentar prever a cobertura de gordura das carcaças. Os resultados obtidos foram interessantes, mas foi constatado um certo viés na rede, por sempre classificar nas três classes de cobertura mais frequentes. No contexto segmentação de carcaças bovinas, redes neurais convolucionais foram utilizadas por [Gonçalves et al. 2020] para destacar a carcaça dos outros elementos presentes na imagem, no intuito de facilitar análises posteriores. Uma arquitetura personalizada foi proposta, combinando a estratégia de segmentação por superpixels e a rede neural.

Quanto à segmentação de carcaças bovinas para classificação de acabamento de gordura, uma estratégia de segmentação em seções de interesse foi utilizada por [De La Iglesia et al. 2020]. Junto a essa estratégia, foi utilizada uma rede de sensores

para coletar dados pertinentes ao contexto da análise proposta. Informações como temperatura, umidade e peso obtidas através dessa rede são utilizadas em conjunto com as saídas de um sistema que classifica cada uma das seções de interesse são fornecidas para um modelo de regressão que realizará a inferência do grau de acabamento de gordura da carcaça.

Diversos estudos já foram realizados para analisar carcaças bovinas, porém, nenhum utilizou a abordagem de segmentar a carcaça bovina para inferir o percentual de gordura. Dito isso, este trabalho foi proposto para cobrir essa lacuna na literatura.

### 3. Metodologia

Ao longo desta seção serão apresentadas as principais tarefas e conhecimentos que possibilitaram a metodologia proposta, sendo eles: (i) criação do conjunto de dados; (ii) escolha da rede neural; (iii) Mask R-CNN; (iv) rotulação dos dados; (v) treinamento da rede neural; e (vi) geração dos resultados.

#### 3.1. Criação do Conjunto de Dados

A coleta das imagens que compõem o conjunto inicial dos dados foi realizada nos dias 26 e 27 de junho de 2019 em um frigorífico local. Ao todo, foram coletadas 2431 imagens, sendo 1311 delas obtidas no primeiro dia e 1120 no segundo. As imagens foram captadas através do sensor Intel RealSense D435 com resolução de 1280x720.

As 2431 fotografias foram analisadas individualmente no intuito de identificar possíveis imperfeições nas imagens, como ocorrências de classes de interesse da carcaça parcial ou totalmente ocultadas por um objeto e/ou pessoa. Uma imagem com uma obstrução parcial é apresentada na Figura 1.



Figura 1. Carcaça com classe de interesse obstruída

Após a análise individual de cada imagem, o conjunto de dados utilizado no trabalho foi criado, com um total de 563 imagens. As outras imagens foram desconsideradas por conter imperfeições que, aos olhos do especialista, influenciariam negativamente no processo de treinamento do modelo. O conjunto de dados criado será utilizado no treinamento do modelo e em sua validação.

### 3.2. Escolha da Rede Neural

O contexto deste trabalho possibilitou a utilização de redes neurais para alcançar o objetivo final, dado que para alcançar esse objetivo é necessário realizar a segmentação de carcaças bovinas por classes de interesse. As redes neurais são modelos matemáticos inspirados no funcionamento dos neurônios de seres inteligentes. Elas são capazes de aprender a partir das experiências que foram submetidas à elas.

No campo das imagens, as RNCs fornecem uma abordagem sofisticada para detectar, classificar e segmentar imagens automaticamente. As RNCs fazem convoluções na matriz de pixels da imagem para extrair padrões, que são convertidos em conhecimento.

Segundo [Albawi et al. 2017], as redes neurais convolucionais têm um excelente desempenho em problemas de aprendizado de máquina, principalmente nos que lidam com imagens. A rede neural utilizada neste estudo foi a Mask R-CNN, proposta por [He et al. 2018]. Ela é uma implementação robusta de uma rede neural convolucional que possibilita identificar múltiplas classes em uma imagem.

### 3.3. Mask R-CNN

A Mask R-CNN consiste de duas principais etapas. A primeira é chamada de *Region Proposal Network*, que delimita com caixas delimitadoras os objetos candidatos a serem reconhecidos pela rede. Essa área presente dentro das caixas delimitadoras é chamada de região de interesse (*Region of Interest – RoI*). Na segunda etapa, é realizada uma extração de características dos candidatos através do RoIPool (*Region of Interest Pooling*). O RoIPool é uma operação de extração de pequenos mapas de características de cada região de interesse.

Após a extração de características, cada objeto é classificado e é realizada uma regressão na caixa delimitadora para verificar se há necessidade de deslocamento. Em paralelo a classificação do objeto e a regressão na caixa delimitadora, a Mask R-CNN também fornece uma máscara binária para cada RoI, o que vai em contraponto aos sistemas recentes que classificam baseado na predição da máscara. O fluxo da Mask R-CNN pode ser conferido na Figura 2.

A rede é composta por múltiplas arquiteturas, porém, para simplificar, é explicitado a diferença entre: (i) arquitetura do *backbone* convolucional, usada para extração de características nas imagens; e (ii) o núcleo da rede, responsável pelos processamentos nas caixas delimitadoras e a predição das máscaras de cada RoI. Na arquitetura do *backbone* convolucional, são utilizadas a ResNet [He et al. 2015] e a ResNeXt [Xie et al. 2017] de 50 e 101 camadas de profundidade. O núcleo da rede é similar ao da *Faster R-CNN* [Ren et al. 2016], porém, com um adendo de um módulo para predição de máscaras através de convoluções na imagem.

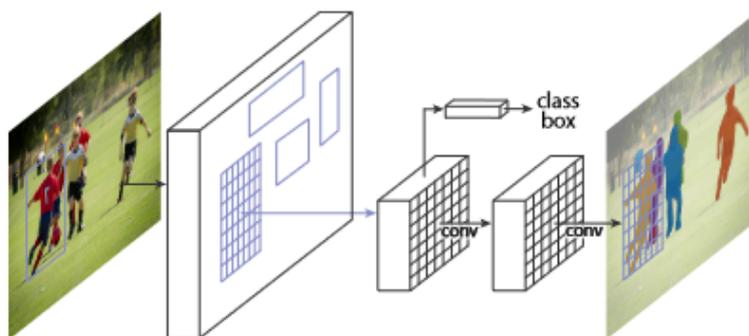


Figura 2. Fluxograma da Mask R-CNN. Adaptado de [He et al. 2018]

### 3.4. Rotulação dos dados

Definiu-se como classes de interesse de uma carcaça bovina as regiões de músculo exposto e de colágeno exposto. Também foi delimitado o contorno da carcaça. Com essa definição e pelo fato de que a rede neural selecionada é de aprendizado supervisionado, foi necessária a rotulação de cada imagem. Para isso, um especialista criou polígonos delimitando a carcaça e suas classes de interesse, com o auxílio da ferramenta LabelMe<sup>1</sup>.

A ferramenta LabelMe gera um arquivo no formato *JSON* para cada imagem, contendo a marcação dos polígonos que representam as classes de interesse. Entretanto, a Mask R-CNN espera como entrada um único arquivo *JSON* com as informações de todas as imagens. Para resolver essa incompatibilidade, foi criado um módulo que recebe como entrada os *JSON*'s de cada imagem, os formata para o padrão de entrada que a rede espera e então os concatena em um único *JSON*.

### 3.5. Treinamento da Rede Neural

Após o processo de rotulação feito pelo especialista, tornou-se possível o início do treinamento da rede neural. A Mask R-CNN originalmente foi implementada para reconhecer até 80 classes distintas com bons percentuais de precisão. A técnica de *Transfer Learning* foi utilizada, para que a rede utilize o conhecimento das 80 classes já treinadas na implementação original para aprimorar os resultados e melhorar o desempenho do treinamento.

A divisão do conjunto de dados em treinamento e validação também é uma parte relevante do processo. Como o conjunto de dados é relativamente pequeno, a divisão utilizada foi 80/20, resultando em 449 imagens para treinamento e 114 imagens para validação.

### 3.6. Geração dos Resultados

A rede foi treinada com as 449 imagens e, posteriormente, as 114 imagens de teste foram submetidas a ela. O fluxo simplificado da Mask R-CNN no modo inferência ocorre da seguinte maneira: (i) a imagem é fornecida para a rede; (ii) a rede realiza diversas convoluções na imagem para identificar regiões que contêm possíveis classes; (iii) as regiões que contêm as classes são delimitadas por uma marcação (caixa delimitadora);

<sup>1</sup><http://labelme.csail.mit.edu/Release3.0/>

(iv) a rede gera as máscaras das classes identificadas; e (v) a rede retorna a imagem original com o acréscimo das marcações das classes identificadas, suas caixas delimitadoras e uma pontuação de cada instância. Na Figura 3, a saída da rede referente à uma imagem com cinco pessoas e três aviões é apresentada.



**Figura 3. Saída da rede com 8 instâncias identificadas. Adaptado de: [https://github.com/matterport/Mask\\_RCNN](https://github.com/matterport/Mask_RCNN)**

Para a geração dos resultados, as marcações das classes, as caixas delimitadoras e as pontuações de cada instância são utilizadas. Esses dados juntamente com os resultados da rede são armazenados, para que posteriormente, seja possível fazer a validação do modelo treinado. As estratégias de validação serão apresentadas e discutidas na Seção 4.1.

## 4. Resultados

Como já mencionado na Seção 3.6, a Mask R-CNN possui um modo inferência, que foi utilizado para a geração dos resultados. O processamento foi realizado em uma máquina com processador Intel(R) Core(TM) i7-7700 CPU @ 3.60GHz, unidade de processamento gráfico GeForce GTX 1070 8Gb e memória principal RAM DDR4 16Gb. O ambiente de desenvolvimento utilizado foi o jupyter <sup>2</sup> e as implementações foram feitas na linguagem de programação Python com auxílio das bibliotecas TensorFlow e Keras. Para verificar o desempenho do modelo treinado, 114 imagens foram submetidas à rede no intuito de validá-lo. As imagens de saída contém a máscara e a pontuação de cada instância encontrada.

### 4.1. Validação dos Resultados

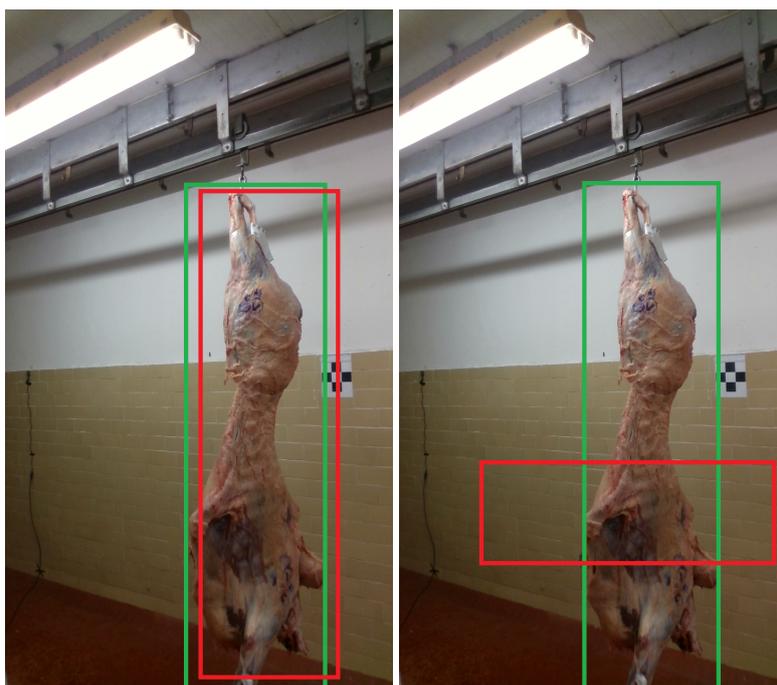
Redes neurais convolucionais exigem métricas de validação dos resultados mais sofisticadas e robustas. Sendo assim, neste estudo os resultados foram validados através da métrica mAP (*mean Average Precision*).

Para cada instância de uma classe identificada em uma imagem, a caixa delimitadora predita é sobreposta à caixa delimitadora real. Dessa sobreposição é calculado o IoU

<sup>2</sup><https://jupyter.org/>

(*Intersection over Union*), o qual é obtido através da divisão da área da interseção pela área da união das caixas delimitadoras. O IoU pode variar de 0 a 1, com 0 significando caixas delimitadoras totalmente diferentes e 1 caixas delimitadoras idênticas.

Através do IoU de cada instância é calculado o AP (*Average Precision*) da imagem. Para isso, é necessário definir um limiar e compará-lo com o IoU. Caso o IoU for maior que o limiar, temos um verdadeiro positivo, caso contrário, temos um falso positivo. Um exemplo ilustrativo é apresentado na Figura 4, considerando o limiar 0.5.



(a) Verdadeiro positivo (IoU > 0.5)    (b) Falso positivo (IoU < 0.5)

**Figura 4. Caixas delimitadora real (verde) e predita (vermelho)**

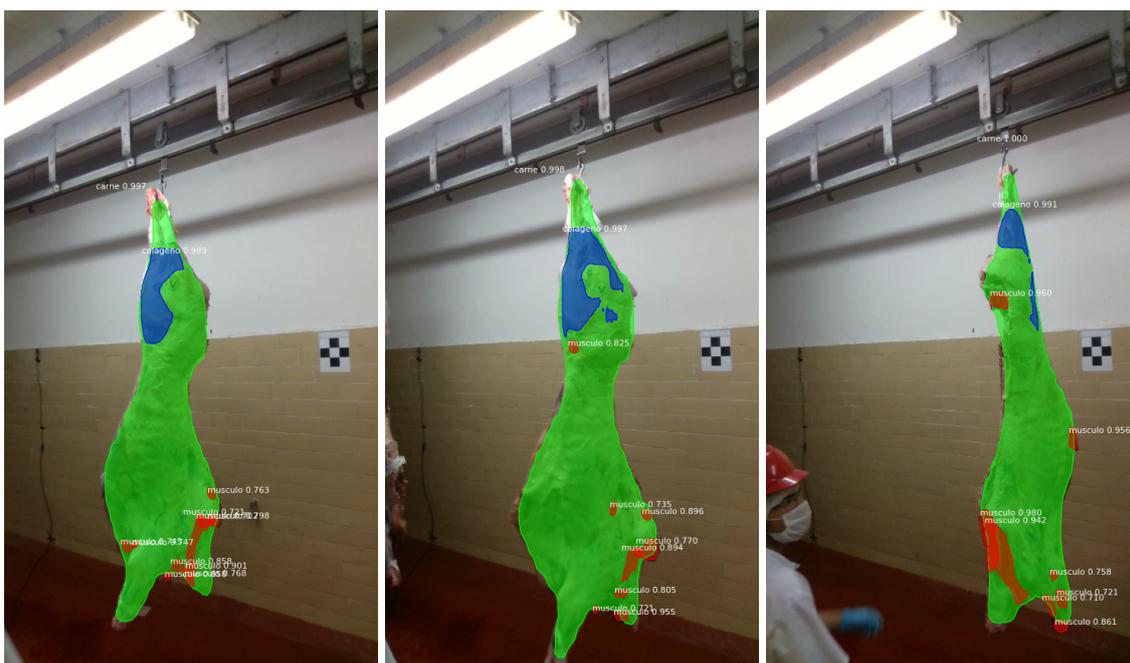
O AP da imagem é a quantidade de verdadeiros positivos dividido pela quantidade de instâncias encontradas na imagem (verdadeiros positivos + falsos positivos). Já o mAP consiste na média aritmética do AP de todas as imagens do conjunto de validação.

## 4.2. Experimentos

As 114 imagens foram submetidas uma por vez ao modelo no modo inferência e a imagem de saída, assim como os dados referentes a imagem (geometria das caixas delimitadoras, pontuações das classificações) foram salvos para validação. Alguns exemplos da segmentação realizada pela rede podem ser conferidos na Figura 5.

Por padronização, as imagens resultantes seguem o mesmo esquema de coloração das classes identificadas, vermelho representa a classe músculo, azul representa a classe colágeno e verde representa a classe carcaça (a identificação da carcaça como um todo). A gordura da carcaça é obtida subtraindo a classe colágeno e a classe músculo da carcaça como um todo.

Como mencionado na Seção 4.1, validar os resultados obtidos por uma rede neural convolucional requer métricas específicas que levam englobam fatores que as métricas



**Figura 5. Amostra dos resultados obtidos**

normais desconsideraram. Para a validação, a métrica mAP foi utilizada, com limiares de 0.25, 0.5 e 0.75. Isso significa que, para o objeto identificado ser considerado um verdadeiro positivo, a caixa delimitadora predita tem que ter 25%, 50% e 75% respectivamente de similaridade com a caixa delimitadora real.

Os resultados da rede serão avaliados de duas maneiras, uma levando em consideração as três classes de interesse e a outra avaliará apenas a classe carcaça, que representa o contorno da carcaça. Os resultados obtidos são apresentados na Tabela 1.

**Tabela 1. mAP obtido na segmentação de todas as classes e da classe carcaça individualmente**

	Três classes de interesse	Classe 'carcaça'
<b>AP25</b>	66.4%	98.2%
<b>AP50</b>	42.7%	98.2%
<b>AP75</b>	11.3%	98.2%

Como pode ser observado na Tabela 1, para a segmentação das três classes de interesse, maiores limiares do AP geraram resultados com o mAP menor. Essa relação faz sentido, pois, o limiar do AP é necessário para definir se o objeto encontrado é um verdadeiro positivo ou um falso positivo. Quando o limiar do AP é mais baixo, a caixa delimitadora predita e a caixa delimitadora real precisam ser menos similares para que o objeto seja um verdadeiro positivo. Ao aumentar o limiar do AP, essa similaridade terá que ser maior para que o objeto seja considerado um verdadeiro positivo, o que explica melhores resultados no AP25 e piores no AP75.

Analisando o mAP quando foi levado em consideração todas as classes, os resultados foram satisfatórios dada a complexidade do problema e a similaridade entre os tecidos expostos bovinos. Já quando olhamos para a classe carcaça individualmente, o

resultado em todas as métricas foi o mesmo, 98.2%. A justificativa para a persistência desse resultado vem do fato de que em 2 das 114 imagens a carcaça não foi identificada corretamente devido a ruídos na imagem.

## 5. Conclusões e Trabalhos Futuros

Neste trabalho, uma extensão da rede neural convolucional Mask R-CNN foi implementada no intuito de obter o percentual de gordura exposta da carcaça através da segmentação automática de carcaças bovinas em regiões de interesse. Para isso, são identificadas as regiões de colágeno e músculo da carcaça e essas regiões são subtraídas da carcaça como um todo, resultando na gordura da carcaça.

Visto que redes neurais convolucionais requerem técnicas de validação mais sofisticadas e robustas, a métrica mAP foi escolhida para validação dos resultados. O resultado obtido através dessa métrica utilizando o limiar 0.5 foi de um mAP de 42.7%, um resultado bem sólido e satisfatório, dado que o problema é complexo até mesmo para profissionais com experiência na área.

Em trabalhos futuros, considera-se a aplicação de técnicas de pré-processamento nas imagens para remover possíveis ruídos e padronizar as imagens. Aumentar o conjunto de dados, seja através da obtenção de novas imagens ou por técnicas computacionais como o data augmentation para melhorar o treinamento é uma possível continuação deste estudo. A avaliação do desempenho do modelo em outros conjuntos de imagens de carcaças bovinas também é um possível estudo futuro. Além disso, a utilização de outras redes neurais convolucionais com características diferentes podem ser avaliadas nesse problema.

## Referências

- Albawi, S., Mohammed, T. A., and Al-Zawi, S. (2017). Understanding of a convolutional neural network. In *2017 International Conference on Engineering and Technology (ICET)*, pages 1–6.
- Coşkun, M., Uçar, A., Yildirim, O., and Demir, Y. (2017). Face recognition based on convolutional neural network. In *2017 International Conference on Modern Electrical and Energy Systems (MEES)*, pages 376–379.
- De La Iglesia, D. H., González, G. V., García, M. V., Rivero, A. J. L., and De Paz, J. F. (2020). Non-invasive automatic beef carcass classification based on sensor network and image analysis. *Future Generation Computer Systems*, 113:318–328.
- Gonçalves, D. N., de Moares Weber, V. A., Pistori, J. G. B., da Costa Gomes, R., de Araujo, A. V., Pereira, M. F., Gonçalves, W. N., and Pistori, H. (2020). Carcass image segmentation using cnn-based methods. *Information Processing in Agriculture*.
- Grzesiak, W., Błaszczuk, P., and Lacroix, R. (2006). Methods of predicting milk yield in dairy cows—predictive capabilities of wood’s lactation curve and artificial neural networks (anns). *Computers and Electronics in Agriculture*, 54(2):69–83.
- Guo, M.-F., Zeng, X.-D., Chen, D.-Y., and Yang, N.-C. (2018). Deep-learning-based earth fault detection using continuous wavelet transform and convolutional neural network in resonant grounding distribution systems. *IEEE Sensors Journal*, 18(3):1291–1300.

- He, K., Gkioxari, G., Dollár, P., and Girshick, R. (2018). Mask r-cnn.
- He, K., Zhang, X., Ren, S., and Sun, J. (2015). Deep residual learning for image recognition.
- Ren, S., He, K., Girshick, R., and Sun, J. (2016). Faster r-cnn: Towards real-time object detection with region proposal networks.
- Xie, S., Girshick, R., Dollár, P., Tu, Z., and He, K. (2017). Aggregated residual transformations for deep neural networks.
- Čandek Potokar, M., Prevolnik, M., Škrlep, M., i Furnols, M. F., and Novič, M. (2015). An attempt to predict conformation and fatness in bulls by means of artificial neural networks using weight, age and breed composition information. *Italian Journal of Animal Science*, 14(1):3198.