



BASES DE DADOS SINTÉTICAS COM ANOTAÇÃO AUTOMÁTICA PARA DETECÇÃO DE FRUTOS

Sara Andrade Veronese dos **Reis**¹; Thiago Teixeira **Santos**²;

Nº 23603

RESUMO – *Sistemas de Inteligência Artificial (IA) precisam de um grande número de anotações, que são trabalhosas de serem produzidas. Esse trabalho pode ser diminuído, substituindo a maior parte dos dados reais por dados sintéticos, anotados automaticamente. Com o objetivo de criar uma base de dados sintéticos para detecção de laranjas, foram utilizadas ferramentas de computação gráfica, a fim de gerar imagens fotorrealistas. Para que houvesse uma anotação automática dessas imagens, tirando proveito do fato de serem artificiais, geradas por computador, utilizou-se de conhecimentos de programação e geometria projetiva. De modo, que através de um script associado a um modelo tridimensional fosse possível coletar as coordenadas e dimensões de cada fruto no cenário 3D. Partindo disso, com os conhecimentos de geometria projetiva, esses dados foram transformados para as posições nas coordenadas 2D, de acordo com um modelo projetivo de formação de imagens, produzindo imagens sintéticas fotorrealistas. Assim, foram desenvolvidas as imagens com anotações automáticas para uma base de dados sintéticos que pode auxiliar o treinamento de redes neurais.*

Palavras-chaves: base dados sintética, laranja, Blender, Python, anotações.

1 Autora, Bolsista CNPq (PIBIC): Graduação em Matemática aplicada e computacional, Unicamp, Campinas-SP; sara.reis@colaborador.embrapa.br

2 Orientador: Pesquisador da Embrapa Agricultura Digital, Campinas-SP; thiago.santos@embrapa.br



ABSTRACT – *Artificial Intelligence (AI) systems need a big number of annotations, but these are laborious to be produced. This labor can be reduced by replacing most of the real data for synthetic data, automatically annotated. In order to create a synthetic dataset for the detection of oranges, features of computer graphics were used to make photo-realistic images. Envisioning the automatic annotation in this image, taking advantage of the fact they were artificial, made by computer, knowledge of projective geometry was used. That way, by a script associated with a three dimensional model it was possible to take the coordinates and dimensions of each fruit in the three dimensional setting. From that, with the projective geometry knowledge, the data was transformed to the 2D images, according to a projective model of image generation, making synthetic photo-realistic images. So, images automatically annotated were developed, for creating a synthetic data set that can help the training of neural networks.*

Keywords: synthetic data set, orange, Blender, Python, annotations.

1. INTRODUÇÃO

No aprendizado de máquina supervisionado, a anotação de bases de dados é um processo longo e cansativo, especialmente no caso de imagens agrícolas. A construção de uma base de dados sintéticos otimizaria o processo, possibilitando combinar uma vasta quantidade de dados sintéticos, anotados automaticamente, a uma menor quantidade de dados reais, anotados manualmente, para validação, prática que se mostrou eficiente em outros meios como de detecção de pessoas e veículos (Farzam et. al, 2019).

A partir de um estudo que utiliza de sistema de detecção usando as redes convolucionais YOLO para detecção de frutos verdes (Camargo Neto et. al, 2019) o objetivo é aprimorar essa prática e outras do gênero de detecção de objetos em imagens agrícolas, através de bases de dados sintéticas.

Visando colocar essa alternativa em prática, através da computação gráfica, utilizando modelos 3D de bibliotecas públicas e de criação própria, foi possível criar um cenário que gera imagens fotorrealistas. Escolhendo a laranjeira e seu fruto como objeto de estudo da pesquisa,



foram simuladas diversas imagens, variando posições, rotação e cor das laranjas, além dos modelos de árvores e perspectiva da câmera.

Com a possibilidade de criação dessas imagens, ainda era necessário realizar as anotações, para ser possível sinalizar onde se encontram as laranjas. Sabendo suas posições no cenário 3D em que foram dispostas, através da geometria projetiva esses dados precisavam ser passados para o plano 2D de imagem e indicadas as coordenadas dos frutos.

Assim, neste trabalho, foi feita a criação de um gerador de imagens fotorrealistas, para produção de uma base de dados sintética e iniciado o processo de coleta e conversão de dados para realização de anotações automáticas que serão concretizadas em trabalho futuro.

2. MATERIAL E MÉTODOS

Na realização do trabalho, a ferramenta utilizada para criação dos modelos 3D foi o Blender, software de computação gráfica utilizado para produção de imagens e vídeos realistas na indústria de efeitos especiais e animação. Via bibliotecas públicas, foram escolhidos diversos modelos de laranjeiras e produzido um modelo de laranja (Figura 1) replicado 100 vezes na imagem para criação do cenário.

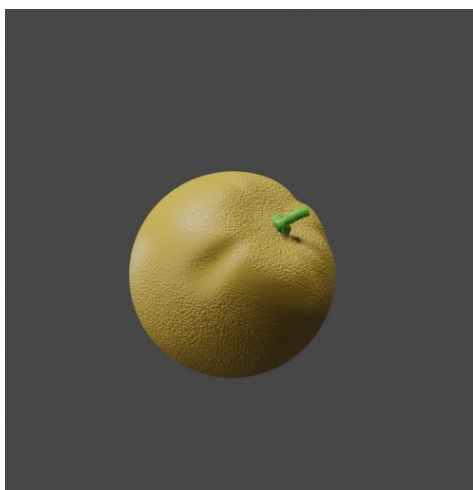


Figura 1: Modelo de laranja criado para os modelos fotorrealistas

No próprio software foi feito um *script* em Python, linguagem aceita pelo Blender, para diversificar as imagens geradas. As 100 laranjas colocadas no cenário eram aleatoriamente tiradas da cena e as que apareciam tinham suas posições, cores e rotações alteradas, produzindo variações realistas. Utilizou-se seis modelos de árvores, gerando um número fixo de cenários para cada, além de serem rotacionadas em torno do seu eixo de altura (eixo z). Para mudança da perspectiva da câmera foram dispostas na cena formas (Figura 2), invisíveis à câmera, pelas quais ela se guiaria, realizando rotações e translações, proporcionando variação de pose nas imagens. Além da disposição de uma fonte de luz, emulando luz solar, que por um arco se movia de leste para oeste, possibilitando diferentes iluminações.

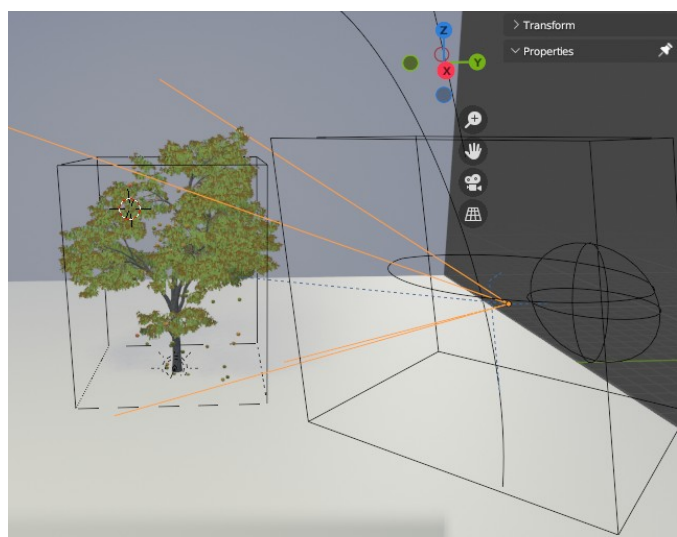


Figura 2: Cenário gerador das imagens

No Blender, tem-se todas as coordenadas do cenário à disposição, porém para as anotações são necessárias as posições na imagem, não no modelo 3D em que está disposta a cena. Para realização dessa conversão, fez-se uso da geometria projetiva. A álgebra linear, junto a conhecimentos a respeito do funcionamento de uma câmera, permitem construir uma matriz que ao multiplicá-la pelas coordenadas no cenário 3D das laranjas, nos fornecem as coordenadas 2D da imagem gerada pela câmera. Trata-se da matriz de câmera P (Equação 1), na qual K é matriz de calibração, Equação 1, composta pela distância focal (convertida para o número de píxeis das dimensões da imagem) e o ponto principal do plano de projeção (Figura 6), e os parâmetros R e t

representam a matriz de rotação e translação de câmera. Essa operação foi realizada no próprio *script* do Blender.

$$K = \begin{bmatrix} \dots \\ \dots \\ \dots \end{bmatrix} \quad (1)$$

$$P = K [R \ v \ t] \quad (2)$$

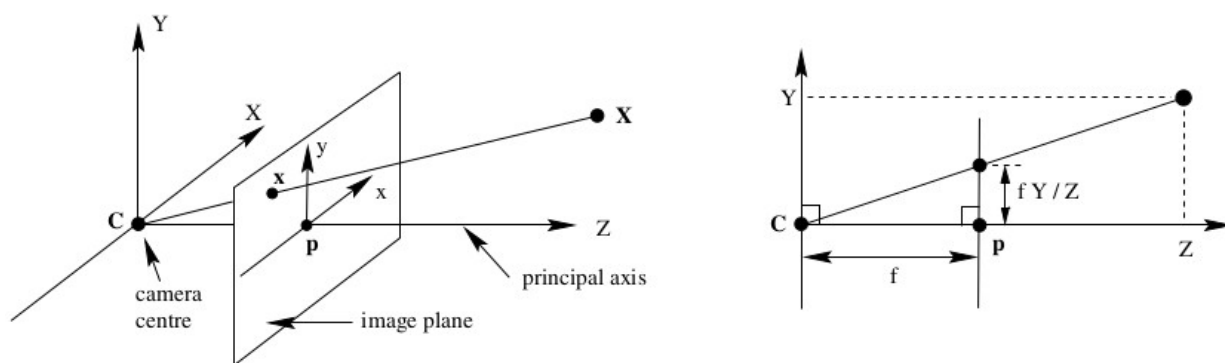


Figura 3: Representação de geometria projetiva com câmera centralizada no eixo z, no qual o ponto p é o ponto principal. A distância focal f é dada pela distância de C a p (câmera até o ponto principal). Imagem retirada de Hartley & Zisserman (2003).

Tendo em vista as coordenadas calculadas no Blender, junto a cada imagem, foi salvo um arquivo texto contendo as posições dos frutos obtidas pelo programa. Assim, foram efetuadas as marcações, utilizando novamente linguagem Python, mas, dessa vez, fora do Blender. Essas indicações consideravam apenas as coordenadas sinalizando as laranjas com um “X” em cada uma delas (Figura 3).

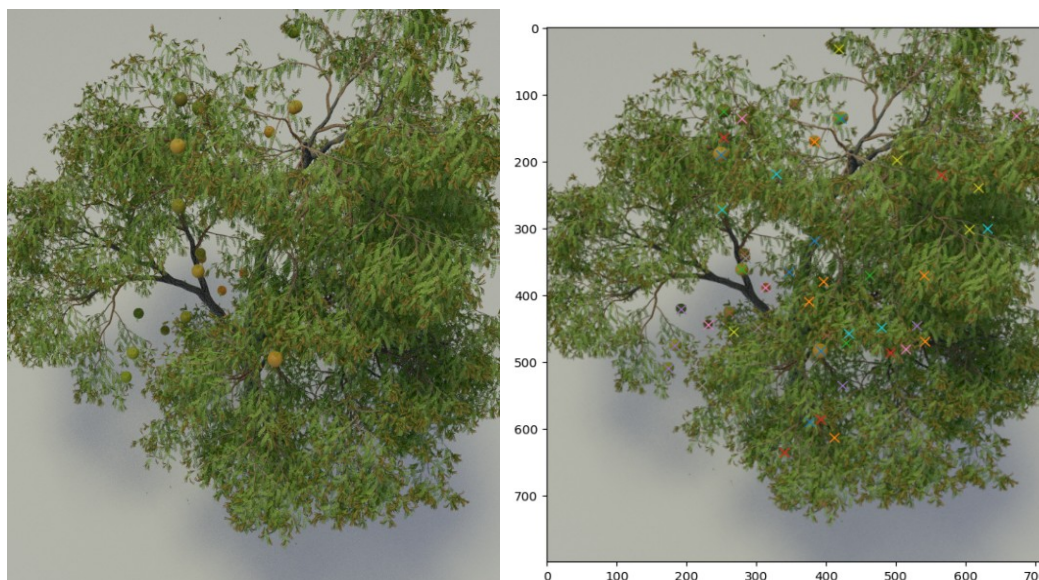


Figura 4: Imagem sem e com a marcação das posições das laranjas

Para uma sinalização ideal, a fim do reconhecimento no treinamento da rede neural, serão utilizadas *bounding boxes*, caixas delimitadoras, importantes para detecção e reconhecimento de objetos, relevantes no treinamento de IA (inteligência artificial). Com objetivo de marcação com esses sinalizadores seria necessário, também, o dimensionamento das laranjas. Os frutos foram aproximados para esferas no cenário 3D e utilizando novamente a geometria projetiva, seus raios seriam convertidos para o plano de imagem 2D.

3. RESULTADOS E DISCUSSÃO

Com a metodologia utilizada foi possível a criação de imagens fotorrealistas, restando o desafio da delimitação das laranjas. A marcação de suas posições foi feita com sucesso (Figura 3), mas, ainda será necessário retirar sinalizações de laranja que se encontram escondidas por galhos e folhas, a fim de não confundir o sistema de IA e obter respostas mais assertivas da rede neural, atividade que será realizada em trabalho futuro.



Além das localizações, já obtidas, era necessário também o raio projetado dos frutos. Na tentativa de encontrar essa projeção, foi realizada uma regra de três com a distância focal parametrizada conforme os píxeis, a (informação extraída da matriz de câmera), a distância da laranja até a câmera, D , e o raio 3D, R_{3D} , conforme a Equação 2. Salvas as informações em um novo modelo de arquivo texto, que incluía as coordenadas e o raio na imagem de cada laranja, foi alterado também o texto que marcava um “X” em cada fruto para marcações com quadrados fornecendo o resultado apresentado na figura 4.

$$R_{2D} = \frac{a \cdot R_{3D}}{D} \quad (3)$$

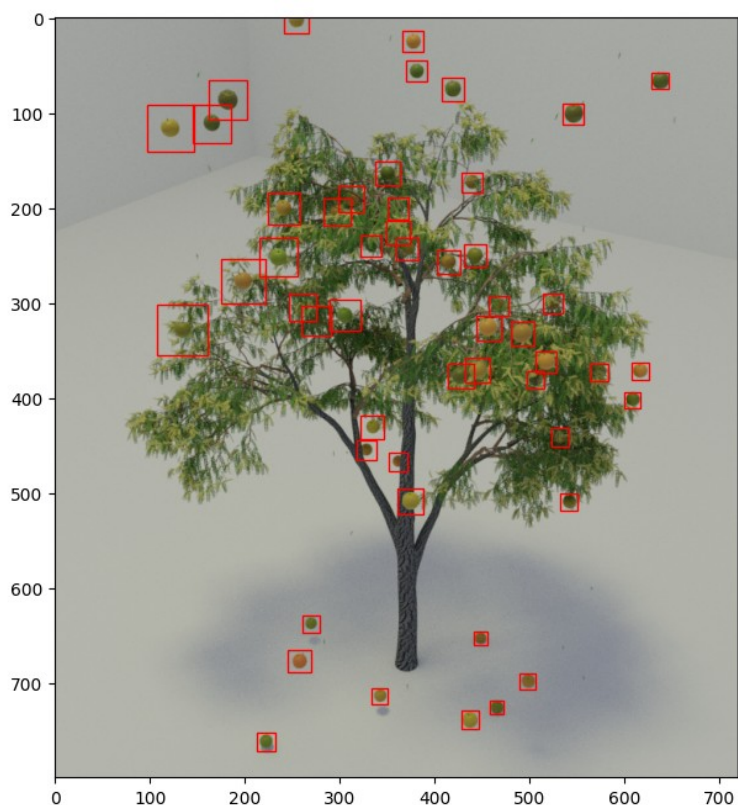


Figura 5: Primeira tentativa de delimitação das laranjas, através do raio e distância da câmera



Contudo, algumas *bounding boxes* ficaram perceptivelmente maiores que as laranjas, mostrando lacunas no método utilizado. Reavaliando o cálculo foi possível perceber que a equação 2 supunha que os frutos se depusessem apenas no eixo principal da câmera, eixo do ponto p na Figura 6, tornando-se necessário duas projeções do plano no qual a laranja se encontra, o que será feito em trabalho futuro.

4. CONCLUSÃO

Com as imagens fotorrealistas, que imitam cor, textura e forma de uma laranjeira e seu respectivo fruto, junto aos dados das coordenadas das laranjas na imagem e os artifícios matemáticos para o dimensionamento das *bounding boxes*, será possível finalizar as anotações da base de dados sintéticas.

Os próximos passos são formar um conjunto de dados sintéticos e combiná-los a dados reais para verificar se eles de fato ajudam a melhorar o desempenho dos detectores de frutos. Em caso positivo, isso possibilitará a otimização do treinamento de uma rede neural por bases de dados sintéticas com anotação automática para detecção de frutos.

5. AGRADECIMENTOS

A autora agradece ao PIBIC/CNPq pelo fomento, com a bolsa concedida, à Embrapa pela oportunidade de trabalhar no projeto, ao orientador Thiago Santos por todo auxílio e conhecimentos passados e aos pais da autora, Andréa e Marcelo, pelo apoio e oportunidade.

6. REFERÊNCIAS

CAMARGO NETO, J., Ternes, S., de Souza, K. X. S., YANO, I., & QUEIROS, L. (2019). Uso de redes neurais convolucionais para detecção de laranjas no campo. In: **CONGRESSO BRASILEIRO DE AGROINFORMÁTICA**, 12., 2019, Indaiatuba. Anais... Ponta Grossa: SBIAGRO, 2019..

HARTLEY, Richard; ZISSERMAN, Andrew. **Multiple view geometry in computer vision**. Cambridge university press, 2003.



17º Congresso Interinstitucional de Iniciação Científica - CIIC 2023
29, 30 e 31 de agosto de 2023
ISSN: 2965-2812

FARZAN, EN. How much real data do we actually need: Analyzing object detection performance using synthetic and real data. **arXiv e-prints**, p. arXiv: 1907.07061, 2019.