

**FEW-SHOT LEARNING PARA ESCORE DE CONDIÇÃO CORPORAL EM
BOVINOS: IMPACTO DO PRÉ-TREINAMENTO COM MASKED
AUTOENCODERS E DINOv2 SOBRE UM VISION TRANSFORMER**

Max Hiroito Tieti (ra868222@ucdb.br)

Roberto Pereira De Freitas Neto (neto2014roberto@gmail.com)

Allan Rios Bezerra (allanrb07@gmail.com)

A avaliação do Escore de Condição Corporal (ECC) é uma ferramenta importante para o manejo nutricional e reprodutivo de rebanhos bovinos, mas sua automatização em larga escala é dificultada pela escassez de dados rotulados e pelo custo da anotação especializada. Este trabalho investiga o uso de pré-treinamento auto-supervisionado como estratégia para mitigar essa limitação em um cenário de aprendizado com poucos exemplos (few-shot learning). Foram comparadas três estratégias de inicialização de um backbone Vision Transformer (ViT-Small): (i) uma linha de base sem pré-treinamento (inicialização aleatória), (ii) pré-treinamento via Masked Autoencoders (MAE) sobre um conjunto de aproximadamente 50 mil imagens não rotuladas de bovinos, e (iii) um modelo DINOv2 pré-treinado em ImageNet, com ajuste fino parcial dos dois últimos blocos do transformer. As estratégias de linha de base e MAE foram avaliadas por meio de uma sonda linear (linear probe) sobre o backbone congelado, enquanto o DINOv2 foi avaliado com ajuste fino parcial e taxa de aprendizado diferenciada. Cada configuração foi treinada com $N = \{20, 30, 50, 100\}$ exemplos rotulados, em cinco repetições com diferentes sementes

aleatórias, sobre um esquema de classificação de ECC em 5 classes (escala inteira de 1 a 5). Os resultados aqui apresentados são preliminares e correspondem a uma rodada inicial de experimentos, mas já indicam tendências consistentes entre as estratégias avaliadas. O modelo DINOv2 obteve desempenho consistentemente superior e crescente com o número de exemplos, atingindo acurácia de 62,8% ($\pm 3,6$ p.p.) com $N=100$, frente a 29,8% ($\pm 3,0$ p.p.) do MAE e 24,5% ($\pm 2,7$ p.p.) da linha de base. O pré-treinamento MAE produziu uma melhora modesta, porém consistente, em relação a características aleatórias (de +2,3 a +5,3 pontos percentuais de acurácia ao longo dos valores de N), sugerindo que o pré-treinamento auto-supervisionado em imagens de bovinos captura informação útil, ainda que limitada, para a tarefa de ECC. Observou-se ainda que as estratégias com backbone congelado (linha de base e MAE) apresentam ganhos marginais com o aumento do número de exemplos rotulados, enquanto o DINOv2, com ajuste fino parcial, continua a melhorar de forma expressiva — sugerindo que a adaptação das representações ao domínio bovino é tão ou mais relevante do que a quantidade de exemplos rotulados disponíveis. Em conjunto, os resultados preliminares reforçam o potencial de modelos de fundação pré-treinados em larga escala (DINOv2) para tarefas pecuárias com poucos dados, ao mesmo tempo em que indicam que o pré-treinamento auto-supervisionado específico de domínio (MAE), nas condições avaliadas, oferece ganhos modestos frente à inicialização aleatória. Os resultados indicam que, nas condições avaliadas, a escala do pré-treinamento importa mais do que sua especificidade de domínio — conclusão com implicações práticas para o custo de implantação de sistemas de avaliação automatizada de ECC.

Palavras-chave: escore de condição corporal; aprendizado auto-supervisionado; few-shot learning; vision transformer; masked autoencoders; dinov2; pecuária de precisão.