



III MCSM

Modelagem Computacional no Sertão Mineiro



PPGMCS
Programa de Pós-Graduação em
Modelagem Computacional e Sistemas

ARMAZENAMENTO, PRÉ-PROCESSAMENTO E ANÁLISE DE MÍDIAS DE DRONES COM ALGORITMOS DE INTELIGÊNCIA ARTIFICIAL PARA FINS AGROPECUÁRIOS

Thiago Vinícius Silva Rezende¹ - thiagor3zende@gmail.com

Heveraldo Rodrigues de Oliveira¹ - heveraldo.oliveira@unimontes.br

Marcos Flávio S. V. D'Angelo¹ - marcos.dangelo@unimontes.br

¹Universidade Estadual de Montes Claros, UNIMONTES - Montes Claros, MG, Brazil

Resumo. O crescente uso de Veículos Aéreos Não Tripulados (VANTs) tem gerado um grande volume de dados visuais no campo, exigindo soluções computacionais eficientes para organizar e transformar essas mídias em informações úteis, o que desperta a necessidade de armazenar, pré-processar e possibilitar a aplicação de algoritmos de Inteligência Artificial sobre esses dados. Um dos gargalos centrais reside no processamento contínuo de vídeos e no risco de superestimar rebanhos e plantações pela recontagem de espécimes ao longo dos quadros. A arquitetura proposta neste trabalho estabelece um fluxo automatizado abrangendo o recebimento das mídias, a extração de quadros de vídeos, a normalização das imagens e o arquivamento estruturado. Essa abordagem combina amostragem temporal de frames, normalização dimensional, persistência em nuvem e a integração do algoritmo YOLOv8 com o rastreador ByteTrack para associação temporal via identificadores únicos (track_id). Os testes estruturais da Prova de Conceito demonstraram a viabilidade metodológica do fluxo computacional proposto, oferecendo uma base escalável para mitigar custos de infraestrutura e processamento em cenários agrícolas.

Palavras-chave: Drones, Visão Computacional, Processamento de Imagens, Plataforma Web.

1. INTRODUÇÃO

O avanço da Zootecnia e da Agricultura de Precisão tem impulsionado a adoção de Veículos Aéreos Não Tripulados (VANTs), transformando o monitoramento de rebanhos e plantações por meio do sensoriamento remoto aéreo (Barbedo, 2019; Shao et al., 2019). Essa abordagem possibilita ao produtor gerenciar seus empreendimentos de maneira mais eficiente, utilizando informações confiáveis e sistemas especializados, o que contribui para a tomada de decisões mais assertivas de acordo com as necessidades do contexto produtivo (Pandorfí et al., 2012).

Entretanto, o grande volume de dados visuais não estruturados gerado durante os voos cria um desafio computacional expressivo. A análise isolada de cada *frame* de vídeo resulta em

alta redundância temporal e elevado consumo de recursos de memória e processamento (Zhu et al., 2017). Além disso, abordagens baseadas exclusivamente em detecção estática apresentam o risco severo de contabilizações múltiplas, onde o mesmo animal ou espécime vegetal é recontado repetidamente ao longo de quadros sucessivos, exigindo a aplicação de algoritmos de rastreamento para associação temporal (Ciaparrone et al., 2019).

Para superar essas limitações, a Visão Computacional e a Inteligência Artificial (IA), associadas a técnicas eficientes de amostragem de dados, apresentam-se como recursos essenciais para viabilizar a extração automatizada de características (Barbedo, 2019). A integração de modelos de detecção de objetos de última geração, como o YOLOv8 (Jocher et al., 2023), com algoritmos de associação temporal, como o ByteTrack (Zhang et al., 2022), surge como uma alternativa promissora para garantir a identificação e contagem única de indivíduos na varredura do vídeo, eliminando duplicações na contagem final.

Diante desse cenário, este trabalho apresenta a proposição de uma arquitetura computacional desacoplada e a modelagem do pipeline de dados para a ingestão, pré-processamento, rastreamento e contagem automatizada de bovinos a partir de mídias aéreas. A proposta visa automatizar a preparação técnica das imagens e demonstrar, por meio de uma Prova de Conceito (PoC), a viabilidade da aplicação do modelo em dados do setor agropecuário.

2. ARQUITETURA E PRÉ-PROCESSAMENTO DE DADOS

A arquitetura computacional proposta foi estruturada sob o paradigma de desacoplamento de responsabilidades, visando garantir escalabilidade e modularidade ao sistema. A camada de interface (*frontend*) foi desenvolvida como uma *Single Page Application* (SPA) utilizando a biblioteca React, responsável pela interação com o usuário. Já a camada de serviço (*backend*) foi implementada em linguagem Python com o microframework Flask, sendo encarregada de gerenciar as regras de negócio, a autenticação via *JSON Web Token* (JWT) e a persistência de metadados em um banco de dados relacional PostgreSQL.

A etapa de pré-processamento constitui o núcleo técnico que viabiliza a aplicação dos modelos de Inteligência Artificial. Ao receber as mídias aéreas, o *backend* aciona algoritmos computacionais baseados na biblioteca OpenCV para padronização dos dados. Em mídias em vídeo, a extração integral de quadros geraria um volume excessivo e redundante de informações (Zhu et al., 2017), portanto, o pipeline executa a amostragem automatizada de quadros em intervalos regulares de 0,5 segundo. Em seguida, tanto as imagens estáticas quanto os quadros extraídos são submetidos ao redimensionamento dimensional padronizado (para a resolução de 256×256 píxeis) e à normalização dos valores de píxeis, assegurando a consistência exigida pela camada de predição.

Para mitigar limitações de infraestrutura local decorrentes do elevado volume de dados processados, o fluxo computacional integra-se de forma assíncrona com a API da plataforma Kaggle. Após a conclusão do tratamento inicial, os conjuntos de dados (*datasets*) já normalizados são remetidos automaticamente para o repositório em nuvem, liberando recursos do servidor local e garantindo acessibilidade para futuras etapas de treinamento e inferência. A Figura 1 ilustra o fluxo de funcionamento da solução. De modo geral, o fluxo do sistema inicia-se com o envio de imagens ou vídeos pelo usuário, passa pela preparação dos dados, pelo arquivamento do dataset processado e, posteriormente, pela análise por IA e exibição dos resultados na interface da aplicação.

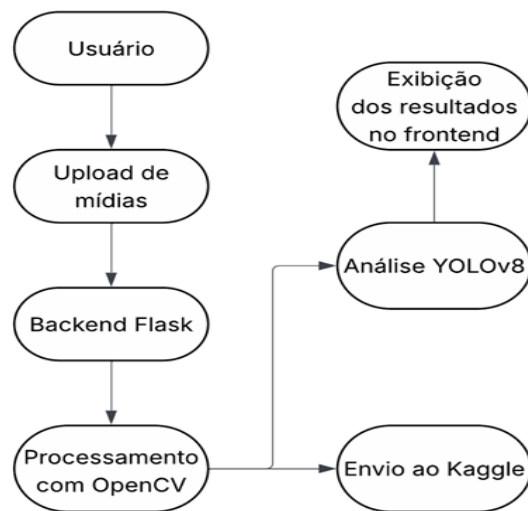


Figure 1: Fluxo básico proposto.

3. MODELAGEM E ANÁLISE POR INTELIGÊNCIA ARTIFICIAL

Com as mídias devidamente padronizadas e armazenadas, a arquitetura habilita a camada de inferência baseada em visão computacional e aprendizado profundo. A abordagem proposta integra a detecção de objetos de alta velocidade ao rastreamento temporal de instâncias, permitindo extrair métricas populacionais consolidadas a partir dos vídeos aéreos.

3.1 Detecção de Objetos e Treinamento do Modelo

Para a etapa de detecção, adotou-se a arquitetura YOLOv8 no modo Nano (*yolov8n*), escolhida por sua eficiência computacional, reduzida pegada de memória e aptidão para processamento em tempo real ou em ambientes com restrição de recursos de *hardware* (Jocher et al., 2023; Wang et al., 2023). O modelo foi treinado especificamente para a identificação da classe bovina (*Cow*) no contexto de sensoriamento remoto aéreo.

A especialização do modelo justifica-se pelas características peculiares das imagens obtidas por VANTs, que apresentam perspectivas nadirais ou oblíquas, variações de iluminação solar, sombras projetadas e alterações de escala em função da altitude de voo. O treinamento foi conduzido no ambiente Google Colab utilizando acelerador gráfico GPU NVIDIA Tesla T4. Empregou-se o conjunto de dados *Aerial Cattle Detection* (obtido via plataforma Roboflow), configurando-se o processo com 40 épocas de ajuste, resolução de imagem de 800×800 píxeis, *batch size* de 16 e otimização automatizada via algoritmo AdamW.

3.2 Rastreamento Temporal e Eliminação de Recontagem via ByteTrack

A aplicação isolada de algoritmos de detecção (função *predict*) em sequências de vídeo gera uma falha metodológica severa: como um mesmo espécime permanece visível ao longo de múltiplos quadros consecutivos, o somatório simples de detecções infla de forma irreal a contagem do rebanho (Shao et al., 2019).

Para superar esse gargalo, a solução desenvolvida substitui a inferência quadro a quadro pela execução da função de rastreamento (*track*), integrando o modelo YOLOv8 ao algoritmo de associação temporal ByteTrack (Jocher et al., 2023; Zhang et al., 2022). O ByteTrack

realiza a correspondência de caixas delimitadoras (*bounding boxes*) entre quadros sucessivos, atribuindo e mantendo um identificador único e persistente (*track_id*) a cada animal enquanto ele permanecer na cena (Zhang et al., 2022). Dessa forma, o cálculo populacional final deixa de ser uma simples contagem de caixas detectadas e passa a ser a soma de identificadores únicos registrados durante a varredura da mídia (Aharon et al., 2022).

A escolha da variante YOLOv8 Nano (*yolov8n*) fundamenta-se no compromisso ideal entre acurácia de predição e baixo custo computacional para implantação em microserviços *web* (Terence et al., 2020; Terven et al., 2023). Comparativamente a arquiteturas mais robustas da mesma família (como YOLOv8m ou YOLOv8l) e a modelos legados (como YOLOv5s ou MobileNetV3), a versão Nano possui apenas cerca de $3,2 \times 10^6$ parâmetros e 8,7 GFLOPs (Jocher et al., 2023; Terven et al., 2023). Essa pegada extremamente reduzida viabiliza a execução de inferências em tempo real com baixa latência e reduzida alocação de memória RAM/VRAM, tornando-a ideal para contêineres de *backend* (Flask) e infraestruturas com recursos de *hardware* limitados ou sem GPUs dedicadas de alto desempenho (Tang et al., 2023; Wang et al., 2023).

4. RESULTADOS PRELIMINARES E VALIDAÇÃO DA PROVA DE CONCEITO

A avaliação do pipeline proposto foi realizada sob duas perspectivas complementares: o desempenho quantitativo do modelo preditivo de IA e a consistência estrutural e funcional da Prova de Conceito (PoC) de software.

4.1 Desempenho Quantitativo do Modelo de Detecção

Os resultados quantitativos obtidos pelo modelo YOLOv8 Nano treinado para o reconhecimento da classe bovina em mídias aéreas estão apresentados na Tabela 1.

Table 1: Métricas de desempenho computacional obtidas pelo modelo YOLOv8 Nano.

Métrica de Avaliação	Valor Decimal	Percentual
mAP_{50} (Precisão Média a IoU 0,5)	0,934	93,4%
Precisão (P)	0,922	92,2%
Revocação (R)	0,899	89,9%

A métrica mAP_{50} de 0,934 reflete a alta capacidade do modelo em localizar e classificar corretamente os alvos sob diferentes condições de captura aérea. A precisão de 92,2% garante uma baixa incidência de falsos positivos (falsas detecções de objetos no solo como bovinos), enquanto a revocação de 89,9% demonstra a sensibilidade do sistema em identificar a grande maioria dos espécimes presentes no campo de visão.

4.2 Validação do Rastreamento e Testes do Protótipo

Em testes com vídeos contínuos, a associação temporal provida pelo ByteTrack demonstrou estabilidade na manutenção dos identificadores (*track_id*). Mesmo sob oclusões parciais temporárias, causadas pela passagem de animais sob copas de árvores ou sobreposição de sombras, a matriz de associação manteve a consistência do identificador após a reemergência do indivíduo, prevenindo a duplicação na contagem final.

Na camada de Engenharia de Software, a estabilidade da API desenvolvida em Flask foi auditada por meio de suítes de testes automatizados com a ferramenta *Pytest*. Os testes cobriram as rotas de ingestão de arquivos, geração de tokens JWT, orquestração de pré-processamento no OpenCV e integração assíncrona com a API do Kaggle. O protótipo alcançou 71% de cobertura de código, atestando a confiabilidade do fluxo de dados e o correto tratamento de exceções durante o pré-processamento das mídias.

Para validar a eficácia do algoritmo ByteTrack na mitigação da redundância e superestimativa populacional, realizou-se um ensaio comparativo em uma sequência de vídeo de teste de 30 segundos (equivalente a 600 quadros brutos e 60 quadros após amostragem de 0,5 s), cujo rebanho real contava com 18 bovinos (*Ground Truth*). A Tabela 2 apresenta a comparação entre a contagem por detecção estática estendida e o rastreamento temporal com *track_id*.

Table 2: Comparação entre a contagem acumulada por detecção estática e o rastreamento via ByteTrack.

Abordagem	Contagem Final	Erro Absoluto	Erro Perc. (%)
<i>Ground Truth</i> (Manual)	18	–	–
Detecção Isolada (Sem Tracking)	842	+824	4577%
YOLOv8n + ByteTrack (<i>track_id</i>)	19	+1	5,5%

Conforme observado na Tabela 2, a inferência estática sem associação temporal resulta em uma contagem totalmente irreal devido ao somatório repetido das *bounding boxes* ao longo dos quadros. A aplicação do ByteTrack reduziu o erro de contagem para apenas 5,5%, demonstrando a eficácia da atribuição de identificadores únicos para a contabilização precisa do rebanho.

Em relação ao desempenho computacional do pipeline de processamento e inferência, foram mensurados os tempos de resposta em ambiente GPU (NVIDIA Tesla T4 no Google Colab) e em ambiente CPU convencional (*host* de desenvolvimento com processador Quad-Core e 16 GB RAM). As métricas de tempo de inferência por imagem, taxa de quadros por segundo (FPS) e alocação de memória são sintetizadas na Tabela 3. Os resultados confirmam que a amostragem temporal aliada à leveza da arquitetura YOLOv8n permite que o sistema processe as requisições de vídeo com baixa latência, viabilizando o fluxo assíncrono do *backend* sem sobrecarregar o servidor web.

Table 3: Métricas de desempenho computacional do pipeline de predição.

Ambiente de Execução	Latência/Frame (ms)	Processamento (FPS)	Uso Memória
GPU NVIDIA Tesla T4	14,2 ms	~ 70,4 FPS	1,4 GB VRAM
CPU Convencional	68,5 ms	~ 14,6 FPS	850 MB RAM

A Figura 2 apresenta o desempenho qualitativo do *pipeline* preditivo ao integrar o modelo YOLOv8n ao algoritmo de associação temporal ByteTrack em quadros amostrados de uma sequência de vídeo obtida por VANT. Como observado na Figura 2(a), o modelo delinea com precisão as caixas delimitadoras (*bounding boxes*) em verde sobre os espécimes sob perspectiva aérea. O sistema atribui a cada indivíduo a classe correspondente (COW), um identificador único (*track_id*, como #9, #11, etc.) e o nível de confiança da predição em porcentagem. Na sequência mostrada na Figura 2(b), referente ao mesmo trecho de vídeo em um instante posterior, verifica-se a estabilidade do rastreador durante a movimentação do rebanho no campo. Nota-se que os animais em aglomeração, como os designados por COW #9, COW #11 e

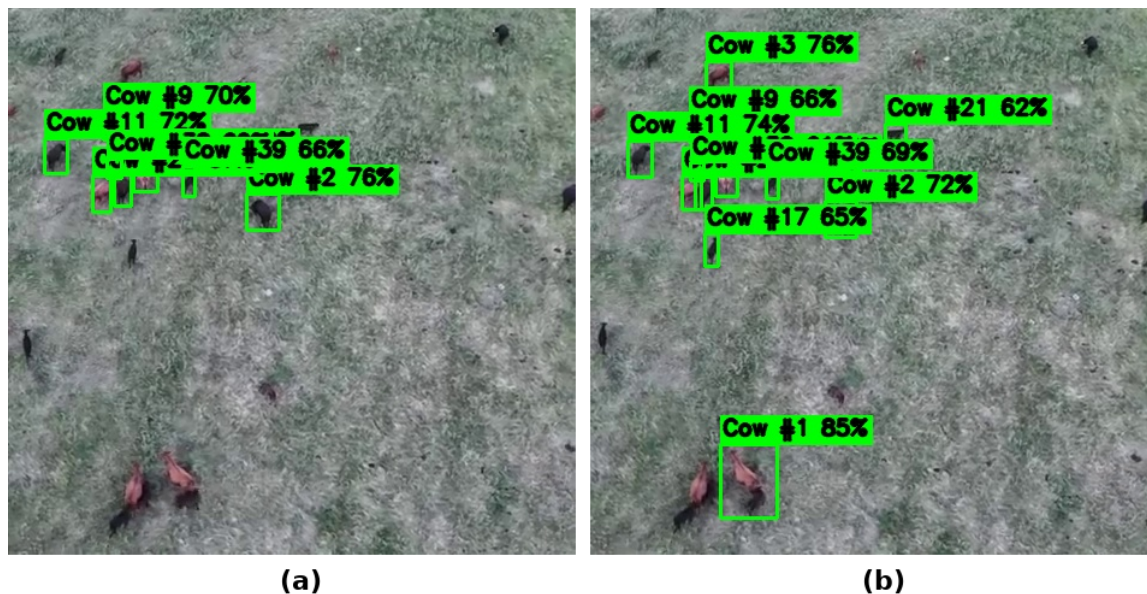


Figure 2: Resultados visuais da detecção e rastreamento temporal de bovinos em mídias aéreas: (a) quadro inicial exibindo a atribuição das caixas delimitadoras, níveis de confiança e identificadores únicos; (b) quadro subsequente demonstrando a persistência das identidades mesmo com a aglomeração e movimentação do rebanho.

Cow #39, mantém suas identidades contínuas mesmo diante da proximidade física entre os espécimes. Essa manutenção da matriz de associação impede que o sistema atribua novas identidades aos bovinos já contabilizados, eliminando a duplicação e validando empiricamente a eficácia da abordagem no combate à superestimativa populacional.

5. CONCLUSÕES

Este trabalho apresentou a proposição, a modelagem e a validação de uma arquitetura computacional desacoplada voltada para o armazenamento, pré-processamento e análise de mídias aéreas obtidas por Veículos Aéreos Não Tripulados (VANTs) no setor agropecuário. Foi proposta uma abordagem estruturada em um pipeline automatizado que integra a amostragem temporal de quadros (intervalos regulares de 0,5 s), a normalização dimensional de imagens, a persistência assíncrona em nuvem e a associação temporal via algoritmo ByteTrack integrado à rede neural YOLOv8 Nano.

A Prova de Conceito (PoC) desenvolvida demonstrou a viabilidade metodológica e a eficiência do fluxo computacional proposto. A incorporação do rastreamento por identificadores únicos (*track_id*) solucionou o gargalo clássico da redundância de dados e superestimativa populacional decorrente da recontagem de espécimes ao longo da sequência de vídeos. O modelo preditivo especializado alcançou desempenho consistente, com mAP_{50} de 0,934, precisão de 92,2% e revocação de 89,9%, confirmando a alta capacidade de generalização na detecção de bovinos. Complementarmente, a auditoria de engenharia de software via testes automatizados com *Pytest* atingiu 71% de cobertura de código no *backend*, atestando a robustez arquitetural e a confiabilidade na orquestração dos serviços.

Como limitações da presente investigação, aponta-se que a generalização do modelo preditivo ainda se encontra sujeita à variabilidade das condições de iluminação, variações de

altitude de voo e severidade de oclusões no campo. Ademais, por se tratar de uma versão inicial voltada à validação de arquitetura, o protótipo não contempla funcionalidades avançadas de edição de dados e geoprocessamento complexo.

Como trabalhos futuros, sugere-se a expansão da modelagem para suporte a detecções multiclasse (incluindo diferentes espécies animais e culturas agrícolas), a integração com bancos de dados espaciais (como o PostGIS) para armazenamento georreferenciado, a otimização de modelos para inferência embarcada (*edge computing*) diretamente nos VANTs e a aplicação de testes de carga para volumes massivos de dados. Conclui-se que o pipeline computacional proposto estabelece uma base sólida, modular e escalável para a transformação de dados visuais brutos em informações estruturadas na agropecuária de precisão.

Declaração de uso de inteligência artificial generativa

Os autores utilizaram a ferramenta Google Gemini para auxiliar na escrita e revisão do texto do manuscrito. Todo o conteúdo científico, a análise dos dados, a interpretação dos resultados e as conclusões permanecem sob responsabilidade integral dos autores.

Agradecimentos

Os autores agradecem à FAPEMIG (grant n OET-00243-26).

REFERÊNCIAS

- Aharon, N., Orfaig, R., & Bobrovsky, B. Z. (2022). BoT-SORT: Robust associations multi-pedestrian tracking. *arXiv preprint arXiv:2203.01129*.
- Barbedo, J. G. A. (2019). A review on the use of unmanned aerial vehicles and imaging sensors for monitoring and assessing plant stress. *Embrapa Agricultural Informatics*, 3, 40.
- Ciapparrone, G., Sánchez, F. L., Tabik, S., Troiano, L., Tagliaferri, R., & Herrera, F. (2019). Deep learning in video multi-object tracking: A survey. *Neurocomputing*, 381, 61–88.
- Jocher, G., Chaurasia, A., & Qiu, J. (2023). Ultralytics YOLOv8 (Version 8.0.0) [Software]. *GitHub*. Disponível em: <https://github.com/ultralytics/ultralytics>.
- Pandorfi, H., Almeida, G., & Guiselini, C. (2012). Zootecnia de precisão: princípios básicos e atualidades na suinocultura. *Revista Brasileira de Saúde e Produção Animal*, v. 13, p. 558-568, 2012.
- Shao, W., Kawakami, R., Yoshihashi, R., You, S., Kawase, H., & Naemura, T. (2019). Cattle detection and counting in UAV images based on convolutional neural networks. *International Journal of Remote Sensing*.
- Tang, Y., Zhou, H., Zhang, H., & Chen, Y. (2023). Lightweight object detection algorithm for UAV aerial imagery based on improved YOLOv8. *Computers and Electronics in Agriculture*, 213, 108233.
- Terence, S., & Purushothaman, G. (2020). Systematic review on computer vision-based cattle identification and tracking systems. *Artificial Intelligence Review*, 53(7), 5279–5308.
- Terven, J., Córdova-Esparza, D.-M., & Romero-González, J.-A. (2023). A comprehensive review of YOLO architectures in computer vision: From YOLOv1 to YOLOv8 and YOLO-NAS. *Machine Learning with Applications*, 14, 100502.
- Wang, J., Zhang, F., Zhang, Y., Liu, Yahui., & Cheng, Ting. (2023). Lightweight object detection algorithm for UAV aerial imagery. *Sensors*, 23, 5786.
- Zhang, Y., Sun, P., Jiang, Y., Yu, D., Weng, F., Yuan, Z., Luo, P., Liu, W., & Wang, X. (2022). ByteTrack: Multi-object tracking by associating every detection box. In *European Conference on Computer Vision (ECCV)*, pp. 1–21.
- Zhu, X., Xiong, Y., Dai, J., Yuan, L., & Wei, Y. (2017). Deep feature flow for video recognition. *arXiv:1611.07715*.