15º Congresso Interinstitucional de Iniciação Científica – CIIC 2021
01 a 02 de setembro de 2021
ISBN 978-65-994972-0-9

IBOJU: UMA FERRAMENTA DE ANOTAÇÃO DE IMAGENS PARA TREINAMENTO DE DETECTORES

Pedro Andrade Ferreira **Sobrinho[1];** Thiago Teixeira **Santos[2]**

Nº 21606

RESUMO *– A automação no contexto da agricultura é significativamente mais complexa do que em contextos urbanos e industriais. Devido à grande variabilidade das estruturas orgânicas, métodos cada vez mais complexos e rebuscados são necessários para um bom reconhecimento de padrões e detecção de imagens, tão necessários para a automação de processos produtivos de atividades ligadas à agropecuária. Contudo, devido à natureza das imagens naturais, existem poucas ferramentas capazes de anotá-las da forma desejada. Neste trabalho, desenvolvemos a Iboju, uma ferramenta de segmentação em imagem natural capaz de anotar, rápida e eficientemente, em imagens capturadas de ambientes naturais ligados à agropecuária, gerando caixas delimitadoras e máscaras do objeto desejado. Para isso, foram usadas técnicas de Deep Learning com redes neurais convolutivas aliadas a algoritmos de segmentação de macropixels. A ideia principal da Iboju é que a Rede Neural gere uma anotação que seria corrigida e aprimorada interativamente pelo usuário através de algoritmos de segmentação. Ao compararmos o uso da Iboju a outras ferramentas de anotação, é notável a maior rapidez e exatidão, sendo possível a criação de bases de dados cada vez maiores que poderão ser usadas para melhorar o desempenho de redes neurais de detecção de imagens. A Iboju alimentaria treinamentos de redes que poderão ser usadas dentro dela. A partir da Iboju, é possível a construção de bases de dados cada vez maiores, que poderão ser usadas para alimentar algoritmos de automação de processos que dependem de detecção, como coleta de frutos e contagem de gado.*

Palavras-chaves: segmentação, aprendizagem profunda, anotação de imagens.

1 Autor, Bolsista CNPq (PIBIC): Graduação em Ciência da Computação, UNICAMP, Campinas-SP; pedroandrade_jp@hotmail.com.
2 Orientador: Pesquisador da Embrapa Informática Agropecuária, Campinas-SP; thiago.santos@embrapa.br.

ISBN 978-65-994972-0-9

ABSTRACT – *Automation in the context of agriculture is significantly more complex than in urban and industrial contexts. Due to the great variability of organic structures, increasingly complex and far-fetched methods are necessary for good pattern recognition and image detection, essential for the automation of productive processes in activities related to agriculture. However, due to the nature of natural images, there are few tools capable of annotating them in the desired way. In this work, we developed Iboju, a natural image segmentation tool capable of quickly and efficiently annotating images captured from natural environments linked to agriculture, generating bounding boxes and segmentation masks of the desired object. For this, Deep Learning techniques were used with convolutional neural networks combined with macro-pixel segmentation algorithms. Iboju's main idea is that the Neural Network generates an annotation that would be interactively corrected and enhanced by the user through segmentation algorithms. When compared Iboju to other annotation tools, the greater speed and accuracy is remarkable, making it possible to create increasingly larger databases that can be used to improve the performance of image detection neural networks. Iboju would feed training on networks that could be used within it. Based on Iboju, it is possible to build an increasingly larger database, which can be used to feed algorithms for automating processes that depend on detection, such as fruit collection and cattle counting.*

Keywords: segmentation, deep learning, image annotation.

1 INTRODUÇÃO

A detecção e reconhecimento de imagens é imprescindível para o aprofundamento da automação de processos no contexto da produção agrária, como a coleta automatizada de frutos, combate a pragas e monitoramento de plantações. Segundo Duckett et al. (2018), “Sistemas robóticos baseados em visão dependem fortemente de aprendizado de máquina, com abordagens como redes neurais ganhando tração e aumentando a possibilidade de robôs compartilharem seus conhecimentos.” Contudo, as técnicas que constituem o estado de arte na detecção de objetos dependem, na sua grande maioria, de grandes bases de dados de imagens naturais com anotações que podem ser usadas para treinar algoritmos de detecção.

ISBN 978-65-994972-0-9

No caso do presente trabalho, anotações são necessárias para a realização de aprendizado de máquina supervisionado no contexto da agricultura, ou seja, necessitamos de bases de dados que, além das imagens dos objetos, disporia de caixas delimitadoras – que indicam a área retangular na qual o objeto se encontra – e de máscaras de segmentação – que determinam com maior precisão quais pixels da imagem constituem cada objeto, como é possível visualizar nas Figuras 1 e 2. O problema é que, devido à diversidade dos objetos no contexto do campo e ao fato da área ainda estar sendo explorada, essas bases de dados não existem e podem ser difíceis e caras de serem geradas.

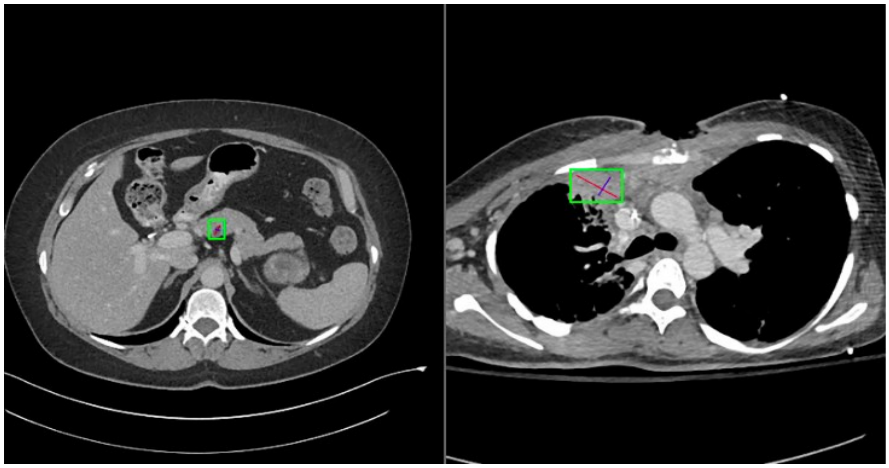

Figura 1. Caixas delimitadoras (*bounding boxes,* em verde) indicando lesões em ossos do corpo humano. Retirado do dataset DeepLesion (YAN et al. 2017)

Algumas ferramentas de anotação de imagens já existem e estão disponíveis e disseminadas na literatura. Entre as mais difundidas temos VGG Image Annotator (VIA) (DUTTA; ZISSERMAN, 2019) e LabelImg (TZUTALIN, 2015). A LabelImg é uma ferramenta que rotula objetos em uma imagem por intermédio de caixas delimitadoras (*bounding boxes*), que delimitam a área na qual o objeto alvo se encontra, não determinando, porém, quais pixels da imagem pertencem ou não ao objeto, permitindo que partes da imagem que não fazem parte do objeto sejam rotulados como se fossem, não podendo ser usado, portanto, para a geração de máscaras.

Já a ferramenta VIA permite que o anotador desenhe através de âncoras para contornar o objeto, criando máscaras poligonais que imitam seu formato.

Contudo, como as imagens naturais têm formatos mais arredondados, contorná-los com a ferramenta VIA torna-se uma tarefa demorada e desgastante, principalmente se levarmos em conta

ISBN 978-65-994972-0-9

que, para um bom treinamento de uma rede neural, precisaríamos de centenas de imagens previamente anotadas.

Portanto, embora ambas as ferramentas sejam usadas amplamente na anotação de bases de dados para o treinamento de modelos de aprendizado de máquina, elas se apresentaram imprecisas ou trabalhosas na anotação de imagens naturais, principalmente na geração de máscaras.

Tendo em vista as necessidade de se produzir grandes bases de dados para serem usadas no desenvolvimento de tecnologias de automação, e a lacuna nessa área, desenvolvemos uma ferramenta interativa de anotação de imagens, precisa e rápida, utilizando algoritmos de segmentação de imagens e de aprendizado de máquina supervisionado.

A ferramenta proposta poderá ser utilizada na anotação de imagens naturais diversas e na criação de bases de dados para serem usadas em processos de aprendizado de máquina.

Figura 2. Máscaras de segmentação diferenciando carros, motos e pessoas no ambiente urbano.

Fonte: Lin et al. (2014)

2 MATERIAL E MÉTODOS

2.1 O APRENDIZADO DE MÁQUINA SUPERVISIONADO

No nosso problema de aprendizado de máquina supervisionado, os algoritmos que compõem o atual estado de arte da área precisam ser alimentados por pares de entrada e saída de exemplos, já resolvidos, do problema, que ele usará para aprender a gerar um modelo que resolva exemplos

ISBN 978-65-994972-0-9

ainda não vistos. Os algoritmos de aprendizado de máquina fazem isso a partir da diferença entre o resultado que o algoritmo gerou (inicialmente de forma aleatória) e o resultado desejado. Para isso, temos que dispor de uma base de dados com centenas de entradas e saídas já rotuladas com o resultado dos exemplos.

A criação de uma base de dados desse tipo, para imagens relacionadas ao contexto da agropecuária, mostrou-se uma tarefa desafiadora utilizando as ferramentas de anotação que normalmente são usadas para essa tarefa. Como podemos ver nas Figuras 1 e 2, fica claro que as ferramentas atuais de anotação são boas para objetos simples, mas muito trabalhosas quando os objetos têm forma complexa (Figuras 3 e 4).

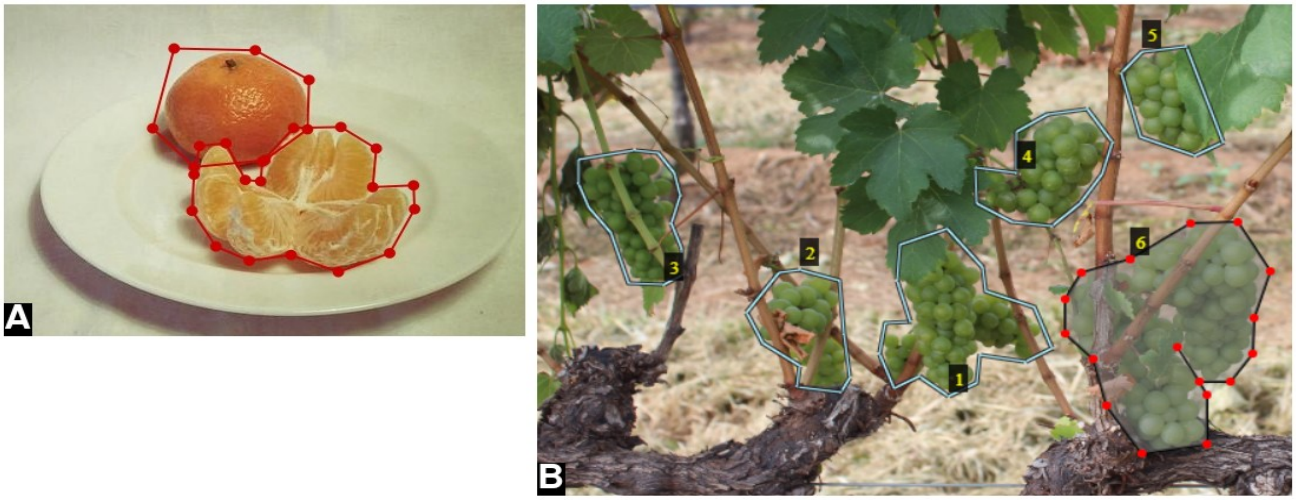

Figura 3. (A) Imagem de laranjas anotadas utilizando a ferramenta LabelImg; (B) Imagem de cachos de uva anotados utilizando a ferramenta VIA.

2.1 INTEGRAÇÃO DA REDE MASK R-CNN

A rede neural utilizada nesse estudo é a Mask R-CNN (HE et al., 2017) desenvolvida pela Facebook AI-Reseach. A Mask R-CNN é uma extensão da rede Faster R-CNN e utiliza redes neurais convolutivas para extrair características de uma imagem e criar regiões de interesse (*RoIs*). Essas regiões de interesse passam por mais algumas camadas de convolução para a classificação de instâncias, previsão de caixas delimitadoras e máscaras de segmentação de forma independente.

Para que a Mask R-CNN funcione bem, ela precisa ser alimentada por uma base de dados de imagens previamente anotadas com máscaras. A partir daí, as ferramentas de anotação citadas

ISBN 978-65-994972-0-9

anteriormente tornam-se cada vez menos eficientes, fazendo cada vez mais necessários algoritmos de segmentação semântica capazes de classificar cada pixel de uma imagem.

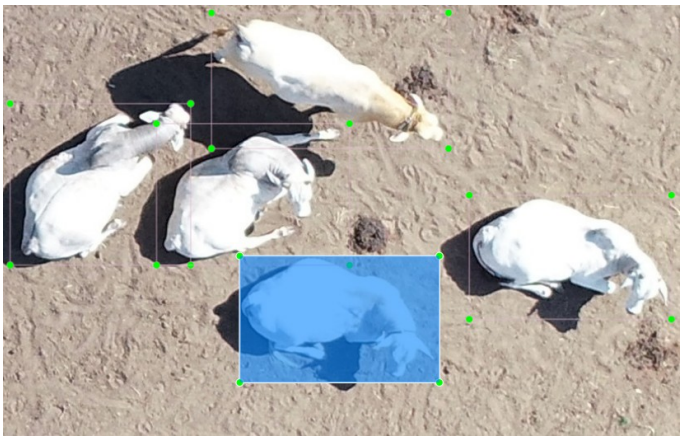

Figure 4. Imagem de bovinos anotada com caixas delimitadoras utilizando a ferramenta LabelImg.

2.2 INTEGRAÇÃO DE ALGORITMOS DE SEGMENTAÇÃO

Os algoritmos de segmentação surgem justamente para atuar nas áreas em que as ferramentas de anotação convencionais não conseguem. Portanto, no desenvolvimento de nossa ferramenta, implementamos o algoritmo proposto por Noma et al. (2012). Em resumo, a ideia central desse algoritmo é que, em vez do processo trabalhoso de criar polígonos que se aproximem do formato do objeto, o usuário só precise fazer alguns rabiscos (*scribbles*) para diferenciá-lo do resto da imagem.

Para isso, rabiscos a mão livre sobre a imagem são feitos pelo usuário para indicar quais são os objetos de interesse. Depois é feita uma segmentação da imagem de entrada em macropixels através de dois grafos, um representado a imagem toda, e outro representando a área que foi rabiscada. A segmentação é feita através de isomorfismo entre esses grafos para propagar os rótulos das regiões marcadas a todas as não marcadas, produzindo uma rotulagem de toda a imagem de entrada.

ISBN 978-65-994972-0-9

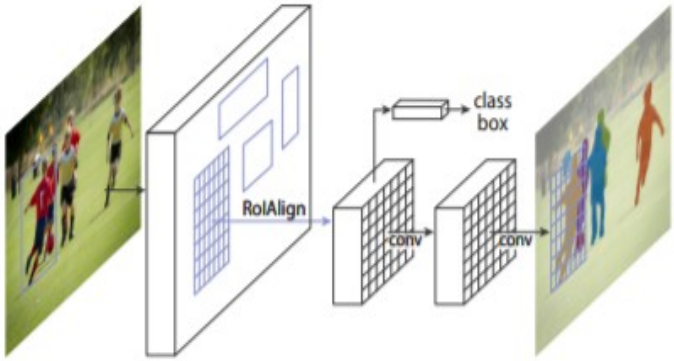

Figure 5. Aplicação do framework Mask R-CNN para segmentação de instâncias.

Fonte: He et al. (2017).

2.3 DESENVOLVIMENTO DA IBOJU

Partindo desse cenário, desenvolvemos a Iboju (máscara no idioma iorubá), uma nova ferramenta de anotação de imagens que utiliza segmentação semântica e redes neurais para anotar imagens mais complexas. A Iboju foi desenvolvida utilizando a linguagem Python e fazendo uso da Interface Gráfica Tkinter. Tendo em vista que estamos criando uma ferramenta interativa de segmentação de imagens, precisamos de uma forma do usuário visualizar as imagens e interagir com elas, e o Tkinter faz isso rápida e eficientemente. Além disso, utilizamos as bibliotecas Pytorch e Torchvision para funcionalidades que envolvem aprendizado de máquina. Utilizamos essas bibliotecas devido à disponibilidade de modelos de aprendizado de máquina já implementados.

3 RESULTADOS E DISCUSSÃO

Ao iniciar a Iboju, é fornecida uma imagem que se quer segmentar, um arquivo texto opcional indicando as caixas delimitadoras já existentes para aquela imagem, e um modelo de aprendizado de máquina já treinado que auxiliará o usuário na automação do processo de anotação.

ISBN 978-65-994972-0-9

A interface da Iboju é dividida em duas sub-interfaces principais, a interface de visualização de instâncias de interesse e a de segmentação (Figura 6). A ideia é que o usuário indique as áreas de interesse nas quais se quer anotar, utilizando caixas delimitadoras nessas áreas, podendo informar sobre as caixas ao programa, através de um arquivo texto no formato YOLO (REDMON et al., 2016). Na interface de visualização de instâncias de interesse, é mostrada toda a imagem com as respectivas áreas sinalizadas. Desenvolvemos a Iboju para ser uma ferramenta completa de anotação de imagens, sendo possível, portanto, a criação de caixas delimitadoras na própria Iboju. Através do botão *Create Bounding Box* nesta área, as caixas serão adicionadas às áreas de interesse.

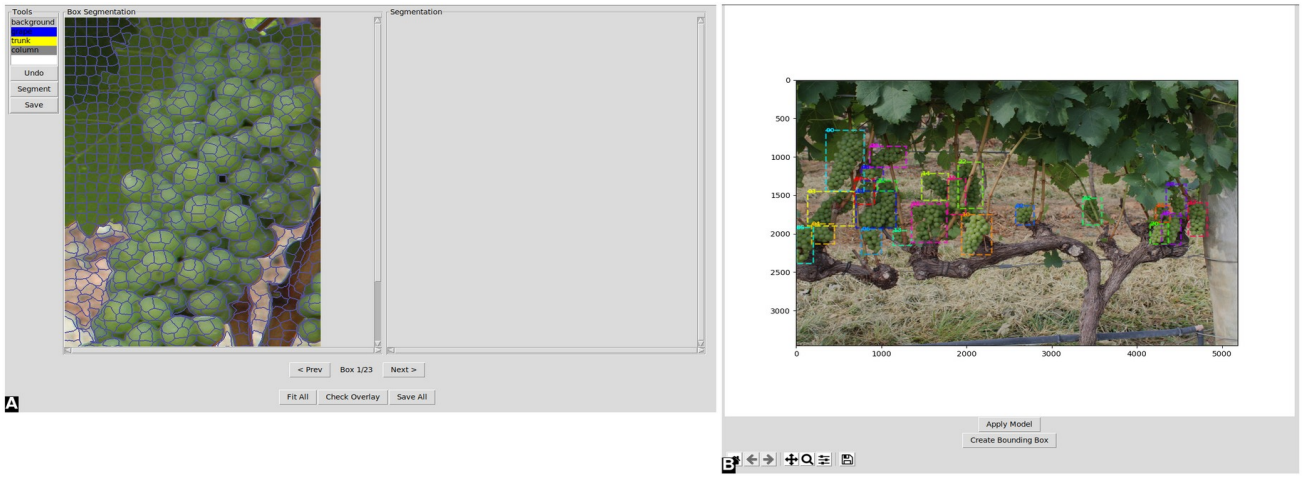

Figura 6. Interface gráfica da Iboju. (A) Interface de segmentação. (B) Interface de visualização de instâncias.

A partir das instâncias de interesse, aplicamos o algoritmo de segmentação proposto por Noma et al. (2012) nas áreas individualmente. O usuário pode desenhar rabiscos (*scribbles*) nos objetos de interesse, gerando máscaras de segmentação para cada instância desejada.

Utilizando uma versão preliminar da Iboju, já com o algoritmo de Noma et al. (2012), porém com menos recursos, foi possível a anotação de fotos de cachos de uvas e a criação da base de dados *Embrapa Wine Grape Instance Segmentation Dataset* (Embrapa WIGISD) (SANTOS et al., 2019) , que contém diversas fotos em alta qualidade de cachos de uvas obtidas na Vinícola Guaspari, localizada em Espírito Santo do Pinhal (SP), além de máscaras e caixas de delimitação no formato YOLO.

ISBN 978-65-994972-0-9

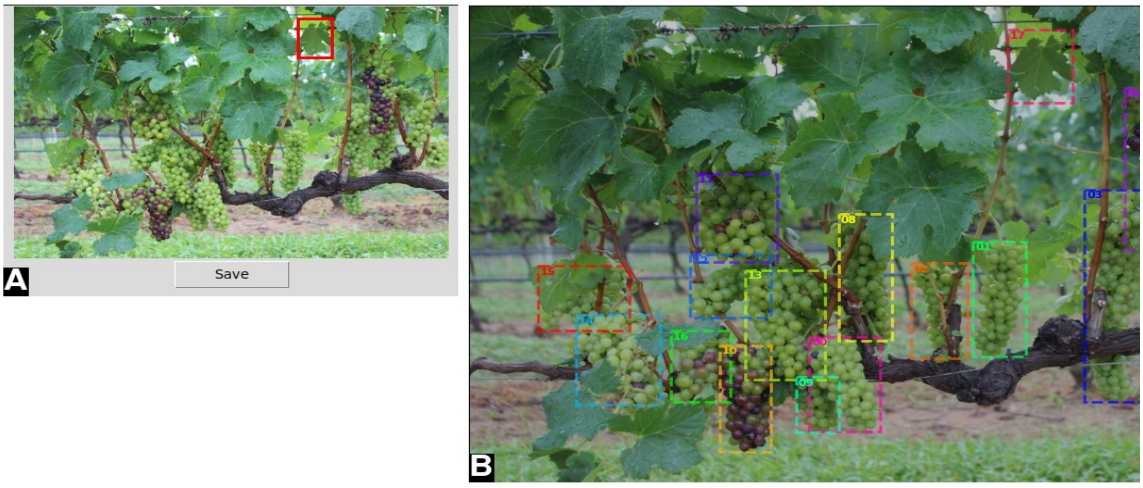

Figura 7. (A) Exemplo de anotação de caixa delimitadora de uma folha utilizando a Iboju; (B) Visualizador de instâncias com a nova área de interesse anotada.

A partir dos dados depositados no WIGISD, uma rede neural Mask R-CNN, implementada pela biblioteca Torchvision (PASZKE et al., 2019), foi treinada para detectar caixa delimitadora e máscaras de cachos de uvas. O modelo foi alimentado por cerca de 130 imagens, ao longo de 25 épocas. Todas as imagens passaram por processos randomizados de aumentação, utilizando o pacote Albumentations (BUSLAEV et al., 2020) implementado em Python. A aumentação é necessária para aumentar a diversidade das imagens, evitando o *overfitting*, que é quando o modelo aprende muito bem as relações existentes no treino, porém, ao receber novas imagens (que não foram usadas no treino) o modelo tem um desempenho ruim.

Durante o treinamento, foi possível observar uma queda e uma posterior estabilização do índice de erro do modelo em porcentagens abaixo de 1. Ao final do treinamento, terminamos com um modelo com acertos médias altos em resposta a entradas ainda não vistas. Porém, o mesmo objeto é detectado pelo modelo múltiplas vezes, aumentando o número de inferências detectadas. Para contornar isso, utilizamos o algoritmo Non-maximum suppression (GIRSHICK et al., 2015) já implementado no Pytorch, que basicamente analisa as áreas que se interseccionam e suprime as com menor possibilidade de conter objetos.

Com o modelo pré-treinado em mãos, foi possível exportá-lo facilmente para dentro da Iboju. A ideia é utilizar a Mask R-CNN dentro de nossa ferramenta para fazer uma segmentação geral da imagem de forma rápida e automática. Como pode ser visto na Figura 9, erros serão produzidos pelo modelo, contudo eles poderão ser corrigidos de forma interativa pelo anotador

ISBN 978-65-994972-0-9

utilizando *scribbles*. As novas anotações produzidas dessa forma podem posteriormente auxiliar o treinamento de redes mais acuradas. Esse processo pode se repetir até atingir a eficiência e a automação desejada.

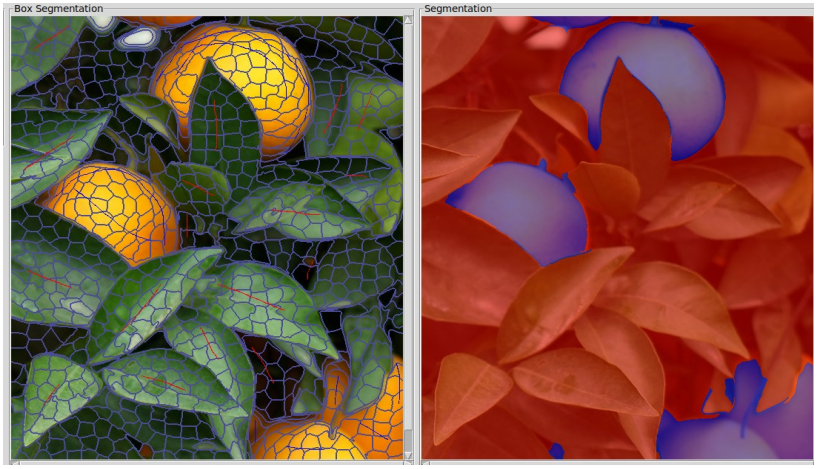

Figura 8. Máscaras de laranjas criadas utilizando a segmentação da Iboju.

A partir do uso de um algoritmo interativo de segmentação semântica junto à aplicação de uma rede neural treinada com cerca de 150 exemplos, foi possível a anotação de imagens de forma rápida e com a exatidão das máscaras exigida para o aprendizado de máquina supervisionado.

Em relação às demais ferramentas de anotação já referenciadas neste trabalho, a Iboju possui vantagens indiscutíveis, como a possibilidade de segmentação semântica mais precisa, o uso auxiliar de uma modelo de Machine Learning e a autonomia a outras ferramentas. Contudo, para que se possa implementar Iboju com todas as suas funcionalidades, ainda é preciso treinar externamente uma rede neural Mask R-CNN e exportá-la para a ferramenta. A exclusividade ao uso do modelo Mask R-CNN dentro da Iboju e a necessidade de treiná-la fora da ferramenta são limitações, as quais se pretende reverter no futuro. Para alcançar uma autonomia maior para a Iboju, o ideal seria que ela fosse capaz de realizar um pequeno treinamento da rede a partir das novas instâncias produzidas, porém teríamos que levar em conta a limitação da máquina do anotador, pois o treinamento de redes voltadas à segmentação de imagens são bastante custosas e normalmente são realizadas por máquinas com GPU.

ISBN 978-65-994972-0-9

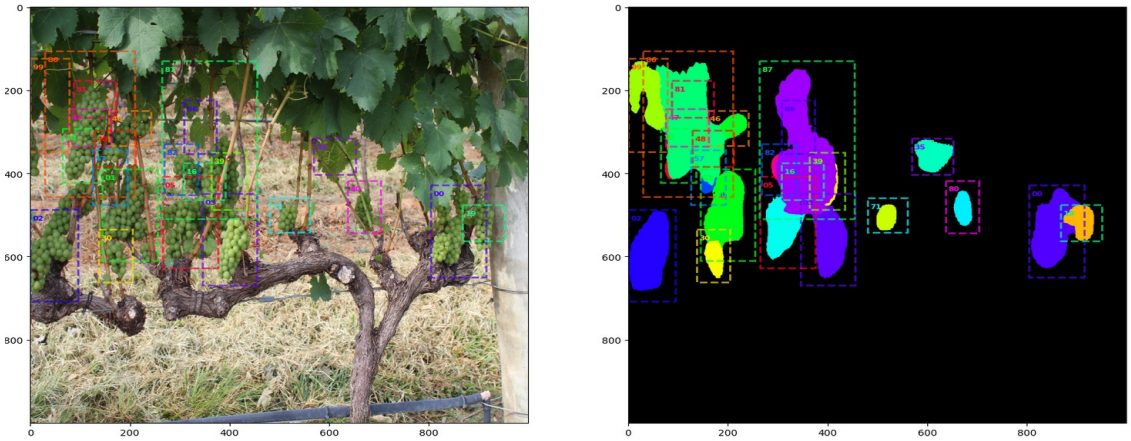

Figura 9. Resultado da aplicação do modelo dentro da Iboju.Explicar aqui o que significam as imagens A e B (inserir identificação A e B nas imagens)

4 CONCLUSÃO

Neste trabalho apresentamos o problema de anotação de bases de dados voltadas ao aprendizado de máquina supervisionado de imagens, focando na detecção e classificação de imagens naturais e orgânicas. Revisamos as ferramentas de anotação já existentes e, devido às lacunas em relação a anotação de imagens naturais, propomos a Iboju, uma nova ferramenta de anotação de imagens que, como consequência do uso conjunto do algoritmo de segmentação semântica de Noma et al. (2012) e da rede Mask R-CNN (HE et al., 2017), permitiu-nos chegar à produção de uma base de dados de forma muito mais rápida e exata, apesar da forte dependência de um treinamento externo da rede.

5 AGRADECIMENTOS

Pedro Sobrinho agradece ao CNPq pela bolsa concedida, à Embrapa pela oportunidade de desenvolver a Iboju e ao orientador Thiago Santos pelos conhecimentos fornecidos, sempre estar presente para responder em mensagens, boas conversas e o prazer de um encontro inusitado em tempos de quarentena. Thiago Santos agradece à FAPESP pelo financiamento do projeto *Agricultura ciente de ambiente: raciocínio sobre estrutura tridimensional no campo de cultivo (AACr3),* processo 17/19282-7, ao qual o presente trabalho se insere.

ISBN 978-65-994972-0-9

1 REFERÊNCIAS

BUSLAEV, A.; IGLOVIKOV, V. I.; KHVEDCHENYA, E.; PARINOV, A.; DRUZHININ, M.; KALININ, A. A. Albumentations: fast and flexible image augmentations. **Information,** v. 11, n. 2, p. 1-20, 2020. DOI: 10.3390/info11020125.

DUCKETT, T.; PEARSON, S.; BLACKMORE, S.; GRIEVE, B.; CHEN, W. H.; CIELNIAK, G.; CLEAVERSMITH, J.; DAI, J.; DAVIS, S.; FOX, C.; FROM, P.; GEORGILAS, I.; GILL, R.; GOULD, I.; HANHEIDE, M.; HUNTER, A.; IIDA, F.; MIHALYOVA, L.; NEFTI-MEZIANI, S.; NEUMANN, G.; PAOLETTI, P.; PRIDMORE, T.; ROSS, D.; SMITH, M.; STOELEN, M.; SWAISON, M.; WANE, S.; WILSON, P.; WRIGHT, I.; YANG, G. Z. **Agricultural robotics**: the future of robotic agriculture. 2018. Disponível em: <http://arxiv.org/abs/1806.06762>. Acesso em: 20 jul. 2021.

DUTTA, A.; ZISSERMAN, A. The VIA annotation software for images, audio and video. In: ACM INTERNATIONAL CONFERENCE ON MULTIMEDIA, 27.; 2019, Nice. **Proceedings...** New York: ACM, 2019. p. 2276-2279. DOI: 10.1145/3343031.3350535.

GIRSHICK, R.; IANDOLA, F.; DARREK, T.; MALIK, J. Deformable part models are convolutional neural networks. In: IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION, 2015, Boston. **Proceedings…** Los Alamitos: IEEE, 2015. p. 437-446. DOI: 10.1109/CVPR.2015.7298641..

HE, K.; GKIOXARI, G.; DOLLAR, P.; GIRSHICK, R. Mask R-CNN. In: IEEE INTERNATIONAL CONFERENCE ON COMPUTER VISION, 2017, Venice. **Proceedings…** Los Alamitos: IEEE, 2017. p. 2961-2969. DOI: 10.1109/ICCV.2017.322.

LIN, T. Y.; MAIRE, M.; BELONGIE, S.; BOURDEV, L.; GIRSHICK, R.; HAYS, J.; PERONA, P.; RAMANAN, D.; ZITNICK, C. L.; DOLLÁR, P. **Microsoft COCO:** common objects in context. 2014. 15 p. Disponível em: <http://arxiv.org/abs/1405.0312>. Acesso em: 28 jun. 2021.

NOMA, A.; GRACIANO, A. B. V.; CESAR JUNIOR, R. M.; CONSULARO, L.; BLOCH, I. Interactive image segmentation by matching attributed relational graphs. **Pattern Recognition,** v. 45, n. 3, p. 1159–1179, Mar. 2012. DOI: 10.1016/j.patcog.2011.08.017.

PASZKE, A.; GROSS, S.; MASSA, F.; LERER, A.; BRADBURY, J.; CHANAN, G.; KILLEEN, T.; LIN, Z.; GIMELSHEIN, N.; ANTIGA, L., et al. **PyTorch**: An imperative style, high-performance deep learning library. In Advances in neural information processing systems, 32, 2019. **Proceedings...** Curran Associates, Inc, 2019. p. 8026–8037.

REDMON, J.; DIVVALA, S.; GIRSHICK, R.; FARHADI, A. You only look once: unified, real-time object detection. In: IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION, 29., 2016, Las Vegas. **Proceedings…** Piscataway: IEEE, 2016. p. 779–788. DOI: 10.1109/CVPR.2016.91.

SANTOS, T. T.; SOUZA, L. L. de; SANTOS, A. A. dos; AVILA, S. **Embrapa Wine Grape Instance Segmentation Dataset – Embrapa WGISD**. 26. Jul. 2019. Zenodo. DOI: 10.5281/zenodo.3361736.

TZUTALIN, D. **LabelImg**. 2015. Disponível em: <https://github.com/tzutalin/labelImg>. Acesso em: 28 jun. 2021.

YAN K.; WANG, X.; LU, L.; SUMMERS, R. M. **DeepLesion**: automated deep mining, categorization and detection of significant radiology image findings using large-scale clinical lesion annotations. 2017. 9 p. Disponível em: <http://arxiv.org/abs/1710.01766>. Acesso em: 28 jun. 2021.