

## Aprendizado auto-supervisionado de representações para monitoramento de pastagens por imagens em alta resolução



*Empresa Brasileira de Pesquisa Agropecuária  
Embrapa Agricultura Digital  
Ministério da Agricultura e Pecuária*

**BOLETIM DE PESQUISA  
E DESENVOLVIMENTO  
57**

Aprendizado auto-supervisionado de  
representações para monitoramento de  
pastagens por imagens em alta resolução

*Thiago Teixeira Santos  
Luciano Vieira Koenigkan  
Celina Maki Takemura  
Patrícia Menezes Santos*

**Embrapa Agricultura Digital** Comitê Local de Publicações

Av. Dr. André Tosello, 209 - Cidade Universitária  
Campinas, SP, Brasil  
CEP. 13083-886  
Fone: (19) 3211-5700  
www.embrapa.br  
www.embrapa.br/fale-conosco/sac

Presidente  
*Carla Geovana do Nascimento Macário*

Secretária-Executiva  
*Maria Fernanda Moura*

Membros  
*Alexandre de Castro, membro indicado, Carla Cristiane Osawa, membro nato, Debora Pignatari Drucker, membro eleito, Graziella Galinari, membro nato, Ivan Mazoni, membro eleito, João Camargo Neto, membro indicado, Joao Francisco Goncalves Antunes, membro eleito, Magda Cruciol, membro nato.*

Revisão de texto  
*Graziella Galinari*

Normalização bibliográfica  
*Carla Cristiane Osawa*

Projeto gráfico da coleção  
*Carlos Eduardo Felice Barbeiro*

Editoração eletrônica  
*Magda Cruciol*

Foto da capa  
*Vant - Thiago Santos*

Publicação digital:PDF (2023)

**Todos os direitos reservados.**

A reprodução não autorizada desta publicação, no todo ou em parte, constitui violação dos direitos autorais (Lei nº 9.610).

**Dados Internacionais de Catalogação na Publicação (CIP)**  
**Embrapa Agricultura Digital**

---

Aprendizado auto-supervisionado de representações para monitoramento de pastagens por imagens em alta resolução / Thiago Teixeira Santos ... [et al.] – Campinas : Embrapa Agricultura Digital, 2023  
PDF (26 p.) : il. color. - (Boletim de pesquisa e desenvolvimento / Embrapa Agricultura Digital, ISSN 2764-2623 ; 57).

1. Aprendizado de máquina. 2. Imagem de pastagem. 3. Metodologia. I. Santos, Thiago Teixeira. II. Embrapa Agricultura Digital. III. Série.

CDD (21. ed.) 006.31

# Sumário

<b>Resumo</b> . . . . .	<b>5</b>
<b>Abstract</b> . . . . .	<b>6</b>
<b>Introdução</b> . . . . .	<b>7</b>
<b>Material e métodos</b> . . . . .	<b>8</b>
<b>Resultados e discussão</b> . . . . .	<b>16</b>
<b>Conclusões</b> . . . . .	<b>23</b>
<b>Agradecimentos</b> . . . . .	<b>23</b>

# Aprendizado auto-supervisionado de representações para monitoramento de pastagens por imagens em alta resolução

Thiago Teixeira Santos<sup>1</sup>  
Luciano Vieira Koenigkan<sup>2</sup>  
Celina Maki Takemura<sup>3</sup>  
Patrícia Menezes Santos<sup>4</sup>

**Resumo** – O monitoramento de forrageiras é um requisito para um bom manejo de pastagem, que produza ganhos de produção e evite degradação da área. O sensoriamento remoto de alta resolução, obtido por imageamento aéreo ou satélites, combinado a sistemas de análise de imagem, são uma alternativa natural a um monitoramento preciso da pastagem. Porém, a complexidade dos padrões de pastagem somada à dificuldade de obtenção de verdade de campo em larga escala para o treinamento de sistemas de aprendizado de máquina são um entrave ao desenvolvimento de sistemas efetivos de monitoramento autônomo. Neste trabalho, apresentamos uma metodologia baseada em aprendizado auto-supervisionado de representações capazes de caracterizar pequenas amostras de pastagem ( $\leq 1 \text{ m}^2$ ), permitindo varreduras aéreas em grande nível de detalhe. Demonstramos que amostras com representações similares podem apresentar interpretações diretas, como maior ou menor presença de solo exposto, ou maior vigor das plantas, por exemplo. Tais representações são potencialmente transferíveis para outras tarefas, para as quais sistemas de aprendizado de máquina podem ser treinados com uma menor quantidade de dados supervisionados (isto é, com aferição em campo).

**Termos para indexação:** pastagens, *masked autoencoders*, drones, sensoriamento remoto.

---

<sup>1</sup>Cientista da Computação, doutor em Ciências da Computação, pesquisador da Embrapa Agricultura Digital, Campinas, SP.

<sup>2</sup>Cientista da Computação, mestre em Engenharia Elétrica, analista da Embrapa Agricultura Digital, Campinas, SP.

<sup>3</sup>Cientista da Computação, doutora em Ciências da Computação, pesquisadora da Embrapa Territorial, Campinas, SP.

<sup>4</sup>Engenheira Agrônoma, doutora em Ciência Animal e Pastagens, pesquisadora da Embrapa Pecuária Sudeste, São Carlos, SP.

## Self-supervised representation learning for high-resolution image-based pasture monitoring

**Abstract** – The monitoring of forage is a requirement for effective pasture management that yields production gains and prevents area degradation. High-resolution remote sensing, obtained through aerial imaging or satellites, combined with image analysis systems, presents a natural alternative for precise pasture monitoring. However, the complexity of pasture patterns coupled with the challenge of obtaining large-scale ground truth data for training machine learning systems poses a barrier to the development of effective autonomous monitoring systems. In this work, we present a methodology based on self-supervised representation learning for automatic able to characterize small pasture samples ( $\leq 1 \text{ m}^2$ ), enabling detailed aerial scans. We demonstrate that samples with similar representations can exhibit direct interpretations, such as higher or lower presence of exposed soil, or stronger plant vitality, for instance. Such representations are potentially transferable to other tasks, allowing machine learning systems to be trained with fewer supervised data (i.e., field measurements).

**Index terms:** pastures, masked autoencoders, drones, remote sensing.

## Introdução

---

Na pecuária, ganhos de produtividade significativos podem ser obtidos com o manejo apropriado de pastagens e rebanhos, devido a uma alocação adequada de alimento aos animais. Quantidades expressivas de pastagens em todo o mundo são consideradas *degradadas* (Dias-Filho, 2014), apresentando solos empobrecidos e desertificação, um processo causado comumente pelo uso demasiadamente intensivo da pastagem decorrente de manejo inadequado (Reinermann et al., 2020).

Um manejo apropriado requer um monitoramento frequente do estado da pastagem, uma tarefa difícil de ser executada em grandes extensões de terra, principalmente quando são necessários níveis mais finos de detalhe (Reis et al., 2020), por exemplo, na estimativa da biomassa disponível ou na avaliação de infestação por espécies invasoras. A necessidade de ganhos de escala na análise detalhada de grandes áreas de pastagem e a automatização do processo de monitoramento faz do sensoriamento remoto por imagens de alta resolução uma opção natural. Aeronaves remotamente pilotadas (ARPs ou *drones*) (Freitas et al., 2022) e satélites de alta resolução (Reis et al., 2020) são duas opções disponíveis para a obtenção de dados de alta resolução (de 1 cm/pixel a 3 m/pixel). Contudo, a análise dessas imagens também necessita ser automatizada e, dada a complexidade da tarefa de análise de imagens naturais, soluções baseadas em *aprendizado de máquina* se mostram mais adequadas para lidar com a complexidade de padrões observados nas imagens em problemas de classificação, regressão e segmentação.

Um problema recorrente no uso de aprendizado de máquina na agropecuária é a anotação dos dados com *verdade de campo* (*ground-truth*), valores necessários para guiar o processo de otimização que busca parâmetros adequados aos modelos de aprendizado, permitindo que eles alcancem desempenho satisfatório na tarefa para a qual estão sendo treinados. No caso da agropecuária, essa verdade de campo consiste em valores com significado agrônômico como quantidade de biomassa, volume de safra, presença de pragas ou plantas daninhas, etc. Em geral, o processo de coleta de verdade de campo em grandes áreas é custoso em termos de trabalho e tempo, dificilmente escalável. Enquanto os dados de imageamento remoto podem ser obtidos em larga escala, a partir de satélites e aeronaves, as anotações de campo associadas são obtidas em quantidade muito menor.

Considere o problema da estimativa de biomassa de forragem em pastagens. Enquanto uma enorme quantidade de dados pode ser obtida com voos periódicos de ARPs (Santos; Koenigkan, 2018; Oliveira et al., 2021; Freitas et al., 2022), valores de biomassa, em kg/m<sup>2</sup>, necessitam ser aferidos manualmente, seguindo protocolos que definem procedimentos de seleção, georreferenciamento, corte, separação e pesagem (Bayma-Silva et al., 2019; Reis et al., 2020). Enquanto centenas de milhares de amostras de imagem podem ser capturadas por ARPs e satélites (Figura 1 B), poucas centenas de amostras com biomassa aferida podem ser obtidas após extensivas campanhas de campo, envolvendo equipes numerosas (Figura 1 C).

O pré-treinamento por aprendizado auto-supervisionado (*self-supervised pre-training*) é uma técnica particularmente adequada para tal tipo de problema: uma abundante quantidade de dados sem anotação somada a uma quantidade menor de dados anotados (Güldenring; Nalpantidis, 2021). A etapa de pré-treinamento envolve um procedimento de anotação automática, para uma tarefa artificial cujo único propósito é permitir ao modelo aprender representações que capturem as regularidades dos padrões observados nos dados. As anotações automáticas guiam o processo de otimização, permitindo o aprendizado auto-supervisionado de representações. A tarefa de pré-treinamento pode ser a reconstrução de imagens artificialmente degradadas (He et al., 2022) ou a associação de imagens similares (Caron et al., 2018), entre outras (Liu et al., 2023). Numa etapa subsequente, os dados manualmente anotados, oriundos das coletas em campo, são empregados no refinamento do modelo para a *tarefa de aplicação* (*downstream task*).

Até onde sabemos, Güldenring; Nalpantidis (2021) foi o primeiro trabalho a utilizar sistematicamente o aprendizado auto-supervisionado com redes neurais em aplicações agrícolas. Esses autores utilizaram o método SwAV (Caron et al., 2018) no pré-treinamento de imagens não anotadas, posteriormente realizando um refinamento (*fine-tuning*) em tarefas de classificação de plantas daninhas e segmentação de anomalias em talhões. Apesar dos resultados promissores, os autores observaram que SwAV pertence ao grupo de métodos que aprendem representações de nível mais alto (imagem), mas não explicitamente ao nível de pixel, o que pode ser crucial em certas tarefas de aplicação.

Este boletim apresenta uma metodologia para o aprendizado auto-supervisionado de representações de nível fino ( $\leq 1$  m) em imagens de pastagens, adequadas a problemas de classificação e regressão como estimativa de biomassa. A coleta de dados com verdade de campo é similar ao visto nos trabalhos de Reis et al. (2020) e Freitas et al. (2022). Contudo, diferentemente desses trabalhos, as amostras de pastagem não são representadas por índices ou descritores de textura. Similarmente a Güldenring; Nalpantidis (2021), empregamos aprendizado automático de representações por redes neurais, mas optamos pelo uso de *masked autoencoders* (He et al., 2022), uma técnica capaz de aprender representações ao nível de pixel e assim explorar as características morfológicas da pastagem observadas na alta resolução de 1 cm/pixel empregada neste trabalho, em que lâminas de folhas são visíveis na imagem.

## Material e métodos

---

As duas áreas de trabalho se encontram na Fazenda Canchim, sede da Embrapa Pecuária Sudeste, localizada no município de São Carlos, SP. Consistem em (1) uma pastagem sob sistema de manejo intensivo com lotação rotacionada, estabelecida em 2012 com *Brachiaria brizantha* e (2) uma pastagem em sistema extensivo, estabelecida em 2007 com *Brachiaria decumbens*. Denominaremos por INT e EXT as áreas em sis-

tema intensivo e extensivo, respectivamente. A Figura 1 exibe a área INT, como vista em 31 de janeiro de 2019 e localizada em -21,97599N, -47,851140E.

Aqui, adaptamos ao estudo de pastagens uma metodologia popularizada recentemente no campo do aprendizado semi-supervisionado. Este arcabouço adaptado consiste no (1) imageamento de alta-resolução das áreas de pastagem com ARPs; (2) montagem automática de bases de dados não-supervisionadas a partir de ortomapas, com a extração de amostras (*patches*) de até 1 m<sup>2</sup>; (3) aprendizado auto-supervisionado utilizando técnicas como aprendizado contrastivo (He et al., 2020) ou mascaramento (He et al., 2022); (4) análise das representações aprendidas com o uso do algoritmo K-médias e (5) *linear probing* (Asano et al., 2020) com dados supervisionados resultantes de coleta ou outra forma de anotação, para avaliação da acurácia do modelo na tarefa de aplicação (estimativa de biomassa, por exemplo).

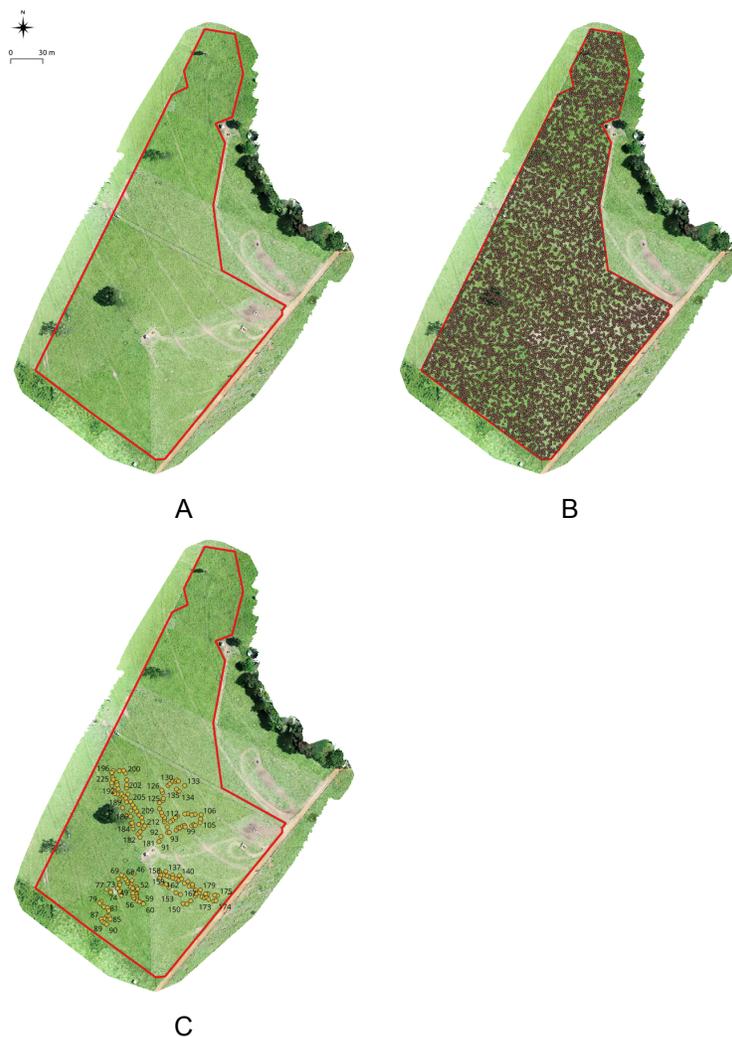
## Bases de dados

Dados foram coletados, seguindo o protocolo descrito em Bayma-Silva et al. (2019), em dois momentos: no dia 17 de abril de 2018 e no dia 31 de janeiro de 2019. Uma aeronave DJI Phantom 4 Pro foi utilizada no imageamento da área. A câmera embarcada na aeronave possui sensor de 1"CMOS com resolução de 20 megapixels (5472 × 3648) em RGB e FOV (campo de visão) de 84°. O sistema OpenDroneMap foi utilizado na construção de ortomosaicos, como o visto na Figura 1 (A), com resolução inferior a 1 cm/pixel, a partir das imagens obtidas, como descrito em Santos; Koenigkan (2018).

Amostras de forragem foram coletadas em diversos pontos da pastagem (Bayma-Silva et al., 2019). Voos foram realizados com a aeronave imediatamente antes das coletas, para obter imagens antes do corte da vegetação. As amostras foram então coletadas em regiões de 0,25 m<sup>2</sup>, demarcadas com o uso de gabarito quadrado de metal rígido com 0,5 m de lado. A forragem era extraída da área demarcada com o uso de uma roçadeira, ensacada, pesada e levada para separação em laboratório. Um marcador de papel era deixado no centro da área, após a extração, para identificação no local de coleta. Novos voos eram então realizados para imageamento das áreas marcadas e identificação dos pontos de coleta, como pode ser visto na Figura 1 (C).

### Base de dados não-supervisionada

Um conjunto de 100.000 pontos aleatoriamente selecionados dentro do perímetro da área de estudo foi utilizado na extração de amostras para o pré-treinamento auto-supervisionado. A Figura 1 (B) exibe um subconjunto dos primeiros 10.000 pontos para fins de visualização. Para cada ponto, é extraída do ortomapa uma imagem correspondente a um quadrado, com centro no ponto e lados paralelos aos eixos Norte-Sul e Leste-Oeste. Foram produzidos conjuntos de imagens para dois tamanhos diferentes: 0,5 × 0,5 m, que corresponde às dimensões do gabarito de metal utilizado nas coletas de campo, e 1,0 × 1,0 m. Para compensar diferenças de luminosidade, causadas por variações de horário e nuvens, as imagens são processadas no espaço de



**Figura 1.** Área do sistema intensivo rotacionado, registrada em janeiro de 2019: ortomapa produzido com área de trabalho em vermelho (A); amostra de 10.000 pontos aleatoriamente escolhidos para treinamento não-supervisionado (B); amostra de 177 pontos nos quais foram realizadas coletas de forragem (C).

**Tabela 1.** Base de dados supervisionada

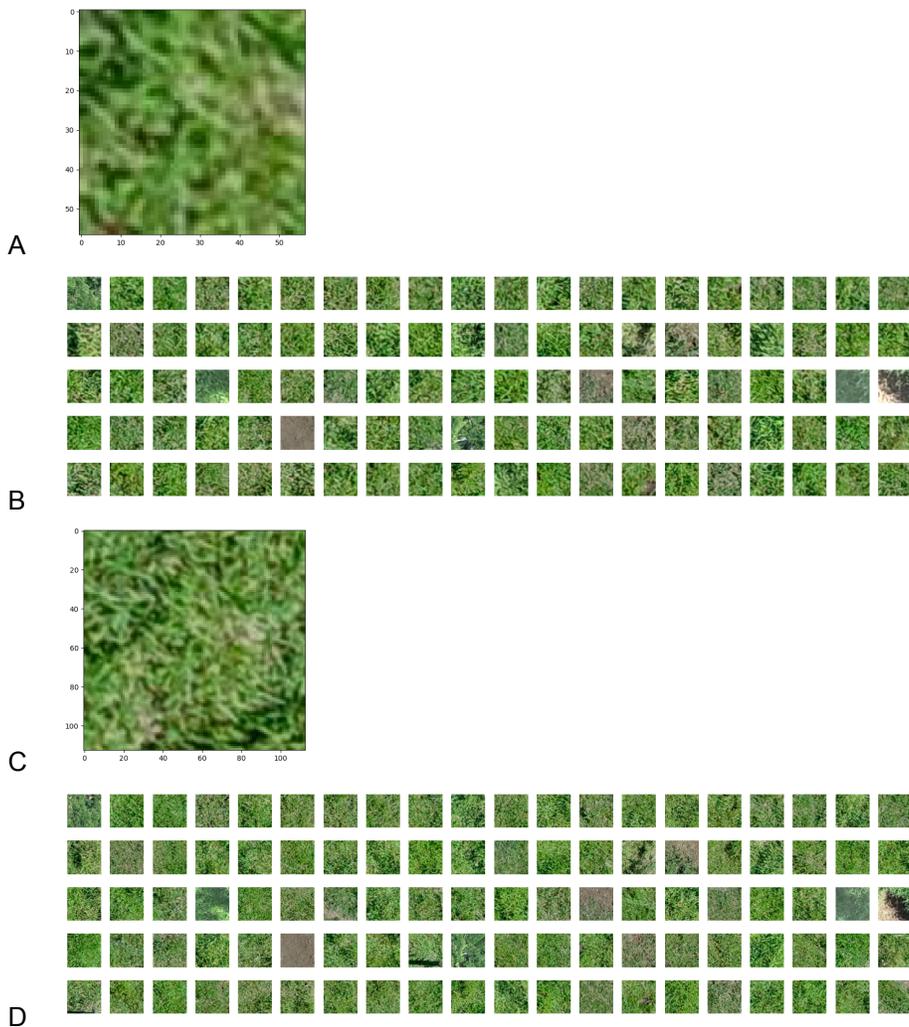
Código	Sistema	Data de coleta	Número de amostras
INT2018	Intensivo	17/04/2018	134
INT2019	Intensivo	31/01/2019	177
EXT2018	Extensivo	17/04/2018	44
EXT2019	Extensivo	31/01/2019	45
Total			400

cor CIE Lab: a média da luminosidade (canal L) é ajustada para 128 e extremos são truncados em 0 e 255 (considerando o intervalo de 8-bits que vai de 0 a 255), posteriormente sendo reconvertidas ao espaço de cor RGB.

Considere, por exemplo, que o ortomapa da área do sistema intensivo (Figura 1) tem GSD de 0.0089 m (8,9 mm por pixel). As amostras de  $0,5 \times 0,5$  m são um conjunto de 100.000 imagens RGB de 8-bits com resolução de  $57 \times 57$  pixels, enquanto as amostras de  $1 \text{ m}^2$  apresentam resolução de  $113 \times 113$  pixels. Exemplos das imagens podem ser vistos na Figura 2.

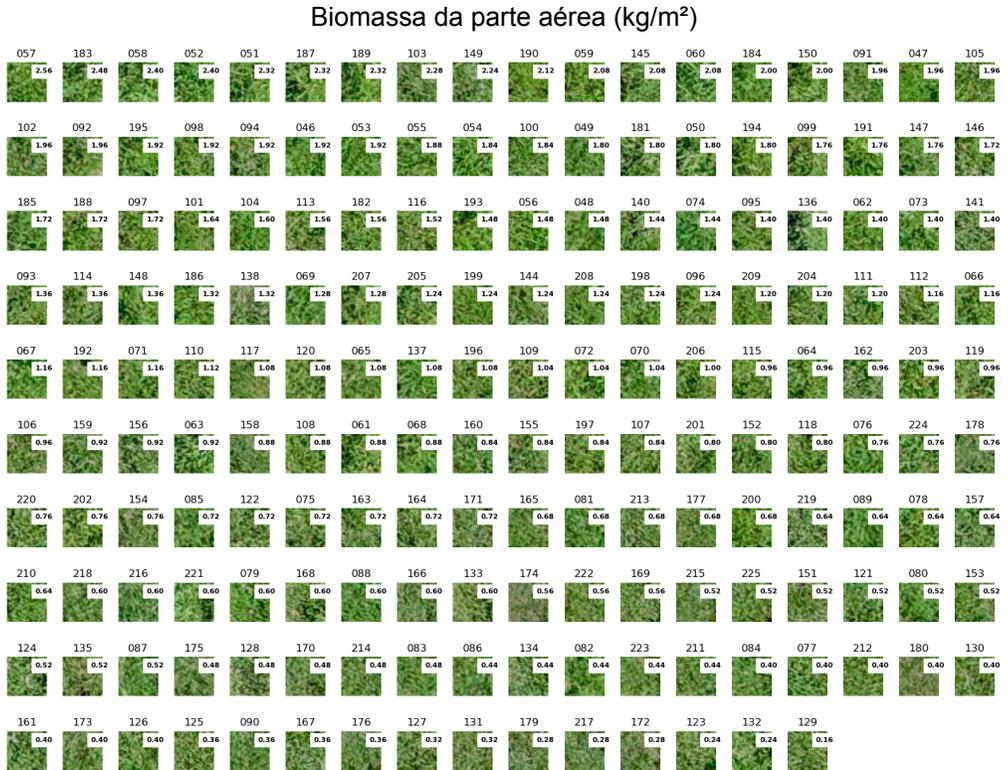
#### Base de dados supervisionada

A base de dados supervisionada foi produzida seguindo o mesmo procedimento descrito anteriormente para a base não-supervisionada, mas ao invés de pontos aleatórios no ortomapa, foram utilizadas as localizações dos pontos de coleta de forragem (Tabela 1), totalizando 400 amostras. Para essas amostras, existem associados valores de biomassa ( $\text{kg}/\text{m}^2$ ), oriundos da aferição em campo: cada amostra possui um identificador numérico que associa a medição em campo com a imagem na base de dados (Figura 1 C). A Figura 3 exibe uma visualização da base INT2019, em que é possível ver para cada amostra seu identificador e o valor de biomassa aferido.



**Figura 2.** Imagens na base de dados não-supervisionada: amostra de  $57 \times 57$  pixels correspondente a uma área de  $0,25 \text{ m}^2$  (A); cem amostras de  $0,25 \text{ m}^2$  (B); amostra de  $113 \times 113$  pixels correspondente a uma área de  $1 \text{ m}^2$  e centrada no mesmo ponto que a amostra em (A) (C); cem amostras de  $1 \text{ m}^2$ , obtidas nos mesmos pontos que os de (B) (D).

Note que as amostras de  $1 \text{ m}^2$  ampliam o entorno observado nas amostras de  $0,25 \text{ m}^2$ .



**Figura 3.** Imagens na base de dados supervisionada: as 177 amostras (0,25 m<sup>2</sup>) do conjunto INT2019.

O número acima de cada imagem é o código numérico da amostra.

O valor no canto superior direito é o valor da biomassa da parte aérea em kg/m<sup>2</sup>, aferido pela coleta em campo.

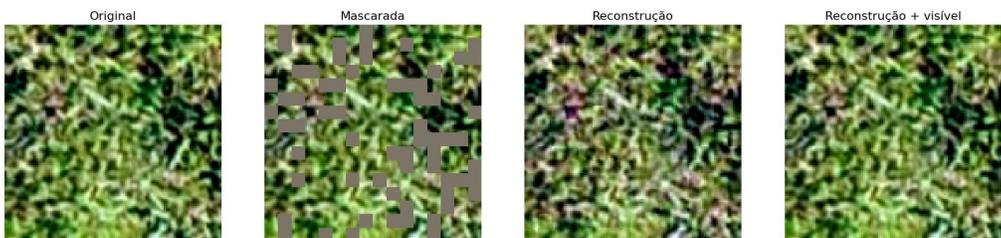
### Pré-treinamento com *masked autoencoders*

*Masked autoencoders* (MAE) são uma técnica para aprendizado de representações de imagens de forma auto-supervisionada, baseada em um conceito simples: a remoção aleatória de parte do conteúdo da imagem, impondo ao pré-treinamento a tarefa de predição do conteúdo removido (He et al., 2022), como pode ser visto na Figura 4. Em 2023, MAE é considerado um método do estado da arte no aprendizado auto-supervisionado de representações, tendo sido empregado recentemente no treinamento de modelos fundamentais em visão computacional (Kirillov et al., 2023).

As representações das imagens são obtidas por uma parte da rede neural chamada de *backbone*, que recebe como entrada imagens na forma de tensores  $H \times W \times C$ ,

em que  $H \times W$  são as dimensões das imagens e  $C$  o número de canais (3 no caso de imagens RGB) e, após o processamento por diversas camadas neurais, produz uma representação na forma de um vetor com  $D$  dimensões. No caso dos MAE, o *backbone* deve ser necessariamente um *Vision Transformer* (ViT) (Dosovitskiy et al., 2021). Neste trabalho, empregamos ViT-base, um *backbone* com 12 camadas que produz um vetor com 728 dimensões. Ou seja, se considerarmos uma imagem RGB de entrada com  $64 \times 64$  pixels de resolução, ViT-base tomará como entrada um tensor  $64 \times 64 \times 3$  (12.288 dimensões) e produzirá um vetor com  $D = 728$  dimensões, obtendo assim uma representação que é não só mais compacta, mas também efetiva (idealmente) na resolução de tarefas visuais (He et al., 2022; Kong et al., 2023). Dois hiperparâmetros são necessários ao treinamento de MAEs, o *tamanho do recorte* (*patch size*) e a taxa percentual de mascaramento, que define a porcentagem dos recortes removidos da imagem. A Figura 4 exibe um exemplo no qual a taxa de mascaramento é de 25%.

Para a implementação de MAE e ViT-base, bem como o processo de treinamento auto-supervisionado, utilizamos a biblioteca MMSelfSup versão 1.0.0 (desenvolvida sobre PyTorch) (OpenMMLab, 2021). O treinamento foi realizado no *cluster* de GPUs da Embrapa Agricultura Digital.

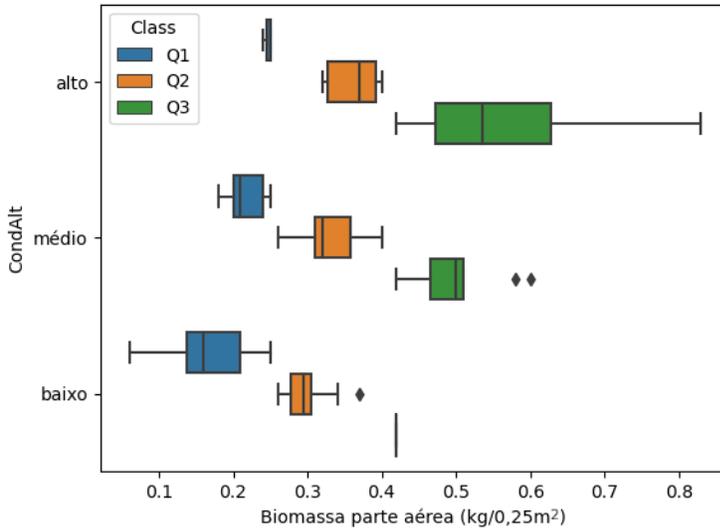


**Figura 4.** *Masked autoencoder*: blocos da imagem original são aleatoriamente removidos, produzindo a imagem mascarada; o *autoencoder* é treinado para produzir uma reconstrução a partir da imagem mascarada que seja a mais próxima possível da imagem original.

## Linear probing

*Linear probing* consiste em abordar uma tarefa de aplicação (classificação, por exemplo) simplesmente treinando um classificador linear sobre representações aprendidas em pré-treinamento, as quais são mantidas fixas. Ou seja, um *backbone* como ViT-base tem seus parâmetros, aprendidos previamente, mantidos fixos durante a tarefa de classificação. Como observado por Asano et al. (2020), classificadores lineares dependem fortemente da qualidade da representação aprendida, uma vez que seu poder discriminativo é baixo. Desta forma, o *linear probing* é uma maneira de avaliar se a representação aprendida pelo modelo no pré-treinamento pode ser efetiva na tarefa de aplicação. Para formular a tarefa de estimativa de biomassa como um problema de

classificação, quantizamos os valores de biomassa em três classes distintas, correspondendo a intervalos de baixa, média e alta quantidade de forragem (Figura 5).



**Figura 5.** Valores de biomassa (INT2019), quantizados em três classes distintas (Q1, Q2 e Q3).

Nota: no eixo das ordenadas, é possível observar o valor de altura da pastagem, aferido em campo juntamente com a biomassa: amostras com maior altura tendem a apresentar maior biomassa.

## Resultados e discussão

Numa análise qualitativa, desejamos saber se as representações obtidas pelo modelo, pré-treinado via MAE, são úteis na análise de pastagem. Utilizamos as primeiras 10.000 amostras aleatórias de pastagem no treinamento de um *backbone* ViT-base. Em seguida, utilizamos o *backbone* para encontrar representações: vetores 728-d para cada uma das 100.000 amostras da base não-supervisionada. Finalmente, utilizamos o algoritmo *K*-médias (Hastie et al., 2009) para encontrar 7 grupos (*classes* ou *clusters*)<sup>5</sup>.

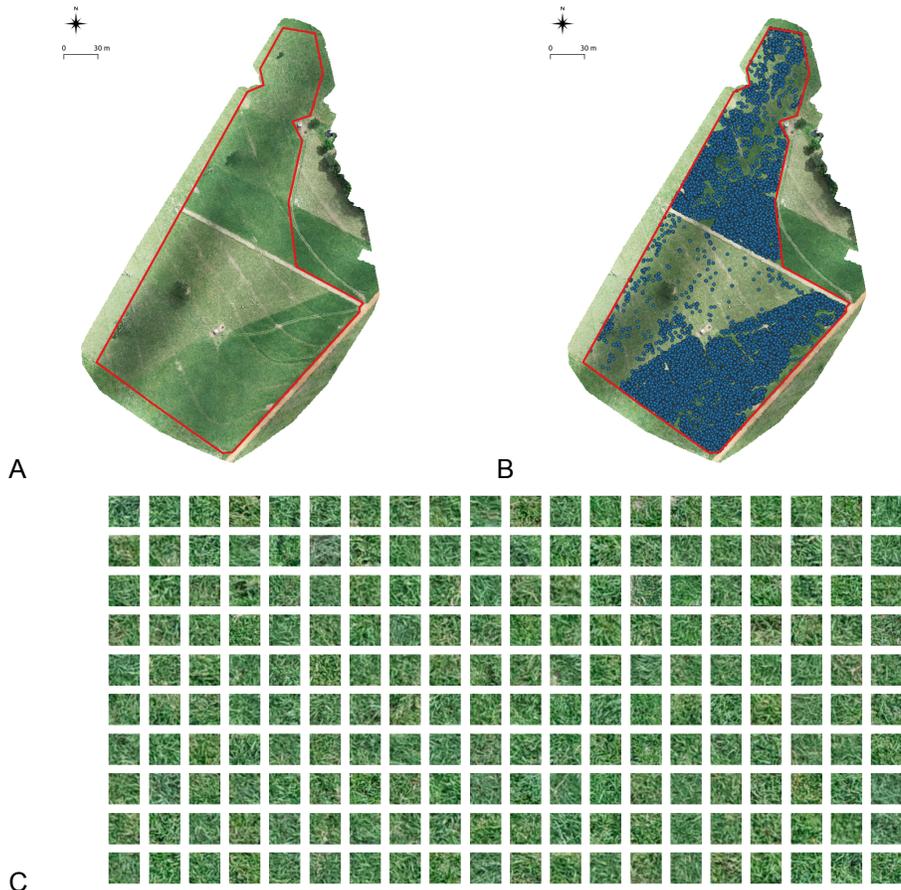
A Figura 6 exibe uma das classes encontradas para INT2018. Os pontos na Figura 6 (B) exibem a localização de todas as amostras agrupadas nessa classe, enquanto a Figura 6 (C) exibe um exemplo de 200 amostras pertencentes ao grupo. As amostras correspondem a piquetes ainda não pastejados, que aguardam a entrada do rebanho e apresentam maior quantidade de alimento disponível. A Figura 7 mostra outra classe encontrada para a mesma área e data, mas com interpretação oposta: corresponde a piquetes recém pastejados com menor disponibilidade de forragem. Nas Figuras 7 (B) e (C) é possível identificar amostras com pasto rente e solo exposto. Uma grande quantidade de solo exposto é identificado nas amostras da classe exibida na Figura 8, sendo que áreas próximas ao bebedouro e à divisão central (onde trafegam os veículos, entre os setores norte e sul da área), são visíveis na Figura 8 (B).

Na área de pecuária extensiva também podemos observar capacidade similar de representação. A Figura 9 exibe amostras que correspondem a setores com maior disponibilidade de forragem em EXT2018, enquanto a Figura 10 exibe amostras associadas a áreas com menor disponibilidade. Já a Figura 11 mostra uma classe identificada com anomalias diversas encontradas nas amostras: locais em que se encontram galhos de árvores, dorsos de animais ou artefatos produzidos no ortomapa pelas linhas de alta-tensão que atravessam a área da pastagem. Em termos qualitativos, é possível afirmar que o *backbone*, treinado de maneira totalmente não-supervisionada, encontra representações úteis à categorização de pequenas áreas da pastagem, inferiores a 1 m<sup>2</sup>, permitindo análises com grande nível de detalhe.

Contudo, na análise quantitativa fornecida pelo *linear probing*, os resultados são inconclusivos. Não foi possível alcançar convergência durante o treinamento da camada de regressão linear, utilizando os dados supervisionados, para nenhuma das áreas de estudo. Uma hipótese é que o número de amostras com aferição de biomassa é insuficiente para o treinamento de um modelo de aprendizado de máquina. Tal hipótese conflita com os resultados de Reis et al. (2020) e Freitas et al. (2022), que apresentaram certo sucesso na estimativa de biomassa com um número similar de amostras supervisionadas. Outra hipótese é a insuficiência do imageamento RGB (espectro visível) na estimativa da biomassa: amostras com aspecto visual similar poderiam apresentar valores distintos de biomassa (e vice-versa), possivelmente devido a estruturas não visíveis do dossel. Reis et al. (2020) e Freitas et al. (2022) empregaram informação fora

<sup>5</sup>Diversos valores de *K* foram testados, valores maiores produzem maior estratificação. O valor *K* = 7 classes é arbitrário, escolhido por produzir um número pequeno de classes discerníveis.

do espectro visível, sendo que Reis et al. (2020) demonstram que valores derivados de NDVI tiveram maior importância no processo de estimativa de biomassa.

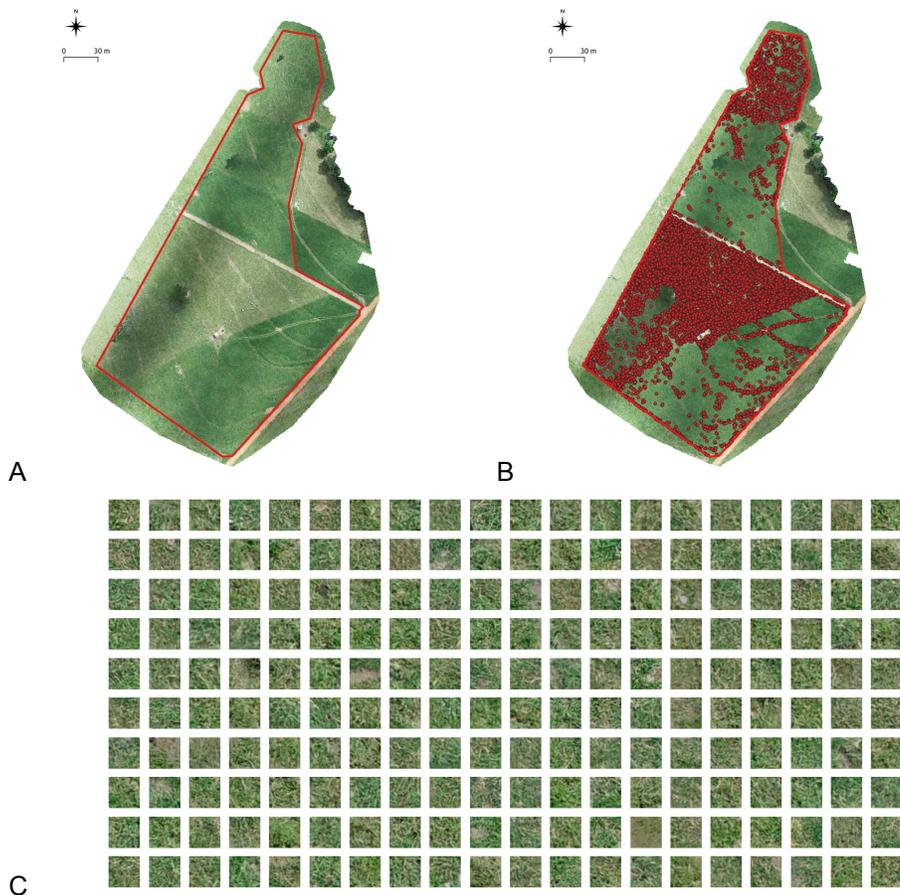


**Figura 6.** Identificação automática de classes: pastagem do sistema intensivo rotacionado (A); localização das amostras pertencentes a classe, correspondendo a piquetes ainda não pastejados (B); exemplos de amostras da classe (C).

Nota: uma das classes identificadas pelo algoritmo K-médias para INT2018 (0,5 m) contém amostras com menos presença de solo exposto e mais folhagem verde.

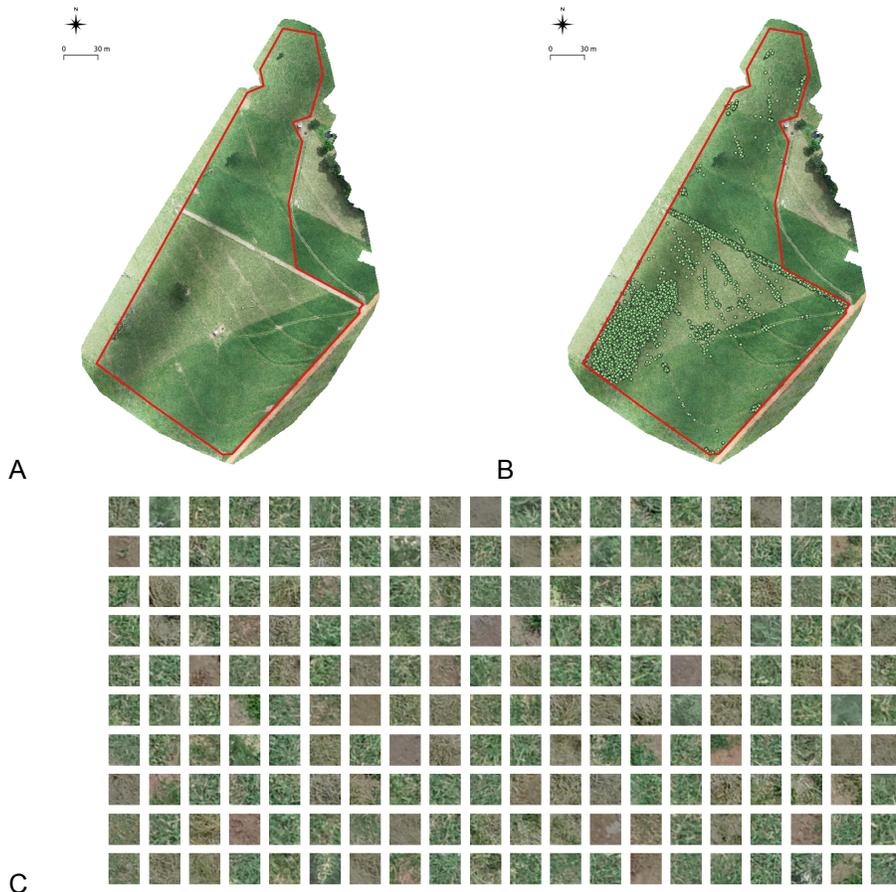
ARPs com câmeras multi-espectrais, capazes de obter imagens para as bandas NIR e Red Edge, além das bandas do espectro visível (RGB), estão se tornando largamente disponíveis. A metodologia apresentada neste boletim poderia ser estendida de maneira trivial, seja pela incorporação dessas bandas em tensores de entrada com 4, 5 ou mais canais, seja pela substituição de um canal RGB por um canal fora do espectro

visível. O aprendizado automático de representações é uma alternativa ao emprego de descritores de textura (Reis et al., 2020; Freitas et al., 2022), como demonstrado neste trabalho e também por Gldenring; Nalpantidis (2021).



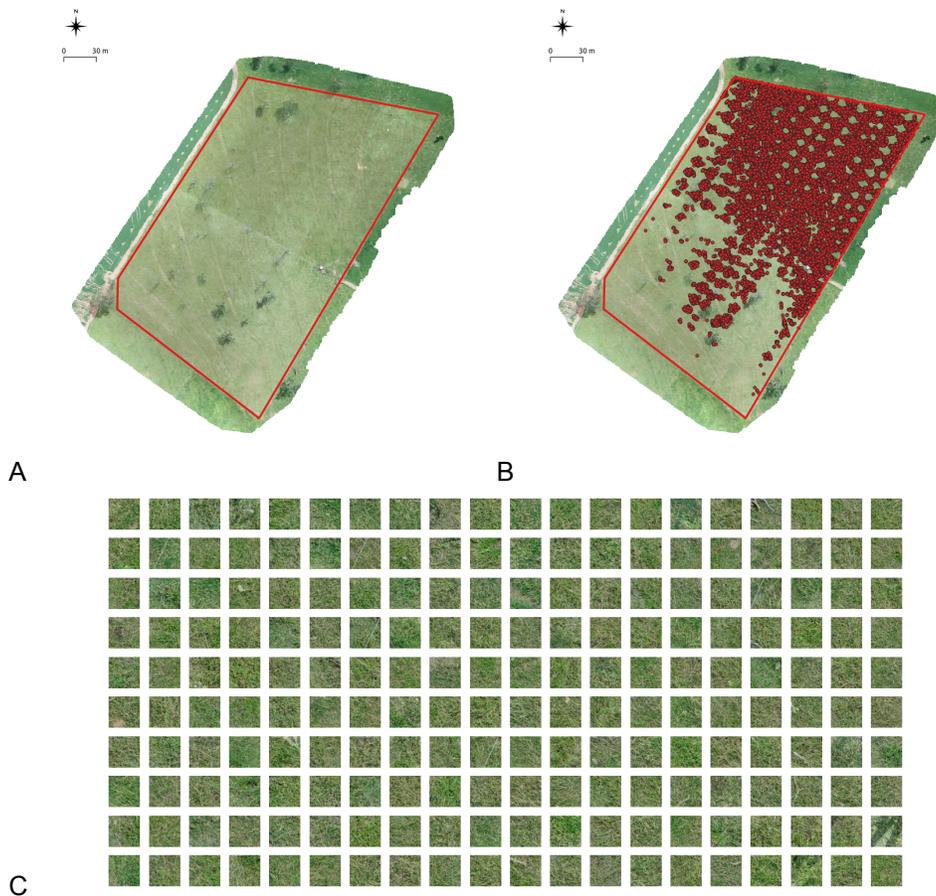
**Figura 7.** Identificao automtica de classes: pastagem do sistema intensivo rotacionado (A); localizao das amostras pertencentes a classe, correspondendo a piquetes j pastejados (B); exemplos de amostras da classe (C).

Nota: uma das classes identificadas pelo algoritmo  $K$ -mdias para INT2018 (0.5 m) contm amostras com menor presena de forragem.



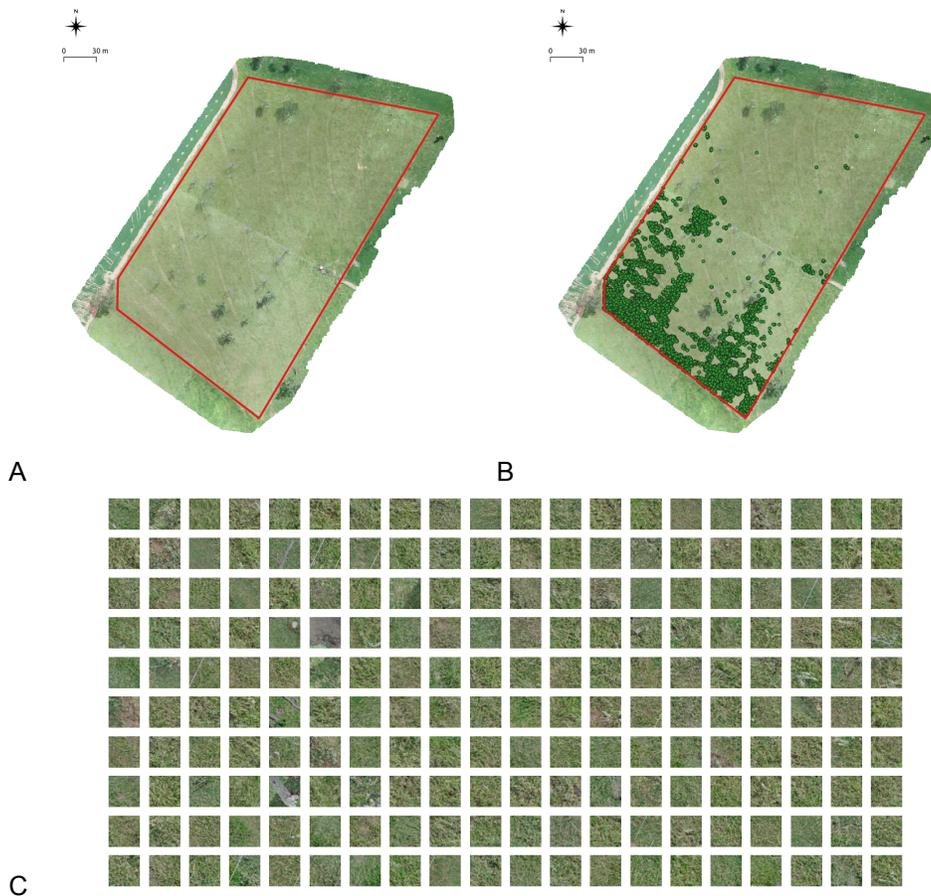
**Figura 8.** Identificação automática de classes: pastagem do sistema intensivo rotacionado (A); localização das amostras pertencentes a classe, concentradas em áreas de solo exposto como passeios e bebedouro (B); exemplos de amostras da classe (C).

Nota: uma das classes identificadas pelo algoritmo  $K$ -médias para INT2018 (0.5 m) apresenta grande quantidade de solo exposto.



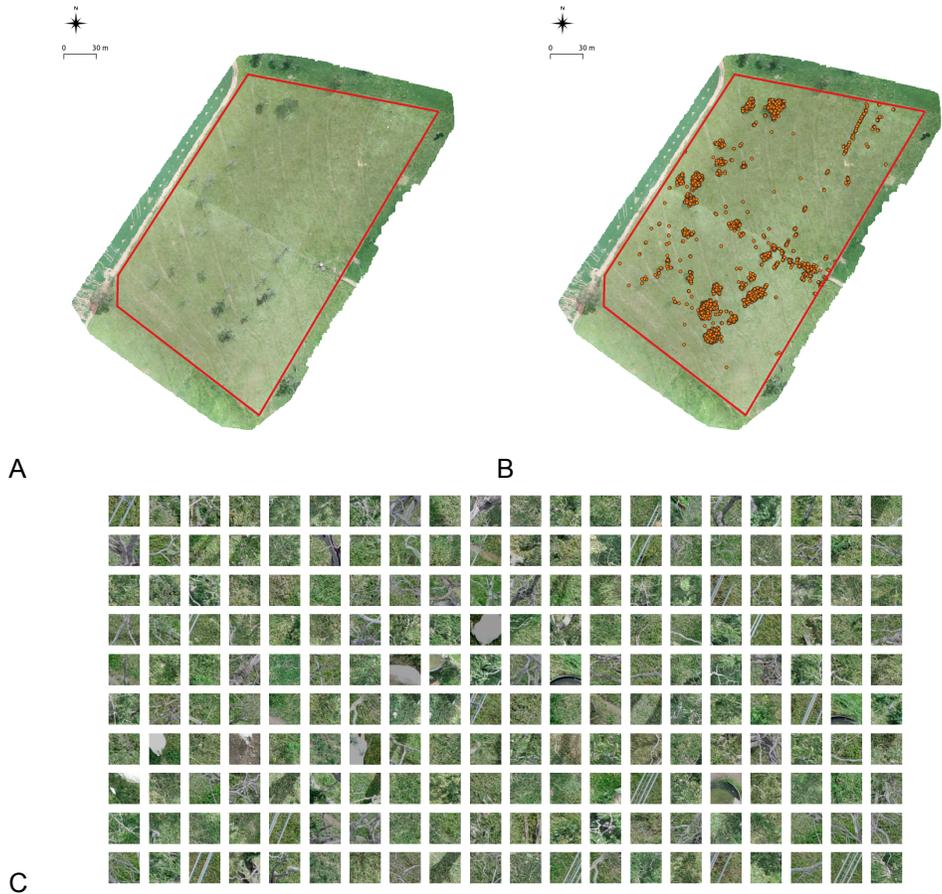
**Figura 9.** Identificação automática de classes: pastagem extensiva (2018) (A); localização das amostras pertencentes a classe (B); exemplos de amostras da classe (C).

Nota: uma das classes identificadas pelo algoritmo  $K$ -médias para EXT2018 (1.0 m) contém amostras com menos presença de solo exposto e mais folhagem verde.



**Figura 10.** Identificação automática de classes: pastagem extensiva (2018) (A); localização das amostras pertencentes a classe (B); exemplos de amostras da classe (C).

Nota: uma das classes identificadas pelo algoritmo  $K$ -médias para EXT2018 (1.0 m) contém amostras com maior presença de solo exposto e menos folhagem verde (ver Figura 9 para comparação).



**Figura 11.** Detecção de anomalias: pastagem extensiva (2018) (A); localização das amostras pertencentes a classe (B); exemplos de amostras da classe (C).

Nota: uma das classes identificadas pelo algoritmo K-médias para EXT2018 (1.0 m) contém amostras anômalas contendo galhos de árvores, fiação elétrica e dorso de animais no pasto.

## Conclusões

---

O pré-treinamento não-supervisionado é uma alternativa apropriada em aplicações de aprendizado de máquina na agropecuária, em que é comum a existência de abundância de dados brutos, mas um número reduzido de dados com verdade de campo agrônômica incorporada, decorrente das dificuldades de aferição no campo em larga escala. A análise de pastagens por imageamento aéreo (por ARPs) é um exemplo desse tipo de situação. Técnicas como MAE conseguem encontrar regularidades estatísticas em grandes bases de imagens e as representações aprendidas têm grande potencial de transferências para diversas tarefas de aplicação, como identificação de degradação e detecção de plantas invasoras. Porém, o uso restrito do espectro visível e imagens aéreas nadir pode ser insuficiente para tarefas como estimativa de biomassa de forragem: nesse caso, o uso de bandas associadas ao vigor de plantas e uma estimativa acurada de altura podem ser necessários. Uma hipótese alternativa é que o número de amostras necessárias para uma estimativa acurada de biomassa a partir do espectro visível seja maior do que o fornecido pela base de dados apresentada. Nesse caso, um esforço maior de coleta de amostras em campo seria necessário.

## Agradecimentos

---

Este trabalho foi desenvolvido dentro dos projetos *Pecuária do futuro* (SEG 22.16.05.-021.00.00) e *PRS – Cerrado*. O projeto *PRS – Cerrado* é financiado via Cooperação Técnica aprovada pelo Banco Interamericano de Desenvolvimento (BID), com recursos do Financiamento Internacional do Clima do Governo do Reino Unido, tendo o Ministério da Agricultura e Pecuária (Mapa) como beneficiário institucional. O Instituto Brasileiro de Desenvolvimento e Sustentabilidade (IABS) é o responsável pela execução e administração do projeto e a Associação Rede ILPF, por meio da Embrapa, pela coordenação científica e apoio técnico.

## Referências

---

ASANO, Y.; RUPPRECHT, C.; VEDALDI, A. A critical analysis of self-supervision, or what we can learn from a single image. In: INTERNATIONAL CONFERENCE ON LEARNING REPRESENTATIONS, 8., 2020, Addis Ababa. **Papers** [...]. 2020. ICLR 2020. Disponível em: <https://openreview.net/forum?id=B1esx6EYvr>. Acesso em: 25 out. 2023.

BAYMA-SILVA, G.; GREGO, C. R.; KOENIGKAN, L. V.; NOGUEIRA, S. F.; PEZZOPANE, J. R. M.; SANTOS, R. C. d.; SANTOS, P. M.; SANTOS, T. T.; TAKEMURA, C. M. **Protocolo de campo para investigação, calibração e validação de métodos para estimativa de massa de forragem baseados em sensoriamento remoto orbital e proximal**. Campinas: Embrapa Informática Agropecuária, 2019. 29 p. il. (Embrapa Informática Agropecuária. Comunicado técnico, 133). Disponível em: <https://www.infoteca.cnptia.embrapa.br/infoteca/handle/doc/1117053>. Acesso em: 25 out. 2023

CARON, M.; BOJANOWSKI, P.; JOULIN, A.; DOUZE, M. Deep clustering for unsupervised learning of visual features. In: EUROPEAN CONFERENCE ON COMPUTER VISION, 15., 2018, Munich. **Proceedings** [...]. Cham: Springer, 2018. pcaronart XIV, p 139-156. Editores: Vittorio Ferrari, Martial Hebert, Cristian Sminchisescu, Yair Weiss. (Lecture notes in computer science, 11218). EECV 2018. DOI: 10.1007/978-3-030-01264-9\_9.

DIAS-FILHO, M. B. **Diagnóstico das pastagens no Brasil**. Belém, PA: Embrapa Amazônia Oriental, 2014. 36 p. (Embrapa Amazônia Oriental. Documentos, 402). Disponível em: <https://www.infoteca.cnptia.embrapa.br/infoteca/handle/doc/986147>. Acesso em: 25 out. 2023.

DOSOVITSKIY, A.; BEYER, L.; KOLESNIKOV, A.; WEISSENBORN, D.; ZHAI, X.; UNTERTHINER, T.; DEGHANI, M.; MINDERER, M.; HEIGOLD, G.; GELLY, S.; USZKOREIT, J.; HOULSBY, N. An image is worth 16x16 words: transformers for image recognition at scale. In: INTERNATIONAL CONFERENCE ON LEARNING REPRESENTATIONS, 9., 2021, Vienna. **Oral presentations**. 2021. ICLR 2021. Disponível em: <https://openreview.net/forum?id=YicbFdNTTy>. Acesso em: 25 out. 2023.

FREITAS, R. G.; PEREIRA, F. R. S.; REIS, A. A. dos; MAGALHÃES, P. S. G.; FIGUEIREDO, G. K. D. A.; AMARAL, L. R. do. Estimating pasture aboveground biomass under na integrated crop-livestock system based on spectral and texture measures derived from UAV images. **Computers and Electronics in Agriculture**, v. 198, 107122, July 2022. DOI: 10.1016/j.compag.2022.107122.

GÜLDENRING, R.; NALPANTIDIS, L. Self-supervised contrastive learning on agricultural images. **Computers and Electronics in Agriculture**, v. 191, 106510, Dec. 2021. DOI: 10.1016/j.compag.2021.106510.

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. **The elements of statistical learning: data mining, inference, and prediction**. 2. ed. New York: Springer, 2009. 745 p. (Springer series in statistics).

HE, K.; CHEN, X.; XIE, S.; LI, Y.; DOLLÁR, P.; GIRSHICK, R. Masked autoencoders are scalable vision learners. In: IEEE/CVF CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION, 2022, New Orleans. **Proceedings [...]**. Los Alamitos: IEEE, 2022. p. 15979–15988. CVPR 2022. DOI: 10.1109/CVPR52688.2022.01553.

HE, K.; FAN, H.; WU, Y.; XIE, S.; GIRSHICK, R. Momentum contrast for unsupervised visual representation learning. In: IEEE/CVF CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION, 2020, Seattle. **Proceedings [...]**. Los Alamitos: IEEE, 2020. p. 9726–9735. CVPR 2020. DOI: 10.1109/CVPR42600.2020.00975.

KIRILLOV, A.; MINTUN, E.; RAVI, N.; MAO, H.; ROLLAND, C.; GUSTAFSON, L.; XIAO, T.; WHITEHEAD, S.; BERG, A. C.; LO, W. Y.; DOLLÁR, P.; GIRSHICK, R. **Segment anything**. 2023. DOI: 10.48550/arXiv.2304.02643.

KONG, L.; MA, M. Q.; CHEN, G.; XING, E. P.; CHI, Y.; MORENCY, L. P.; ZHANG, K. Understanding masked autoencoders via hierarchical latent variable models. In: IEEE/CVF CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION, 2023, Vancouver. **Proceedings [...]**. Los Alamitos: IEEE, 2023. p. 7918-7928. CVPR 2023. DOI: 10.1109/CVPR52729.2023.00765.

LIU, X.; ZHANG, F.; HOU, Z.; MIAN, L.; WANG, Z.; ZHANG, J.; TANG, J. Selfsupervised learning: Generative or contrastive. **IEEE Transactions on Knowledge and Data Engineering**, v. 35, n. 1, p. 857–876, 2023. DOI: 10.1109/TKDE.2021.3090866.

OLIVEIRA, G. S. de; MARCATO JUNIOR, J.; POLIDORO, C.; OSCO, L. P.; SIQUEIRA, H.; RODRIGUES, L.; JANK, L.; BARRIOS, S.; VALLE, C.; SIMEÃO, R.; CARROMEU, C.; SILVEIRA, E.; JORGE, L. A. de C.; GONÇALVES, W.; SANTOS, M.; MATSUBARA, E. Convolutional neural networks to estimate dry matter yield in a guineagrass breeding program using UAV remote sensing. **Sensors**, v. 21, n. 12, June 2021. DOI: 10.3390/s21123971.

OPENMMLAB. **MMSelfSup**: OpenMMLab self-supervised learning toolbox and benchmark. 2021. Disponível em: <https://github.com/open-mmlab/mmselfsup>. Acesso em: 25 out. 2023.

REINERMANN, S.; ASAM, S.; KUENZER, C. Remote sensing of grassland production and management—a review. **Remote Sensing**, v. 12, n. 12, 1949, 2020. DOI: doi.org/10.3390/rs12121949.

REIS, A. A. dos; WERNER, J. P. S.; SILVA, B. C.; FIGUEIREDO, G. K. D. A.; ANTUNES, J. F. G.; ESQUERDO, J. C. D. M.; COUTINHO, A. C.; LAMPARELLI, R. A. C.; ROCHA, J. V.; MAGALHÃES, P. S. Monitoring pasture aboveground biomass and canopy height in an integrated crop-livestock system using textural information from planetscope imagery. **Remote Sensing**, v. 12, n. 16, 2534, Aug. 2020. DOI: 10.3390/rs12162534.

SANTOS, T. T.; KOENIGKAN, L. V. **Produção de ortomapas com VANTs e OpenDroneMap**. Campinas: Embrapa Informática Agropecuária, 2018. 21 p. (Embrapa Informática Agropecuária. Circular técnica, 05). Disponível em: <https://www.infoteca.cnptia.embrapa.br/infoteca/handle/doc/1102033>. Acesso em: 25 out. 2023.



---

*Agricultura Digital*