Method Applied to Animal Monitoring Through VANT Images

B. Vasconcellos, J. Trindade, L. Volk, and L. Pinho

***Abstract*—One of the necessary demands in extensive livestock systems is the counting of animals in areas of tens of hectares, costly when carried out manually and locally. In this context, this work proposes and discusses the efficacy of a semi-autonomous, non-invasive method for remote identification of animals in the field, applicable to precision livestock systems. The method was conceived from an exploratory research methodology based on remote sensing techniques that include image collection processes by aerial surveying with RGB camera embedded in unmanned aerial vehicle, persistence of images obtained by means of storage in space-time databases and processing of stored images for the construction of a rural property orthomosaic succeeded by the application of patterns discovery processes, making use of deep learning, especially convolutional neural networks. According to the experiments carried out, the method was effective, being able to identify and count animals from the collection of images made at 100 m height, with an accuracy of up to 95%, including the approximate geographical position of the animals to field.**

***Index Terms*—Cattle herding, Deep learning, Convolutional neural network.**

I. Introdução

A computação aplicada se destaca nos mais diversos sistemas produtivos, inclusive tendo papel cada vez mais fundamental para o aumento da eficiência na agropecuária. Sobretudo na pecuária extensiva, a observação humana das diferentes variáveis que compõem este complexo sistema demanda tempo significativo, dependendo das características da propriedade (principalmente a extensão) e dos animais (em especial a quantidade/densidade). Neste contexto, princípios de Pecuária de Precisão – "gestão da produção que, a partir do uso de conhecimentos e tecnologias variadas, busca entender e intervir na variação que existe, tanto no ambiente de produção, como no rebanho" [1] – são implantados buscando melhorar processos anteriormente considerados como homogêneos, em relação aos diversos estados existentes na produção de animais.

Entre as tecnologias disponíveis, com potencial para auxiliar pecuaristas na implantação deste princípios, destacam-se os Veículos Aéreos Não Tripulados (VANT), capazes de sobrevoar áreas de interesse, transportando diversos tipos de sensores, de forma semiautônoma. Em especial, VANT pode tornar mais eficiente o controle do estoque de animais, o qual se caracteriza como uma das atividades mais importantes, tendo em vista a necessidade de realizar, de acordo com o caso, inventários de animais com frequência diária ou semanal, sendo as variações no estoque decorrentes de morte, nascimento, abigeato (furto ou roubo de animais) e transferência entre produtores, por meio de venda [2]. Assim sendo, mostra-se relevante desenvolver métodos capazes de obter, de forma não invasiva, a contagem e posicionamento dos animais no campo com uma acurácia e desempenho suficientes para esta e outras aplicações, em especial para estudos que permitam uma melhor compreensão sobre o comportamento do rebanho.

Diante desse problema, o presente trabalho[1] propõe e avalia um método, baseado em aprendizado de máquina, capaz de monitorar remotamente animais e realizar, de forma eficaz, a contagem autônoma do rebanho a campo, a partir de imagens *Red Green Blue* (RGB) coletadas por meio de VANT. Posteriormente processadas com a aplicação de uma técnica de visão computacional em que o reconhecimento de padrões se dá por aprendizagem profunda ou Convolutional Neural Network (CNN), denominada *Faster Region-based Convolutional Neural Network* (*Faster R-CNN*).

II. Trabalhos Relacionados

Buscando identificar trabalhos relacionados publicados em periódicos científicos, conduziu-se uma revisão sistemática da literatura com os termos *animal monitoring*, *unmanned aerial vehicle*, *livestock* e *deep learning* por meio da ferramenta Google Scholar, tendo como critério de exclusão artigos publicados antes de 2016, resultando em três artigos relevantes de 2019. De forma complementar, buscou-se outros trabalhos envolvendo a aplicação de CNN no monitoramento de animais, com destaque para os dois trabalhos abordados na sequência.

Barbedo *et al.* [4] realizaram uma pesquisa utilizando aprendizagem profunda com CNN e monitoramento de bovinos com VANT em São Carlos, Brasil. Nesta pesquisa, os objetivos foram três: (1) determinar a maior precisão possível que poderia ser alcançada na detecção de animais da raça Canchim, visualmente semelhante à raça Nelore (Bos taurus indicus); (2) determinar a distância ideal da amostra do solo (Ground Sample Distance - GSD) para a detecção de animais; (3) para determinar a arquitetura CNN mais precisa para esse problema específico. Os experimentos envolveram 1853 imagens contendo 8629 amostras de animais e 15 diferentes arquiteturas CNN. Foram treinados 900 modelos, permitindo uma análise de vários aspectos que impactam a detecção de bovinos, usando imagens aéreas capturadas com VANT. Os resultados revelaram que muitas arquiteturas CNN são robustas o suficiente para detectar animais de forma confiável em

B. C. Vasconcellos, Universidade Federal do Pampa, Bagé, Rio Grande do Sul, Brasil, brunoc.vasconcellos@gmail.com.

J. P. P. Trindade, Embrapa, Bagé, Rio Grande do Sul, Brasil, jose.pereira-trindade@embrapa.br.

L. B. S. Volk, Embrapa, Bagé, Rio Grande do Sul, Brasil, leandro.volk@embrapa.br.

L. B. Pinho, Universidade Federal do Pampa, Bagé, Rio Grande do Sul, Brasil, leonardo.pinho@unipampa.edu.br.

[1]Mais detalhes podem ser observados em [3].

imagens aéreas, mesmo em condições não ideais, indicando a viabilidade do uso de VANT para monitoramento de gado.

Wang *et al.* [5] fizeram pesquisas sobre animais selvagens com base em múltiplas plataformas, incluindo satélites, aeronaves tripuladas e VANT, avaliando métodos de detecção de animais e suas precisões. Foram discutidas vantagens e limitações de cada tipo de dados de sensoriamento remoto. As imagens espaciais com resolução muito alta têm potencial para modelar a dinâmica populacional de animais selvagens grandes ($>$ 0,6 m) em grandes escalas espaciais e temporais, mas têm dificuldade em discernir animais pequenos ($<$ 0,6 m) no nível de espécies, embora os satélites comerciais de alta resolução, como o *WorldView*, tenham conseguido coletar imagens com uma resolução no solo de até 0,31 m no modo pancromático. Essa situação não será alterada, a menos que a resolução da imagem de satélite melhore bastante no futuro. Pesquisas aéreas tripuladas são utilizadas há muito tempo para capturar as imagens em escala de centímetros necessárias para censos de animais em grandes áreas. No entanto, esses levantamentos aéreos são caros para serem implementados em pequenas áreas e podem causar distúrbios significativos aos animais selvagens devido ao seu ruído. Por outro lado, os levantamentos com VANT são vistos como uma alternativa segura, conveniente e menos dispendiosa aos levantamentos aéreos tripulados terrestres e convencionais, mas a maioria dos VANT pode cobrir apenas pequenas áreas. Destacam também que o desenvolvimento de sistemas de software para produzir automaticamente mosaicos de imagens e reconhecer animais selvagens melhorará ainda mais a eficiência da pesquisa.

Vayssade *et al.* [6] propuseram um método para processar imagens capturadas por um VANT, a fim de automatizar a detecção e rastreamento de atividades com animais. Este método detecta automaticamente cabras em imagens e rastreia sua atividade usando uma combinação de limiar e métodos de classificação supervisionados. Foram testadas 571 imagens de VANT realizadas ao longo de 11 dias, com acurácia de 74% para detecção de animais e 78,3% para detecção de atividade.

Andrew, Greatwood e Burghardt [7] usaram redes neurais convolucionais para reconhecer de forma individual, através de imagens, a raça Holstein Friesian do gado a campo. Inicialmente, foi fixada uma câmera a 2 m de distância acima dos animais em uma área restrita onde os animais passavam pelo espaço e registraram 900 imagens. Após treinar o classificador, foram capturados vídeos aéreos a 15 m de altura, por meio de VANT a campo para utilizar como validação dos testes. Os autores concluíram que para a identificação em particular, as arquiteturas baseadas em convoluções são significativamente adequadas para aprender e distinguir as propriedades de padrão e estrutura dorsal únicas exibidas individualmente pelas espécies. Importante ressaltar que este processo pode ocorrer de forma não invasiva, em contraste com a maioria das estruturas de identificação existentes girando em torno de um equipamento instalado no animal.

Rey [8] buscou identificar a quantidade de animais em savanas semiáridas, justificando as ameaças recebidas pelas mudanças no equilíbrio frágil entre chuvas, incêndios e pressão de pastagem exercida por animais selvagens ou gado. Para evitar a invasão dos arbustos e o declínio da grama perene, os administradores de terras devem prestar atenção para manter a quantidade de gado e vida selvagem em equilíbrio com a oferta de alimento. Em grandes fazendas e parques de conservação, estimar as populações de animais é, portanto, um importante aspecto de gestão. Métodos tradicionais de recenseamento animal, tais como contagens usando helicóptero, ou marca / recaptura, são muito caras e trabalhosas para serem conduzidas regularmente. O autor destacou que VANT aparecem como uma ferramenta para a detecção de animais. Eles podem ser facilmente implantados, por um custo menor e maior segurança. O aspecto negativo é interpretar visualmente o grande número de imagens de alta resolução que elas adquirem. Os recentes avanços nas técnicas de aprendizado de máquina podem permitir a automação da detecção de animais nessas imagens aéreas. Neste contexto, este autor buscou implementar diferentes algoritmos, a fim de investigar a viabilidade e os benefícios potenciais da combinação de aprendizado de máquina e VANT para a detecção de animais. As técnicas de aprendizado de máquina envolvidas foram Bags of visual Words (BOW), Support-vector machine (SVM) e aprendizado ativo. Os resultados foram promissores mostram que taxas de recuperação na faixa de 60 a 80% são possíveis, se uma baixa precisão (5 a 20%) for aceita.

O uso de CNN torna o método proposto neste trabalho similar aos autores Barbedo *et al.* [4], contudo, este trabalho já confirma a hipótese do uso de CNN na detecção de animais com VANT, diferenciando-se por apresentar detecção independente de raça e posicionamento em pé ou deitado, apresentando, ainda, a contagem de animais com *bouding boxes* e posicionamento geográfico. Wang *et al.* [5] confirmam o uso de detecção de animais com VANT em pequenas áreas, e ainda reforçam o uso de CNN no uso de imagens capturados por VANT e a detecção de objetos (animais). Visto que é uma pesquisa de revisão, não demonstra aplicações práticas. Vayssade *et al.* [6] utilizaram técnicas sem aprendizado de máquina, tornando o processo mais manual, e suas acurácia fica em torno de 70%. Abaixo dos resultados apresentados neste trabalho. Andrew, Greatwood e Burghardt [7] usaram câmera fixa para captação das imagens, o que torna o processo mais restrito a uma área específica de animais. Estes autores também utilizaram apenas uma raça específica. Portanto, a pesquisa deste trabalho diferencia-se por usar apenas VANT, independente de raça, e ainda com altitude de 100 m, que aumenta a abrangência do monitoramento. Rey [8] usou técnicas de aprendizado de máquina mais simples como BOW e SVM, que são técnicas que necessitam mais pré-processamento manual, e ainda assumindo um resultado final com menos acurácia.

O método proposto no trabalho atual demonstra o uso de redes neurais convolucionais, uma técnica dentro do estado da arte em detecção de objetos em imagens, que mesmo a 100 m de altitude (aumento na área de abrangência em relação a altitudes menores) com um VANT, foi capaz de aplicar filtros de características de bordas, sombra, relevo e destaque de objetos durante o processamento e detecção de objetos. Além disso, cabe destacar que está apto ao uso em GPU e ambientes de *Cloud Computing*, os quais fornecem recursos computacionais significativos para métodos que envolvem processamento de imagens [9].

III. Referencial Teórico

O objetivo da visão computacional é extrair informação útil das imagens, ajudando a tomar decisões sobre objetos físicos e cenas através de imagens. Para tomar decisões sobre objetos reais, é quase sempre necessário construir alguma descrição ou modelo deles a partir das imagens. Para construir um modelo de treinamento, é preciso ensinar a máquina com outras imagens pré-processadas que contenham os padrões desejados. Para tal tarefa existe uma área da inteligência artificial chamada aprendizado de máquina, a qual se divide em supervisionada e não supervisionada. Na supervisionada, o processo de aprendizagem de um conjunto de regras ocorre a partir de instâncias ou exemplos de um conjunto de treinamento com a intervenção humana, enquanto que, na não supervisionada, algoritmos, depois de treinados com dados de entrada, tentam criar clusters (agrupamentos) de forma automática – sem intervenção humana – classificando dados de acordo com o aprendizado. Também pode ser definido como a criação de um classificador, que pode ser generalizado para novas instâncias, não presentes no conjunto de treinamento. Um classificador é um modelo que, após treinado, pode ser utilizado para associar classes a instâncias de testes, cujas classes são desconhecidas, utilizando a informação dos seus atributos [10].

Existem diversas técnicas de aprendizado de máquina para reconhecimento de padrões: algumas com necessidades de realização de um pré-processamento maior, outras com características específicas para voz e imagens, com custo diferentes de processamento computacional. Em especial, o aprendizado profundo é definido como um conjunto de técnicas de aprendizado de máquina que exploram grandes quantidades de camadas de processamento de informação não linear para extração e transformação, supervisionada ou não supervisionada, para análise de padrões e classificação. A sua utilização vem crescendo progressivamente, principalmente no reconhecimento de objetos em imagens, que ficou dependente de técnicas como *Scale-Invariant Feature Transform* (SIFT) e *Histogram of Oriented Gradients* (HOG). No entanto, estas técnicas têm dificuldade de obter maior nível de informação nas imagens como detecção de bordas e fragmentos. A partir deste problema, o aprendizado profundo, visa superar as dificuldades obtendo um nível de detalhamento maior dos dados da imagem [11]. Dentro deste contexto, destacam-se as CNN.

As CNN são arquiteturas biologicamente inspiradas capazes de serem treinadas e aprenderem representações invariantes à escala, translação, rotação e transformações afins. As CNN são projetadas para uso com dados em duas dimensões tornando-as uma boa candidata para a solução de problemas envolvendo reconhecimento de imagens. Por definição, uma arquitetura profunda é uma estrutura hierárquica de múltiplas etapas, onde cada etapa é formada por uma rede neuronal de, pelo menos, três camadas, e cada etapa é treinada pelo algoritmo *Backpropagation*. Com as fontes em larga escala de dados de treinamento e implementação eficiente em *Graphics Processing Unit* (GPU), as CNN recentemente superaram alguns outros métodos convencionais, até mesmo o desempenho humano, em muitas tarefas relacionadas à visão, incluindo classificação de imagens, detecção de objetos, rotulagem de cena e reconhecimento facial [12]. Na arquitetura, a própria rede teria que detectar as dependências existentes na estrutura espacial da distribuição subjacente às imagens de entrada. Além disso, devido à conectividade muito grande, esse tipo de arquitetura sofre da denominada Maldição da Dimensionalidade (*Curse of Dimensionality*) e, portanto, não é adequado a imagens de alta resolução, por conta do alto potencial de sobreajuste aos dados. Isso sem considerar o tempo para computar as pré-ativações de todas as unidades em cada camada [13]. No contexto de arquiteturas de extração de características de imagens, CNN são pré-treinadas com um conjunto de dados de imagens (banco de dados ImageNet) para descritores de imagem genérica distintos e podem ser aplicados para extrair características discriminativas de imagens baseadas na teoria de *Transfer Learning* - neste tipo de CNN, ”treinada em uma grande imagem da natureza conjunto de dados antes de ser usado como um extrator de recurso em um pequeno conjunto de dados”. As características extraídas de CNN pré-treinadas são genéricas e aplicáveis a outros conjuntos de dados [14]. Destacaram-se na revisão da literatura duas arquiteturas: *Inception* e *Inception-ResNet-V2*. A primeira é capaz de processar recursos espaciais mais ricos e aumentar a diversidade de recursos. Já a segunda arquitetura une o potencial da Inception com a capacidade maior de reconhecimento de objetos, melhorando acurácia. As duas suportam *Transfer Learning* [15]. Dentro do atual estado da arte da localização e detecção de objetos em imagens estão os algoritmos *Faster R-CNN* e *Single Shot MultiBox Detector* (SSD).

O *Faster R-CNN* recebe como entrada uma imagem inteira e um conjunto de objetos propostos. A rede primeiro processa a imagem inteira com várias camadas convolucionais e de pool máximo para produzir um mapa de atributos. Então, para cada objeto, uma camada de agrupamento da *Region of Interest* (ROI) extrai um vetor de características de comprimento fixo do mapa de atributos. Cada vetor de características é alimentado em uma sequência de camadas totalmente conectadas, que, finalmente se ramificam para duas camadas de saída: uma que produz estimativas de probabilidade de softmax em classes de objetos, mais uma classe de “fundo”, e, por fim, a camada que produz quatro números de valor real para cada uma das classes de objetos. Cada conjunto de quatro valores codifica posições refinadas de caixa (*bounding box*) delimitadora para cada uma das classes. No caso do algoritmo SSD, a abordagem é baseada em uma rede convolutiva *feed-forward* que produz uma coleção de tamanho fixo de *bounding boxes* e seus escores para a presença de classe de objeto nessas caixas. Após, ocorre uma supressão das camadas para produzir as detecções finais. As primeiras camadas de rede são baseadas em uma arquitetura padrão usada para classificação de imagem de alta qualidade (truncada antes de qualquer camada de classificação). Em seguida, é adicionada a estrutura auxiliar à rede para produzir detecções com os mapas de características de escala múltipla para detecção de objeto. Essas camadas diminuem gradualmente de tamanho e permitem previsões de detecções em escalas múltiplas. O modelo convolucional para prever detecções é diferente para cada camada de recurso que

opera em um único mapa de atributos de escala [16].

Em uma pesquisa neste contexto, as CNN foram aplicadas no reconhecimento em imagens de marcas em gado com imagens em escalas de cinza. Os autores atingiram acurácia de 93%, no entanto concluíram a necessidade de usar uma quantidade maior de imagens [17]. O algoritmo Faster R-CNN foi usado para classificar padrões de imagens de humanos em estado correndo, caminhando e correndo muito rápido a partir de vídeos feitos por smartphone com resolução 1920x1080p [18]. O dataset foi montado convertendo seus *bounding boxes* em XML para treinamento. O autor conseguiu 3000 imagens randomicamente para treinamento, a partir de cortes de tamanhos de 200 a 400 px quadrados, e obteve uma acurácia de 98,9% nas classes correndo e caminhando e correndo muito rápido 97,7%. O algoritmo SSD, com imagens coletadas por meio de VANT, foi usado para identificar isoladores em linhas de transmissão de energia [19]. Foi coletado um banco de dados de 6700 imagens fazendo cortes de tamanhos de 150 a 400 px quadrados, sendo destacado que a quantidade não foi suficiente para obter bons resultados. Obteve-se acurácia de 93,75% e 85,29% nas duas classificações de linhas previstas. Com estas variações de algoritmos percebe-se a necessidade de uma quantidade de imagens acima de 3000 exemplos para treinamento. Além disso, identificou-se que Faster R-CNN tente a produzir melhores resultados com tamanhos de imagens diferentes e quantidades menores de imagens.

IV. Método Proposto

Tendo como base as "tecnologias de apoio ao manejo extensivo de rebanhos em sistemas de produção baseados em conceitos de Pecuária de Precisão", o método de identificação de animais a campo desenvolvido neste trabalho, conforme ilustra a Fig. 1, possui quatro etapas.

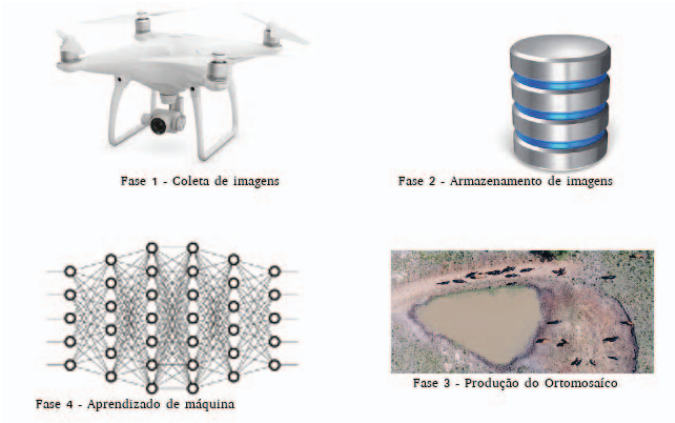


Fig. 1. Passo a passo do método proposto.

Na primeira etapa ocorre o processo de coleta por aerolevantamento, o qual envolve a definição de um plano de voo em uma área que contenha animais e seja acessível por um VANT. A coleta de imagens do espectro visível – parcela do espectro eletromagnético na qual a radiação é composta por fótons capazes de sensibilizar o olho humano – é realizada por meio de uma câmera digital padrão RGB embarcada no VANT, posicionada de forma perpendicular ao veículo.

Na segunda etapa, dá-se o processo de persistência, por meio do qual a sequência de imagens coletadas é armazenada em um sistema gerenciador de banco de dados espaço-temporal a fim de manter a temporalidade e localização das imagens para análises futuras e extração de padrões.

A terceira etapa do processo inicia pelo acesso às imagens armazenadas na base de dados espaço-temporal com metadados associados produzindo-se um ortomosaico georreferenciado, tendo como objetivo aumentar a eficiência a partir da minimização da duplicidade na contagem causada pela potencial movimentação dos animais durante o aerolevantamento, para tanto defini-se uma área de interesse delimitada por um polígono, incluindo a aplicação do algoritmo SIFT sobre o conjunto de imagens parcialmente sobrepostas. A partir de anotações de polígonos de animais no ortomosaico, por meio da definição de Regiões de Interesse (Regions of Interest - ROI), são rotulados animais, por classe, suficientes para a primeira etapa de treinamento do classificador baseado em técnica supervisionada.

Como quarta e última etapa do processo, as imagens rotuladas na etapa anterior são submetidas a um classificador pré-treinado, que tem como objetivo melhorar a performance tendo como referência a métrica tempo de treinamento. A partir do resultado do classificador, as imagens não rotuladas na etapa três são classificadas e são avaliadas as métricas acurácia, que neste caso sintetiza a relação entre a quantidade de erros e acertos do método na detecção dos animais, e a Intersection-over-Union (IoU), a qual avalia a porcentagem da sobreposição da predição feita corretamente sobre o objeto real na imagem, em complemento ao resultado a contagem dos animais de forma automatizada, associada à identificação da posição geográfica aproximada de cada animal.

V. Materias e Métodos

Como estudo de caso de aplicação do método proposto, foi mapeada uma área de 20 hectares (ha) de pastagem que faz parte dos campos experimentais da Embrapa Pecuária Sul, na cidade de Bagé, localizada na região sul do Brasil, caracterizada pela presença de animais manejados de forma extensiva e em campo nativo. Para o processamento do método, utilizou-se uma máquina virtual (Virtual Machine - VM) do Google Colaboratory (Colab), disponibilizada com Linux e ambiente de desenvolvimento Python 2 e 3 já preparados para pesquisas voltadas para inteligência artificial com CPU e GPU, de forma gratuita. No Colab uma VM com ou sem GPU é acionada por demanda, permitindo avaliar a execução de um programa completo ou em blocos de código. A plataforma pode ser útil na aceleração de processamentos computacionais de projetos de aprendizado de máquina nos quais o fator tempo é determinante. Tem como base o ambiente web Jupyter Notebook, o qual permite acesso remoto sem qualquer tipo de instalação necessária e o compartilhamento de scripts de execução de linguagens interpretadas, principalmente projetos Python, podendo ser mais rápida que 20 nucleos físicos em aplicações de aprendizado profundo, eliminando a etapa de configuração de ambiente de execução para o treinamento [20].

A. Experimentos

Hiperparâmetros precisam ser definidos para qualquer CNN, com destaque para *batch size*, *learning rate* e épocas. O *batch size* determina o número de imagens transmitidas por uma única passagem para *forward/backward* da rede. A *learning rate* é uma constante que determina com que rapidez a rede pode abandonar o que aprendeu antes para obter novas informações. Uma época refere-se a uma única apresentação de todos os dados de treinamento através da rede. Os hiperparâmetros podem ser ajustados intuitivamente, dependendo da resposta do modelo aos dados de treinamento [21].

Considerando estes princípios, utilizou-se o algoritmo *Faster R-CNN*, com a arquitetura *Inception Resnet v2*, com os seguintes hiperparâmetros: variação de escala e proporção de 0,25 a 3,0, *stride* de tamanho 16, *batch size* 12, dividido em épocas de 600 e 1100, e *threshold* mínimo de 60% e com capacidade máxima de detecção definida para 300 *bounding boxes* por imagem, *learning rate* 0.0002 a cada 200 épocas.

1) Primeiro Experimento: A execução da primeira etapa do método envolveu a coleta das imagens RGB na área com animais a campo, na posição em pé, por volta das 16 h (condições de luminosidade não ideais), com plano de voo definido para 100 m de altitude, com sobreposição horizontal e vertical de 60% com resolução espacial (GSD) de 5 cm por pixel quadrado. Na terceira etapa, foram unidas as imagens sequenciais a partir de suas bordas semelhantes por meio do algoritmo SIFT na ferramenta OpenDroneMap. Foi escolhida uma das imagens (com animais) em seu formato original, dividida em subimagens de 400 por 400 px, e a partir destas foram selecionados 52 animais - definidos como classe "vaca" e anotados com sua respectiva bounding box. Estas anotações de animais, ilustradas na Fig. 2, foram definidas como conjunto de treinamento para a CNN, incluindo o uso da técnica Data Augmentation com rotação das imagens a 45, 90, 135 e 180 graus, representando 60% dos 87 animais encontrados no aerolevantamento. O conjunto de imagens de teste foi dos 40% restantes (todos os experimentos deste trabalho usaram a mesma proporção). A imagem total da área foi dividida em tamanhos de 200 px até 800 px quadrados, com intervalos de 100 px, com a intenção de diminuir o tamanho total da imagem e identificar as variações na acurácia.

Fig. 2. Imagens a campo com anotações de animais.

O desenvolvimento do classificador se deu em Python 2.4 dentro do Google Colab. Foi criado um script para instalação do TensorFlow 1.12, importação de imagens de treinamento e classificação. As imagens da etapa três foram submetidas ao algoritmo classificador Faster R-CNN realizando um treinamento de acordo com os hiperparâmetros.

2) Segundo Experimento: foi realizado outro aerolevantamento, as 12 h (condição teoricamente ideal de posição solar), com 50 animais dispostos de forma aleatória na área, sendo que estes são outros animais em relação ao primeiro experimento, mas desta vez em posições em pé e deitado, com as mesmas configurações da CNN do primeiro experimento, demandando a criação uma classe complementar denominada "vaca deitada". Dos 60% (30 animais) das imagens necessárias para treinamento, o resultado gerado no primeiro experimento foi capaz de automatizar a classificação de 44%. Os 16% exigiram anotação manual das *bounding boxes*.

3) Terceiro Experimento: As imagens do primeiro e segundo experimento foram submetidas em conjunto ao treinamento da rede neural, mantendo as mesmas configurações e proporções, com o objetivo de aumentar acurácia do primeiro experimento e avaliar a influência da quantidade de imagens.

As imagens usadas nos experimentos estão disponíveis na seguinte url: https://drive.google.com/uc?id=1FSgRpn0KX5COY0v4yQyk8jc51UeIA3cl&export=download

VI. Resultados e Discussões

A. Métricas Avalidas nos Resultados

Para avaliar a eficácia do método, foram adotadas duas formas de avaliação anteriormente definidas: Acurácia e IoU. A acurácia é calculada conforme a Eq.1:

$$\frac{VPR + VN}{VP + FP + VN + FN} \tag{1}$$

onde, Verdadeiro Positivo Real (VPR): foi detectado corretamente, excluindo dupla detecção; Verdadeiro Negativo (VN): não foi detectado corretamente; Verdadeiro Positivo (VP): foi detectado corretamente; Falso Positivo (FP): foi detectado, mas não é um animal; Falso Negativo (FN): existe e não foi detectado corretamente. IoU mede quanto há de área de sobreposição (AS) dos objeto predito dividido pela área de união (AU) a região predita mais região verdadeira do objeto a ser buscado, conforme definido na Eq.2:

$$\frac{AS}{AU} \tag{2}$$

B. Resultados e Discussões

Os resultados do experimento 1, expostos na Tabela I, mostram que o método foi capaz de detectar com uma acurácia de 86% com 1000 épocas e sobreposição (IoU) mínima de 75% de intersecção correta entre a bounding box. A variação de IoU interfere na acurácia do método, já que o animal pode estar fragmentado entre duas imagens separadas. Outra tentativa foi feita para aumentar a acurácia diminuindo o IoU para 65%, e por fim a 60%, sendo obtida a acurácia de 91%.

Já no experimento 2, como pode ser visto nos resultados apresentados na Tabela II, o método foi capaz de detectar com uma acurácia de 95% com 1000 épocas e IoU de 60%, também confirmando que o uso de um IoU menor melhora a acurácia. Outro aspecto relevante desse experimento é a hora do voo (12 h), não havendo interferência de sombras.

TABELA I
EXPERIMENTO INCLUINDO IMAGENS DO VOO 1 COM 1000 ÉPOCAS

Total	VPR	FP	VN	FN	DD	Threshold	Acurácia
35	30	0	0	5	0	75%	86%
35	31	0	0	4	0	65%	88%
32	29	0	0	3	0	60%	91%

TABELA II
EXPERIMENTO INCLUINDO IMAGENS DO VOO 2 COM 1000 ÉPOCAS

Total	VPR	FP	VN	FN	DD	Threshold	Acurácia
22	17	0	0	5	0	75%	77%
22	19	0	0	3	0	65%	86%
22	21	0	0	1	0	60%	95%

O experimento 3, mostrado na Tabela III, teve a acurácia de 92%, mas houve aumento na contagem de animais com a variação do IoU. Neste experimento ocorreu uma contagem duplicada. Quando ocorreram duplicações de detecção de animais, foi necessário estabelecer um critério de exclusão para os casos nos quais há sobreposição de pelo menos 50% entre as *bounding boxes* no mesmo animal.

TABELA III
EXPERIMENTO INCLUINDO IMAGENS DOS VOOS 1 E 2 COM 1000 ÉPOCAS

Total	VPR	FP	VN	FN	DD	Threshold	Acurácia
28	25	0	0	3	0	75%	90%
28	26	0	0	2	0	65%	92%
28	26	0	0	2	0	60%	92%

Como forma de identificar onde o treinamento deve ser interrompido, como mostra a Fig. 3, nos três experimentos, observou-se que, a partir de 1200 épocas, houve queda no aprendizado, indicando que 1000 épocas mostrou-se ideal.

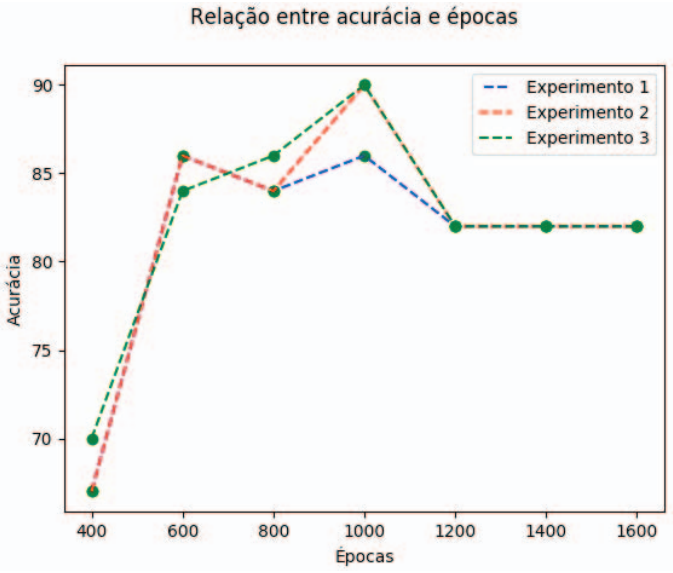


Fig. 3. Efeito da quantidade de épocas nos três experimentos.

Todos os experimentos apresentaram maior acurácia em 300 px e 400 px quadrados em razão do tamanho das imagens e capacidade de detecção da rede neural. A Fig. 4 ilustra a variação nos três experimentos, de modo que quanto menor a IoU do objeto (animal) a ser encontrado, maior tende a ser a eficácia da contagem. O IoU é fator determinante para acurácia já que é possível encontrar animais abaixo de 100% de sobreposição, levando a um resultado eficaz com animais vistos sob outros ângulos, característica que ocorre em aerolevantamentos.

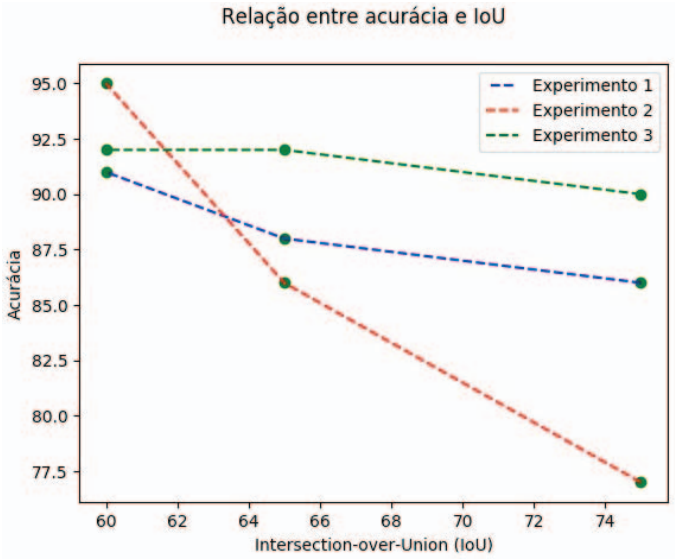


Fig. 4. Efeito do IoU nos três experimentos.

A partir das *bounding boxes* obtidas no primeiro voo, foi possível identificar a localização geográfica aproximada dos animais, sobrepondo o arquivo JPG sobre o arquivo TIFF (arquivo que contém informações sobre as posições geográficas) correspondente ao ortomosaico.

Fig. 5. Disposição geográfica de animais no experimento 1.

Na Fig. 5, é apresentado o mapa temático retirado do Google

Maps com a área onde foi realizado o voo do experimento 1. As posições geográficas foram retiradas das imagens e são indicadas pelos ícones azuis, facilitando a compreensão da distribuição dos animais na área.

VII. Conclusões

Com o método apresentado, foi possível identificar animais à distância de 100 m e contá-los através de imagens registradas por VANT com câmera RGB. O método atingiu melhores resultados quando animais estiveram em posição deitada, com maior ângulo de incidência de luz solar e alterações no IoU para 60%. Embora a contagem de animais não tenha se aproximado de 99% ou mais, foi próxima ao total, demonstrando que o método tem capacidade de ser usado não apenas na contagem de animais mas também na identificação de sua localização no campo, e ainda pode ser aprimorado coletando mais imagens de outros voos em séries temporais, inclusive em outras áreas de tamanhos superiores a 20 ha. Além disso, conforme sugerem os resultados obtidos, o processo de contagem e localização de animais, desde que melhorada a acurácia na contagem coletando mais imagens, pode ser base para uma solução automatizada, desde o treinamento do classificador até o resultado final da contagem, ficando manual apenas o aerolevantamento da área que contém animais, por questões de segurança regulamentar.

Agradecimentos

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001, tendo em vista o vínculo dos autores ao Programa de Pós-Graduação em Computação Aplicada (PPGCAP), resultado da cooperação entre UNIPAMPA e EMBRAPA. Destaca-se também o fundamental suporte recebido de Naylor Bastiani Perez e Ânderson Fischoeder Soares na viabilização dos aerolevantamentos. Adicionalmente, cabe agradecer aos revisores da IEEE Latin America Transactions pelas suas relevantes contribuições que permitiram qualificar a versão final deste trabalho.

Referências

[1] "Pecuária de precisão ganha destaque na 40a Expointer," Sep. 5, 2017. Accessed on: Jan. 15, 2016. [Online]. Available: https://www.embrapa.br/busca-de-noticias/-/noticia/26426446/pecuaria-de-precisao-ganha-destaque-na-40-expointer

[2] V. F. Barbalho, S. P. C. Casa Nova, A. C. Pereira, and A. B. S. Oliveira, "O controle de estoque de animais na pecuária bovina de corte: uma questão de continuidade", in *Anais do XII Congresso Brasileiro de Custos*, Florianópolis, SC, Brasil, 2005, Art. no. 1927.

[3] B. C. Vasconcellos, "Método aplicado ao monitoramento remoto de animais vaseado em aerolevantamento com VANT e aprendizagem profunda," M.S. thesis, Programa de Pós-Graduação em Computação Aplicada, Universidade Federal do Pampa, Bagé, RS, Brasil, 2019.

[4] J. G. A. Barbedo, L. V. Koenigkan, T. T. Santos, and P. M. Santos, "A study on the detection of cattle in UAV images using deep learning," *Sensors*, vol. 19, no. 24, Dec. 2019, Art. no. 5439, DOI: 10.3390/s19245436.

[5] D. Wang, Q. Shao, and H. Yue, "Surveying wild animals from satellites, manned aircraft and unmanned aerial systems (UASs): a review," *Remote Sensing*, vol. 11, Jun. 2019, 10.3390/rs11111308, Art. no. 1308.

[6] J. A. Vayssade, R. Arquet, and M. Bonneau, "Automatic activity tracking of goats using drone camera," *Computers and Electronics in Agriculture*, vol. 162, pp. 767-772, Jul. 2019, DOI: 10.1016/j.compag.2019.05.021.

[7] W. Andrew, C. Greatwood, and T. Burghardt, "Visual localisation and individual identification of Holstein Friesian cattle via deep learning," in *Proc. of IEEE International Conference on Computer Vision Workshop (ICCVW)*, Venice, 2017, pp. 2850-2859, DOI: 10.1109/ICCVW.2017.336.

[8] N. Rey, "Combining UAV-imageryand machine learning for wildlife conservation," M.S. thesis, Laboratory of Geographic Information Systems, Ecole Polytechnique Federale De Lausanne, Switzerland, 2016.

[9] B. C. Vasconcellos, J. P. Trindade, L. Volk, and L. B. Pinho, "Análise de Desempenho da Execução Remota de Método Aplicado ao Monitoramento de Animais com VANT," in *Anais da XIX Escola Regional de Alto Desempenho da Região Sul (ERAD)*, Três de Maio, RS, Brasil, 2019, Art. no. 7067.

[10] A. S. Ferreira, "Redes neurais convolucionais profundas na detecção de plantas daninhas em lavoura de soja," M.S. thesis, Faculdade de Computação, Universidade Federal do Mato Grosso do Sul, Campo Grande, MS, Brasil, 2017.

[11] C. Szegedy, A. Toshev, and D. Erhan, "Deep neural networks for object detection," in *Proc. of the 26th International Conference on Neural Information Processing Systems*, vol. 2, Lake Tahoe, Nevada, USA, 2013, pp. 2553-2561, DOI: 10.5555/2999792.2999897.

[12] W. Hu, Y. Huang, L. Wei, F. Zhang, and H. Li, "Deep convolutional neural networks for hyperspectral image classification," *Journal of Sensors*, vol. 2015, pp. 1-12, Jan. 2015, Art. no. 258619, DOI: 10.1155/2015/258619.

[13] E. Bezerra, "Introdução à aprendizagem profunda," in Tópicos em gerenciamento de dados e informações, Simpósio Brasileiro de Banco de Dados, E. Ogasawara and V. Vieira, eds., 1st ed. Salvador, BA, Brasil: Sociedade Brasileira de Computação, 2016, pp. 57-86. [Online]. Available: http://sbbd2016.fpc.ufba.br/e-book/minicursos.pdf.

[14] L. D. Nguyen, D. Lin, Z. Lin, and J. Cao, "Deep CNNs for microscopic image classification by exploiting transfer learning and feature concatenation," in *Proc. IEEE ISCAS*, Florence, 2018, pp. 1-5, DOI: 10.1109/ISCAS.2018.8351550.

[15] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "SSD: Single shot multibox detector," in Computer Vision – ECCV 2016, Lecture Notes in Computer Science, vol. 9905, B. Leibe and J. Matas, eds., Cham, Switzerland: Springer, 2016, pp. 21–37, DOI: 10.1007/978-3-319-46448-0_2.

[16] Y. Zhao, J. Li, X. Li, and Y. Hu,"Low-altitude UAV imagery based cross-section geological feature recognition via deep transfer learning," in *Proc. of 3rd International Conference on Robotics and Automation Engineering (ICRAE)*, Guangzhou, China, 2018, pp. 253-257, DOI: 10.1109/ICRAE.2018.8586733.

[17] C. Silva, D. Welfer, F. P. Gioda, and C. Dornelles, "Cattle brand recognition using convolutional neural network and support vector machines," *IEEE Latin America Transactions*, vol. 15, no. 2, pp. 310-316, Feb. 2017, DOI: 10.1109/TLA.2017.7854627.

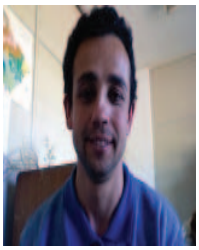
[18] K. Yang and F. Geng, "Application of faster R-CNN model on human running pattern recognition," Nov. 2018. [Online]. Available: arXiv:1811.05147 [cs.CV]

[19] X. Miao, X. Liu, J. Chen, S. Zhuang, J. Fan, and H. Jiang, "Insulator detection in aerial images for transmission line inspection using Single Shot Multibox Detector," *IEEE Access*, vol. 7, pp. 9945-9956, 2019, DOI: 10.1109/ACCESS.2019.2891123

[20] T. Carneiro, R. V. M. Nóbrega, T. Nepomuceno, G. Bian, V. H. C. Albuquerque, and P. P. R. Filho, "Performance analysis of Google Colaboratory as a tool for accelerating deep learning applications," *IEEE Access*, vol. 6, pp. 61677-61685, 2018, DOI: 10.1109/ACCESS.2018.2874767.

[21] D. Griffiths and J. Boehm, "Rapid object detection systems, utilising deep learning and unmanned aerial systems (UAS) for civil engineering applications," *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XLII-2, pp. 391-398, DOI: 10.5194/isprs-archives-XLII-2-391-2018.

Bruno Campos de Vasconcellos Possui graduação em Sistemas de Informação (2004) URCAMP e Especialização em Sistemas Distribuídos (2013) pela Universidade Federal do Pampa. Possui experiência em desenvolvimento de software, banco de dados, aplicações web, mobile e desktop, tem interesse em inteligência artificial e visão computacional.



José Pedro Pereira Trindade Possui graduação em Agronomia pela Universidade Federal de Santa Maria (1996), mestrado em Zootecnia pela Universidade Federal de Santa Maria (1999) e doutorado em Zootecnia pela Universidade Federal do Rio Grande do Sul (2003). Atualmente é pesquisador da Empresa Brasileira de Pesquisa Agropecuária. Tem experiência na área de Ecologia, com ênfase em Ecologia de ecossistemas campestres, atuando principalmente nos seguintes temas: padrões e processos em ecossistemas campestres, ecologia de comunidades, ecologia de paisagem e análise multivariada de dados biológicos.

Leandro Bochi da Silva Volk Engenheiro Agrônomo pela Universidade Federal do Rio Grande do Sul (1998), possui mestrado (2002) e doutorado (2006) em Ciência do Solo pela Universidade Federal do Rio Grande do Sul. Atuou como professor adjunto na Universidade Estadual de Maringá de 2005 a 2011. Tem experiência na área de Agronomia, com ênfase em Manejo e Conservação do Solo, nos temas: erosão hídrica, conservação, manejo e qualidade do solo. Atualmente é Professor de Solos no curso de Agronomia da Faculdades IDEAU e Pesquisador da Embrapa Pecuária Sul, ligado ao Laboratório de Estudos em Agroecologia e Recursos Naturais, com atuação no entendimento de processos edáficos na relação solo-planta-animal em sistemas pecuários.

Leonardo Bidese de Pinho Bacharel em Ciência da Computação, pela Universidade Católica de Pelotas (UCPEL); Mestre e Doutor em Engenharia de Sistemas e Computação, pelo Programa de Engenharia de Sistemas e Computação (PESC) da COPPE-Universidade Federal do Rio de Janeiro (UFRJ). Realiza PD&I nas áreas de Engenharia e Ciência da Computação, principalmente em Computação de Alto Desempenho e Alta Eficiência (duas ênfases: sistemas baseados em redes de sensores sem fio, particularmente sistemas com sensores e atuadores móveis e estáticos aplicados à pecuária de precisão; e sistemas distribuídos multimídia, em especial sistemas de distribuição de vídeo sob demanda escaláveis e custo-efetivos).