

## 1. Introdução

O manejo adequado da saúde animal é essencial para a produtividade e sustentabilidade na pecuária moderna. A detecção precoce de doenças em bovinos, em particular, desempenha um papel crucial na melhoria do bem-estar animal e na eficiência operacional. Tradicionalmente, o monitoramento da saúde dos animais é realizado por meio de medições manuais, um método comum, mas que apresenta limitações significativas quando aplicado a grandes rebanhos.

A termografia vem sendo investigado como uma alternativa a formas de medição mais invasivas, demonstrando correlações entre as medições e oferecendo vantagens significativas em termos de redução do estresse animal durante o monitoramento [1]. A capacidade de detectar variações na temperatura superficial dos animais, relacionadas à identificação de condições de saúde por meio de alterações fisiológicas, pode ser crucial para a detecção precoce de doenças.

Paralelamente, a visão computacional tem se destacado no gerenciamento de gado, permitindo medições precisas de parâmetros corporais e a identificação de padrões anormais que podem indicar problemas de saúde [2]. A integração dessas tecnologias não só aprimora a precisão no monitoramento da saúde animal, mas também automatiza processos que tradicionalmente exigem intervenção humana, contribuindo para a redução do estresse dos animais e aumentando a eficiência dessas operações.

A escolha dos olhos como a principal região para medição da temperatura em bovinos se dá pela sua forte correlação com a temperatura interna do animal. Diferente de outras partes do corpo, os olhos apresentam uma resposta térmica mais rápida e precisa às variações fisiológicas internas, o que os torna ideais para detecção de febres ou outros sinais de infecção [12]. Além disso, outras regiões, como o couro ou o lombo do boi, podem estar expostas a diferentes condições ambientais, como luz solar direta ou em regiões cobertas, o que compromete a precisão da leitura e não reflete adequadamente a temperatura interna do animal.

Neste contexto, foi realizada uma pesquisa detalhada sobre o desenvolvimento e a aplicação de um sistema de visão computacional para a leitura automatizada da temperatura de bovinos, com o objetivo de detectar precocemente sinais de doenças. A abordagem proposta utiliza vídeos de câmeras térmicas que passam por algoritmos de processamento de imagens e modelos de detecção e segmentação dos olhos do gado para a medição precisa da temperatura corporal dos ani-

mais. Além disso, são discutidos os desafios técnicos e os impactos no âmbito da pecuária.

## 2. Metodologia

A metodologia adotada envolveu as seguintes etapas: aquisição de imagens, seleção de frames representativos, anotação e criação de máscaras para o treinamento de modelos de detecção.

### 2.1 Aquisição das Imagens

Os vídeos foram coletados em um confinamento bovino por profissionais treinados, usando uma câmera infravermelha. As gravações ocorreram a uma distância de aproximadamente 1 metro do curral, em diferentes momentos ao longo de vários dias.

### 2.2 Seleção de Frames Diversos

Após a aquisição, foi aplicado um algoritmo de seleção espaçada de frames garantindo a diversidade dos frames escolhidos, minimizando a necessidade de anotar todas as imagens individualmente.

### 2.3 Anotação e Criação de Máscaras

As imagens selecionadas foram carregadas no CVAT (Computer Vision Annotation Tool) para anotação das cabeças e olhos dos bovinos [3]. O uso do CVAT foi fundamentado por sua eficácia em tarefas de rotulagem de imagens em aplicações de visão computacional [4]. Com as anotações exportadas em formato XML, foi desenvolvido um script para recortar as imagens em 128x128 pixels e gerar máscaras correspondentes para as regiões de interesse.



Figure 1: Frame com as labels.

## 3. Pré-processamento

As imagens foram divididas em conjuntos de treino e teste. No conjunto de treino, foram aplicadas técnicas de data augmentation e filtros de contraste,

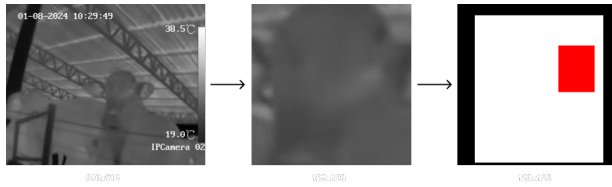


Figure 2: Processo de criação das máscaras.



Figure 3: Overlay da cabeça do boi com a máscara.

incluindo CLAHE (Contrast Limited Adaptive Histogram Equalization), para melhorar a qualidade e variabilidade das imagens. O método CLAHE foi escolhido por sua capacidade de melhorar o contraste em imagens sem amplificar o ruído, especialmente em regiões homogêneas [5].

#### 4. Desenvolvimento do Modelo de Segmentação

Nesta etapa, foram desenvolvidos e comparados quatro modelos de segmentação baseados em arquiteturas de redes neurais profundas: VGG16, ResNet [6], VGG16 + UNet [7], e ResNet + UNet. Esses modelos foram escolhidos por suas capacidades comprovadas em tarefas de segmentação de imagens e por permitirem a combinação entre uma rede preditiva (VGG16 ou ResNet) e uma rede de segmentação (UNet) [8].

**4.1 Divisão e Configuração dos Dados** As imagens pré-processadas foram divididas em três conjuntos: 70% para treinamento, 15% para validação e 15% para teste. Essa divisão foi realizada para garantir que o modelo fosse avaliado em dados não vistos durante o treinamento, permitindo uma validação eficaz do desempenho. O pré-processamento incluiu normalização dos valores dos pixels, e as imagens foram redimensionadas para 128x128 pixels para garantir compatibilidade com as arquiteturas escolhidas.

**4.2 Arquiteturas Utilizadas** Foram configuradas quatro arquiteturas distintas para a segmentação dos olhos dos bovinos:

**VGG16:** Utilizada como base de uma rede convolucional, com camadas fully connected substituí-

das por camadas convolucionais para segmentação. **ResNet:** Implementada para explorar sua capacidade de aprendizagem profunda, utilizando skip connections para mitigar o problema de gradientes desaparecendo. **VGG16 + UNet[6]:** Combinação da VGG16 como extrator de características, seguido por uma arquitetura UNet para segmentação precisa. **ResNet + UNet:** Combinação da ResNet como extrator de características, seguido pela UNet para melhorar a segmentação.

**4.3 Otimização de Hiperparâmetros** Foi realizada uma busca de hiperparâmetros utilizando Grid Search para cada arquitetura, buscando otimizar os seguintes parâmetros:

**Taxa de aprendizado:** Avaliando diferentes valores para garantir uma convergência adequada. **Número de camadas convolucionais e filtros:** Ajustando o número de camadas e a profundidade das redes. **Regularização:** Implementando técnicas como dropout e L2 regularization para evitar overfitting. **Técnicas de augmentação de dados:** Aplicadas especificamente ao conjunto de treino para aumentar a variabilidade dos dados.

**4.4 Treinamento e Validação** Os modelos foram treinados usando os mesmos dados e pelo mesmo número de épocas, com early stopping baseado na métrica de validação para evitar overfitting. O treinamento foi realizado em um ambiente controlado, utilizando GPUs para acelerar o processo. Durante o treinamento, foram monitoradas as seguintes métricas:

**Intersection over Union (IoU):** Para medir a sobreposição entre as máscaras preditas e as máscaras reais. **Dice Coefficient:** Para avaliar a similaridade entre as predições e as máscaras verdadeiras. **Boundary F1-Score:** Para analisar a precisão na segmentação dos contornos dos olhos dos bovinos.

**4.5 Validação Cruzada** Para garantir a robustez dos modelos, foi utilizada validação cruzada com k-folds. Essa técnica permitiu avaliar o desempenho dos modelos em diferentes subconjuntos dos dados, aumentando a generalização dos resultados e reduzindo o viés.

#### 5. Detecção de Objetos com YOLO

Após o processo de aquisição, seleção e anotação das imagens, o próximo passo foi implementar um modelo de detecção de objetos utilizando o YOLO para identificar as cabeças e os olhos dos bois nas imagens térmicas. O uso de YOLO é uma escolha adequada[14], pois

sua arquitetura de detecção em estágio único permite identificar múltiplos objetos em uma única passada pela rede neural, mantendo um equilíbrio entre velocidade e precisão.[13] Essa abordagem foi aplicada em duas fases: a primeira focada na detecção das cabeças e a segunda na detecção dos olhos, restrita às áreas previamente identificadas como cabeças.

Para preparar os dados de entrada do modelo, as coordenadas das regiões anotadas foram extraídas e transformadas no formato YOLO, que requer a normalização das coordenadas das caixas delimitadoras (bounding boxes) em relação às dimensões da imagem. As cabeças foram rotuladas como classe 0 e os olhos como classe 1. Abaixo está um exemplo do código utilizado para gerar esses arquivos de rótulo:

```

...
with open(label_file, 'w') as f:
    for (x, y, w, h) in head_bboxes:
        x_center = (x + w / 2) / width
        y_center = (y + h / 2) / height
        w_norm = w / width
        h_norm = h / height
        f.write(f"0 {x_center} {y_center} {w_norm} {h_norm}\\n")

    for (x, y, w, h) in eyes_bboxes:
        x_center = (x + w / 2) / width
        y_center = (y + h / 2) / height
        w_norm = w / width
        h_norm = h / height
        f.write(f"1 {x_center} {y_center} {w_norm} {h_norm}\\n")
...

```

**5.1 Treinamento Faseado com YOLOv5 e YOLOv8** Na primeira fase, o YOLOv5 foi utilizado para detectar as cabeças dos bois. Esta fase foi fundamental para garantir que, durante a detecção dos olhos, o modelo focasse apenas nas regiões dentro das cabeças, evitando falsos positivos em outras áreas da imagem. O treinamento com YOLOv5 utilizou as imagens rotuladas com as coordenadas das cabeças, e a performance foi medida utilizando métricas como Interseção sobre União (IoU), coeficiente Dice[15], e Boundary F1-Score, amplamente utilizadas para avaliar a precisão de modelos de detecção de objetos.

Após a detecção bem-sucedida das cabeças, a segunda fase foi implementada utilizando o YOLOv8. Nesta fase, o objetivo foi refinar a detecção dos olhos dentro das áreas delimitadas pelas cabeças já identificadas. Essa abordagem é particularmente eficaz para cenários em que a precisão em detectar objetos pequenos, como os olhos, depende de um pré-processamento robusto. O YOLOv8 foi escolhido nesta fase por apresentar melhorias em relação às versões anteriores, especialmente

para a detecção de objetos pequenos em imagens de baixa qualidade, como é o caso de imagens térmicas.

Este processo é corroborado por estudos que aplicam YOLO em imagens em escala de cinza, onde o contraste e a precisão são críticos. [9] demonstraram a eficácia do YOLO em imagens médicas de baixo contraste, [10] aplicaram técnicas semelhantes para melhorar a detecção em ambientes com baixa visibilidade, como em neblina. Esses estudos mostram que a implementação de YOLO em imagens térmicas segue uma linha bem fundamentada de aprimoramento da detecção de objetos em condições adversas.

Além disso, trabalhos recentes sobre o YOLOv8 destacam seu desempenho superior em imagens de sensoriamento remoto, que apresentam desafios semelhantes às imagens térmicas em termos de contraste e definição[11]. O modelo YOLOv8 foi treinado para refinar a detecção dos olhos, e as métricas de avaliação indicaram um desempenho robusto, com alta precisão e baixa taxa de falsos positivos.

Em suma, a abordagem faseada com YOLOv5 e YOLOv8, combinada com a estruturação adequada dos dados de entrada, permitiu uma detecção precisa tanto das cabeças quanto dos olhos dos bois. Este processo representa uma aplicação prática e eficaz de técnicas modernas de detecção de objetos em imagens térmicas, alinhada com o estado da arte da pesquisa em visão computacional e aprendizado profundo.



Figure 4: Imagem object detection olho e cabeças.

## 6. Análise e Comparação dos Resultados

Após o treinamento, os resultados foram analisados e comparados em termos das métricas mencionadas anteriormente (IoU, Dice Coefficient, Boundary F1-Score). Além disso, foram coletados dados sobre:

**Número de Parâmetros:** Para avaliar a complexidade e os recursos computacionais necessários para cada modelo. **Tempo de Treinamento:** Para medir a eficiência de cada arquitetura. **Uso de Memória:** Para analisar a viabilidade de implementar os modelos em sistemas com recursos limitados. Os resultados foram compilados e comparados para identificar qual arquitetura oferece o melhor desempenho com o menor número de parâmetros.

## 7. Resultados

**Apresentação de Resultados:** O modelo **YOLOv8n** foi treinado para a detecção de cabeças e olhos de bovinos ao longo de **50 épocas** utilizando um conjunto de dados customizado. A arquitetura escolhida foi a **YOLOv8n**, uma versão leve da rede YOLO, com um *batch size* de 16 e uma resolução de imagem de 640x640. O treinamento foi realizado em uma GPU **NVIDIA A100-SXM4-40GB**, com **Automatic Mixed Precision (AMP)** ativada para otimização do tempo de treinamento e redução de memória.

Durante o treinamento, foi observada uma melhora contínua nas métricas de desempenho do modelo. As perdas de caixa delimitadora (**box loss**), classificação (**cls loss**) e regressão distribuída (**dfl loss**) apresentaram uma queda significativa, indicando que o modelo aprendeu a detectar com precisão os objetos no conjunto de dados. Ao final das 50 épocas, os resultados demonstraram que o modelo foi ajustado corretamente, atingindo um bom equilíbrio entre precisão e recall.

Nos testes de detecção em vídeo, o modelo identificou **102 cabeças** e **89 olhos** de bovinos em diversos frames. A inferência foi realizada com um tempo médio de **8.4ms a 9.6ms por frame**, o que permite seu uso em aplicações em tempo real. Durante esses testes, foram também reconhecidas variações de temperatura corporal dos animais, com valores de **39.7°C** (alta) e **19.8°C** (baixa).

**Gráficos e Tabelas:** A **Tabela 1** apresenta os resultados do modelo ao longo das épocas, destacando as perdas e as métricas de avaliação (precisão, recall e mAP):

Época	Box Loss	Class Loss	Dfl Loss	Precisão (P)	Recall (R)	mAP@50	mAP@50-95
1/50	2.039	2.959	1.707	0.876	0.274	0.358	0.178
10/50	1.525	1.131	1.274	0.567	0.644	0.607	0.312
20/50	1.328	0.822	1.188	0.726	0.719	0.741	0.368
30/50	1.183	0.753	1.120	0.759	0.793	0.788	0.394
40/50	1.040	0.637	1.050	0.751	0.735	0.752	0.385
50/50	0.8057	0.4571	0.9443	0.725	0.784	0.758	0.396

Figure 5: Resultados ao longo das épocas

Os resultados da época **50/50** indicam que o modelo atingiu uma **precisão (P)** de **0.725**, **recall (R)** de **0.784**, com **mAP@50** de **0.758** e **mAP@50-95** de **0.396**, demonstrando uma evolução consistente até a última época. A perda de caixa delimitadora (**box loss**) foi de **0.8057**, enquanto a perda de classificação (**cls loss**) e a perda de regressão distribuída (**dfl loss**) foram de **0.4571** e **0.9443**, respectivamente.

**Desempenho em vídeo:** O modelo foi testado em vídeo, obtendo os seguintes resultados: - **Total de cabeças detectadas:** 102 - **Total de olhos detectados:** 89 - **Tempo médio de inferência por frame:** 8.4ms a 9.6ms - **Temperatura mais alta registrada:** 39.7°C - **Temperatura mais baixa registrada:** 19.8°C

Esses resultados indicam uma boa capacidade de detecção em tempo real, com molduras precisas ao redor das cabeças e olhos dos bovinos, além de um desempenho estável mesmo em condições de variação térmica.

**Clareza:** Os resultados evidenciam uma evolução significativa nas métricas de desempenho ao longo do treinamento. A partir da **época 20**, a precisão e o recall estabilizaram em torno de **0.7**, atingindo uma precisão máxima de **0.759** na **época 30**. O **mAP@50-95**, que avalia a precisão em diferentes limiares de IoU, alcançou **0.396** na **época 50**, demonstrando um bom nível de detalhamento na detecção.

A estabilidade das métricas nas últimas 20 épocas, com variação mínima nas perdas e ganhos de desempenho, sugere que o modelo foi capaz de aprender as principais características do conjunto de dados sem superajuste (**overfitting**). O desempenho em vídeo, aliado ao tempo de inferência reduzido, reforça a aplicabilidade do modelo em cenários práticos de monitoramento de gado.

## 8. Análise e Discussão

O desempenho observado no YOLOv8n demonstra que o modelo atingiu os objetivos propostos, alcançando um equilíbrio adequado entre precisão e velocidade, o que o torna viável para o monitoramento em tempo

real da saúde de bovinos. Ao longo das 50 épocas de treinamento, o Box Loss e o Class Loss diminuíram progressivamente, indicando que o modelo foi aprimorado na detecção de regiões de interesse, como olhos e cabeças dos animais. O Box Loss reduziu de 2.039 para 0.8057, enquanto o Class Loss caiu de 2.959 para 0.4571, refletindo uma melhoria substancial na qualidade das predições

A evolução das métricas de precisão e recall também merece destaque. Inicialmente, a precisão era elevada (0.876), sugerindo que o modelo fazia previsões corretas com alta confiança, mas com um recall baixo (0.274), o que indica que muitos olhos e cabeças não estavam sendo detectados. No decorrer das épocas, o recall aumentou consistentemente, alcançando 0.784 na última época, sugerindo que o modelo passou a capturar mais detecções verdadeiras. A melhoria contínua no mAP@50 (de 0.358 para 0.758) e no mAP@50-95 (de 0.178 para 0.396) reflete a capacidade do modelo de aumentar a precisão em predições mais desafiadoras e complexas.

Esses achados estão alinhados com pesquisas prévias que demonstram a eficácia da aplicação de sistemas automatizados de detecção na pecuária de precisão, especialmente no uso de tecnologias de termografia infravermelha e visão computacional [1]. A combinação dessas tecnologias para o monitoramento contínuo da saúde animal oferece uma abordagem não invasiva, possibilitando a detecção precoce de condições anormais por meio da leitura da temperatura na região dos olhos, identificando variações corporais associadas a febres e outras doenças.

A relação entre os resultados e os objetivos iniciais do projeto é evidente. O modelo alcançou um desempenho robusto, com alta capacidade de detecção e agilidade, o que é crucial para monitorar grandes rebanhos de forma eficiente. As implicações para a pecuária de precisão são significativas, pois a automação desses processos pode melhorar a gestão da saúde animal, reduzir custos e prevenir surtos de doenças que impactam diretamente a produtividade[16].

No entanto, algumas limitações devem ser consideradas. O modelo foi testado em condições controladas, o que pode não refletir seu desempenho em cenários de campo, onde variações de luminosidade e condições climáticas adversas podem influenciar a qualidade das imagens capturadas e a acurácia das detecções [17]. Para aumentar a robustez do sistema, são necessários testes com vídeos em ambientes diversos e sob condições variadas. A adaptação a esses fatores pode envolver a normalização das imagens, o uso de técnicas de data augmentation para aumentar

a diversidade do conjunto de dados e a ampliação do conjunto de imagens, permitindo a implementação e refinamento de outras técnicas de aprendizado profundo, o que pode melhorar significativamente o desempenho do modelo.



## Referências

- Hoffman, A.A., Long, N.S., Carroll, J.A., Burdick Sanchez, N.C., Broadway, P.R., Richeson, J.T., Jackson, T.C., Hales, K.E. Infrared thermography as an alternative technique for measuring body temperature in cattle. *Applied Animal Science*, 2023.
- Dórea, J.R., Bresolin, T., Ferreira, R.E.P., Pereira, L.G.R. Harnessing the Power of Computer Vision System to Improve Management Decisions in Livestock Operations. *Journal of Animal Science*, 2020.
- “CVAT Overview.” CVAT. Disponível em: [https://docs.cvat.ai/docs/getting\\_started/overview/](https://docs.cvat.ai/docs/getting_started/overview/). Acesso em: 15 de agosto de 2024.
- Sager, C., Janiesch, C., & Zschech, P. A survey of image labelling for computer vision applications. *Journal of Business Analytics*, 4(2), 91–110. <https://doi.org/10.1080/2573234X.2021.1908861>
- ŞENGÖZ, Nilgün et al. Importance of preprocessing in histopathology image classification using deep convolutional neural network. *Advances in Artificial Intelligence Research*, v. 2, n. 1, p. 1-6, 2022.
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 770-778). <https://doi.org/10.1109/CVPR.2016.90>
- Ghosh, S., Chaki, A., & Santosh, K. (2021). Improved U-Net architecture with VGG-16 for brain tumor segmentation. *Physical and Engineering Sciences in Medicine*, 44. <https://doi.org/10.1007/s13246-021-01019-w>
- Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015* (Vol. 9351, pp. 234-241). Springer, Cham. [https://doi.org/10.1007/978-3-319-24574-4\\_28](https://doi.org/10.1007/978-3-319-24574-4_28)
- Sheng, W., & Zhu, Y. “Object Detection Using YOLO in Grayscale Medical Images.” *International Journal of Machine Learning and Cybernetics*, 2023.
- Liang, Z., Li, L., & Chen, J. “AODs-CLYOLO: An Object Detection Method Integrating Fog Removal and Detection in Haze Environments.” *Applied Sciences*, vol. 14, n. 16, 2024.
- Swathi, Y., & Challa, M. “YOLOv8: Advancements in Object Detection for Grayscale Remote Sensing.” *Multimedia Tools and Applications*, 2024.
- Jaddoa, M. A., Gonzalez, L., Cuthbertson, H., & Al-Jumaily, A. “Multiview Eye Localisation to Measure Cattle Body Temperature Based on Automated Thermal Image Processing and Computer Vision.” *Infrared Physics & Technology*, vol. 119, 2021, p. 103932. <https://doi.org/10.1016/j.infrared.2021.103932>.
- Diwan, T., Anirudh, G., & Tembhurne, J. V. “Object Detection Using YOLO: Challenges, Architectural Successors, Datasets and Applications.” *Multimedia Tools and Applications*, vol. 82, 2023, pp. 23897–23930. <https://doi.org/10.1007/s11042-022-13644-y>.
- Guo, S.-S., Lee, K.-H., Chang, L., Tseng, C.-D., Sie, S.-J., Lin, G.-Z., Chen, J.-Y., Yeh, Y.-H., Huang, Y.-J., & Lee, T.-F. “Development of an Automated Body Temperature Detection Platform for Face Recognition in Cattle with YOLO V3-Tiny Deep Learning and Infrared Thermal Imaging.” *Applied Sciences*, vol. 12, n. 8, 2022, p. 4036. <https://doi.org/10.3390/app12084036>.
- Bello, R. W., Ikeremo, E. S., Otobo, F. N., Olubummo, D. A., & Enuma, O. C. “Cattle Segmentation and Contour Detection Based on SOLO for Precision Livestock Husbandry.” *Journal of Applied Sciences and Environmental Management*, vol. 26, no. 10, 2022, pp. 1713-1720. <https://dx.doi.org/10.4314/jasem.v26i10.15>.
- Jaddoa, M. A., González, L., Cuthbertson, H., & Al-Jumaily, A. “Multi View Face Detection in Cattle Using Infrared Thermography.” *Lecture Notes in Computer Science*, vol. 11951, 2019, pp. 190–200. [https://doi.org/10.1007/978-3-030-38752-5\\_18](https://doi.org/10.1007/978-3-030-38752-5_18).
- Church, J., Hegadoren, P. R., Paetkau, M., Miller, C., Regev-Shoshani, G., Schaefer, A., & Schwartzkopf-Genswein, K. “Influence of Environmental Factors on Infrared Eye Temperature Measurements in Cattle.” *Research in Veterinary Science*, vol. 96, no. 1, 2014, pp. 220–226. <https://doi.org/10.1016/j.rvsc.2013.11.006>.