



RECONSTRUÇÃO DE CENAS 3D PARA GERAÇÃO DE DADOS SINTÉTICOS PARA DETECÇÃO DE FRUTOS

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

Nº 24603

RESUMO – A anotação de bases de dados em aprendizado de máquina supervisionado, especialmente para imagens agrícolas, é um processo exaustivo. Uma estratégia promissora para otimizar essa tarefa é a construção de bases de dados sintéticas. Este trabalho propõe gerar dados sintéticos a partir de dados reais usando tecnologias atuais, expandindo a base de dados. O estudo explora técnicas avançadas de renderização que utilizam imagens reais e dados de posicionamento de câmera para recriar graficamente a cena, permitindo a geração de novas visualizações a partir de diferentes perspectivas. O primeiro método examinado é o NeRF (Neural Radiance Fields), que se destaca por modelar cenas complexas através de redes neurais, mapeando coordenadas 3D para radiância. Esta abordagem permite a criação de visualizações detalhadas e realistas a partir de diversas perspectivas. Outra técnica é o Gaussian Splatting, que utiliza processos de computação gráfica para distribuir dados da cena sobre uma grade de pixels usando uma função gaussiana, resultando em uma representação suave e contínua da cena, sem depender diretamente de redes neurais ou aprendizado de máquina.

Palavras-chaves: Renderização, Campos de radiação, Base de dados sintéticas, Gaussian Splatting, Agricultura digital, Fruticultura.

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

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

ABSTRACT – Annotating databases in supervised machine learning, especially for agricultural images, is an exhaustive process. A promising strategy to optimize this task is the construction of synthetic databases. This work proposes generating synthetic data from real data using current technologies, expanding the database. The study explores advanced rendering techniques that use real images and camera positioning data to graphically recreate the scene, allowing the generation of new visualizations from different perspectives. The first method examined is Nerf (Neural Radiance Fields), which stands out for modeling complex scenes through neural networks, mapping



3D coordinates to radiance. This approach allows the creation of detailed and realistic visualizations from various perspectives. Another technique is Gaussian Splatting, which uses computer graphics processes to distribute scene data over a grid of pixels using a Gaussian function, resulting in a smooth and continuous representation of the scene, without directly relying on neural networks or machine learning.

Keywords: Rendering, Radiation fields, Synthetic database, Gaussian Splatting, Digital agriculture, Fruit growing.



1. INTRODUÇÃO

Como discutido em trabalho anterior (Reis et al, 2023), na área do aprendizado de máquina supervisionado, a tarefa de anotar bases de dados é conhecida por sua extensão e esgotamento, especialmente quando se trata de imagens agrícolas. Uma estratégia promissora para otimizar esse processo é a construção de bases de dados sintéticas, já explorada anteriormente.

Este trabalho propõe uma abordagem que permite a geração de dados sintéticos a partir dos dados reais, fazendo uso de tecnologias atuais para expandir a base de dados. A prática de utilização de dados sintéticos, misturados a dados reais já demonstrou eficácia em contextos semelhantes, como na detecção de pessoas e veículos (Farzam et al., 2019).

A proposta deste estudo reside na exploração de técnicas avançadas de renderização, que empregam imagens reais da cena e dados de posicionamento de câmera para recriar graficamente a cena, permitindo a geração de novas visualizações a partir de perspectivas diversas, expandindo assim a base de dados disponível.

O primeiro método examinado é o NeRF (Neural Radiance Fields), uma técnica que se destaca por sua capacidade de modelar cenas complexas através de redes neurais. O NeRF funciona ao treinar uma rede neural para aprender uma função contínua que mapeia coordenadas 3D para radiância, permitindo assim a geração de novas visualizações da cena a partir de diferentes perspectivas. Sua abordagem baseada em inteligência artificial oferece uma maneira eficiente e poderosa de recriar ambientes tridimensionais detalhados e realistas.

Por outro lado, o Gaussian Splatting é uma técnica de renderização que se baseia em processos de computação gráfica. Neste método, os dados da cena são distribuídos sobre uma grade de pixels usando uma função gaussiana, resultando em uma representação suave e contínua da cena. Embora não dependa diretamente de redes neurais ou aprendizado de máquina, o Gaussian Splatting oferece uma abordagem robusta e eficaz para a renderização de cenas tridimensionais.

A análise comparativa desses dois métodos, baseada em pesquisas anteriores e em treinamentos com dados da própria Embrapa, revelou que o método de Gaussian Splatting produziu imagens de qualidade superior ao NeRF. No entanto, é importante destacar que o método NeRF procurou representar os arredores da imagem, mesmo com pouca informação disponível.

2. MATERIAL E MÉTODOS

Visando a criação de cenas 3D através de campos de radiação, representação da distribuição de luz em um espaço tridimensional que explica a interação da luz com o ambiente foram comparadas a performance de diferentes métodos utilizando o software Nerfstudio, programa desenvolvido para gerar ambientes 3D.

Para treinar e comparar o desempenho dos métodos de renderização, foram utilizados dados de uma base de imagens de laranjeiras e anotações previamente geradas pela Embrapa (Base de dados MOrangeT), disponível publicamente. Especificamente, foram explorados os modelos "nerfacto", uma variante do Nerf (Neural Radiance Field), e "splacto", que se baseia na técnica de Gaussian Splatting.

2.1 Nerfacto

Métodos Nerf no geral representam cenas como campos de radiação neural para a criação de visualizações. Partindo de um conjunto de imagens e das coordenadas de perspectiva da câmera, os métodos NERF permitem gerar novas visualizações da cena. Eles fazem isso, mapeando um cenário 3D através do processamento de coordenadas 5D, que incluem a localização espacial e a direção de visualização conforme a Equação 1.

$$(x, y, z, \theta, \varphi) \rightarrow F \rightarrow (RGB\sigma) \quad (1)$$

Para cada entrada o método retorna uma densidade de volume e a radiância emitida, quantidade de energia luminosa que passa, é emitida ou refletida de uma superfície por unidade de área por unidade de ângulo sólido por unidade de tempo (Figura 1). Dependendo da visão naquele ponto espacial específico

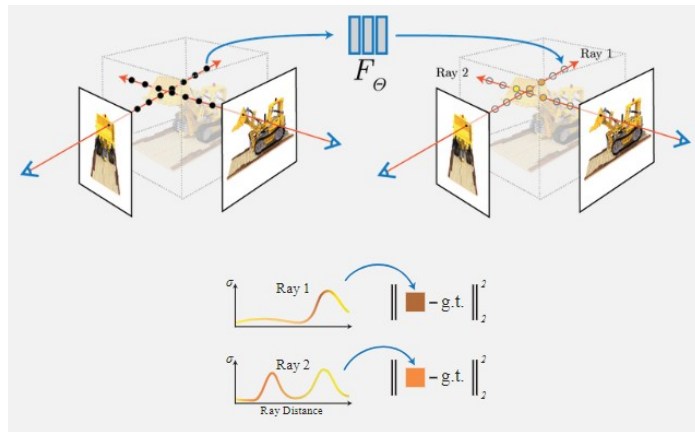


Figura 1. Representação visual do mapeamento da cena a partir das 5 coordenadas

O algoritmo utiliza técnicas clássicas de renderização de volume para projetar cores e densidades na imagem. Como a renderização de volume é naturalmente diferenciável, é

necessário apenas um conjunto de imagens e as respectivas matrizes de câmera (que contém informações de translação e rotação da câmera em relação à origem da cena) conforme as Equações 2 e 3.

$$K = \begin{bmatrix} f_x & s & x_0 \\ 0 & f_y & y_0 \\ 0 & 0 & 1 \end{bmatrix} \quad (2)$$

$$P = K [R | t] \quad (3)$$

Mais especificamente, o método Nerfacto realiza uma série de métodos para aperfeiçoar o processo de geração de campos de radiação neural. Começando pelo refinamento de pose para evitar resultados embaçados e com perda de detalhes, a estrutura do método usa da estratégia de retropropagação, método de redes neurais onde as camadas são acessadas de trás pra frente, a partir do *output* e sua diferença com os dados reais para otimização na atribuição dos pesos nas funções utilizadas na rede neural.

Após o refinamento de pose os dados são divididos por um amostrador que divide as amostras por partes, a primeira parte é distribuída uniformemente até uma distância específica da câmera (por exemplo, uma distância de 1 unidade). Isso garante que as amostras sejam concentradas perto da câmera, proporcionando detalhes suficientes para objetos mais próximos. A segunda parte das amostras é distribuída com tamanho de passo crescente. Isso significa que a distância entre as amostras aumenta à medida que se afastam da câmera. Tal abordagem é útil para capturar objetos mais distantes sem sobrecarregar a quantidade de amostras, garantindo eficiência no uso dos recursos computacionais.

Novamente é realizada uma amostragem, dessa vez o amostrador de proposto consolida a localização das amostras para as regiões da cena que mais contribuem com a renderização, através de uma função de densidade para a cena, aumentando sua qualidade.

Por fim os dados são passados para um campo de densidade onde há uma representação grosseira da cena que guia a amostragem conforme a Figura 2.

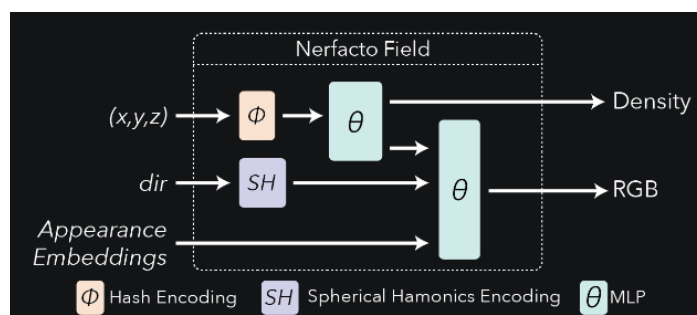


Figura 2. Arquitetura do método nerfacto (Documentação nerfacto)

2.2 Gaussian Splatting

Neste método as cenas são renderizadas através de Gaussianas tridimensionais caracterizadas pela média μ (posição central), opacidade α , matriz de covariância Σ e cor c , computadas através de retropropagação.

Primeiro é realizado uma seleção das informações fornecidas pelas imagens que serão utilizadas, excluindo o que está fora do frustum da câmera, a forma trapezoidal que representa a região visível da cena a partir da perspectiva da câmera. O que economiza recursos computacionais, já que as gaussianas que não estão visíveis na imagem final não precisam ser processados.

Posteriormente é realizado o “Splatting”, etapa onde as gaussianas começam a ser processadas através da projeção de gaussianas tridimensionais no espaço bidimensional das imagens, computando a matriz de covariância bidimensional projetada. Então é calculada a distância de cada pixel para todas as gaussianas sobrepostas ordenadas por profundidade, o que permite a determinação da cor final do pixel com base na cor e opacidade de cada gaussianas.

Uma forte característica do Gaussian Splatting é a computação paralela. No processo de renderização dos pixels a imagem é dividida em trechos sem sobreposição e gaussianas que cobrem diferentes trechos são replicadas. Assim, ocorre uma renderização paralela, onde cada gaussianas é pareada com seu trecho de pixels.

Aprofundando mais nos mecanismos do modelo, para inicializar o modelo são utilizados um conjunto inicial de pontos esparsos e então é feita densificação e seleção com esses pontos.

A densificação dos pontos é crucial para melhorar a qualidade da reconstrução da cena, pois ajusta a densidade das gaussianas para capturar melhor os detalhes da cena. Durante este processo, o algoritmo identifica áreas na cena onde estão faltando detalhes ou onde as gaussianas estão muito espalhadas para representar adequadamente a geometria da cena. São realizadas clonagens e/ou divisões das gaussianas para alcançar uma distribuição ótima na cena, com o objetivo de que cada detalhe importante seja adequadamente capturado e representado.

Posteriormente é feita a seleção dos pontos, que é essencial para remover gaussianas redundantes e garantir que o modelo permaneça preciso e eficaz na representação da cena. O que é feito através do descarte de gaussianas transparentes ou excessivamente grandes, além do controle da densidade das gaussianas que estiverem próximas às posições das câmeras. Proporcionando uma representação mais eficiente da cena.

Para o ajuste dos parâmetros que caracterizam as gaussianas tridimensionais é realizada uma otimização. Essencialmente, a otimização visa ajustar os parâmetros das gaussianas para que a renderização resultante se aproxime o máximo possível da imagem real, e isso é alcançado otimizando os parâmetros através de uma função de perda e garantindo que os gradientes sejam calculados de forma eficiente para evitar custos computacionais desnecessários.

Todas essas etapas resultam no funcionamento do Gaussian Splatting conforme a Figura 3

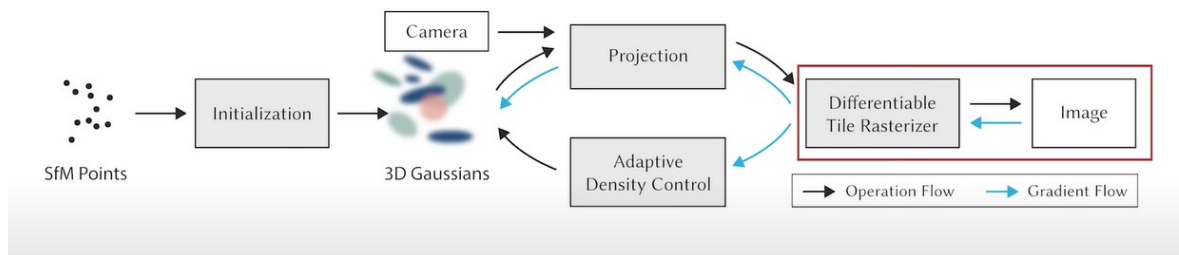


Figura 3. Arquitetura do Gaussian Splatting (Kerbl, et al. 2023)

3. RESULTADOS E DISCUSSÃO

Buscando entender o melhor método para renderização do cenário utilizado para o treinamento de redes neurais convolucionais de detecção de frutos, foram comparados resultados de pesquisas e de treinamentos realizados com dados próprios, através do NerfStudio.

A avaliação da performance do Gaussian Splatting, testada em diferentes cenários renderizados em tempo real (Kerbl, et al. 2023), foi dividida em índice de medida da similaridade estrutural (SSIM), relação sinal-ruído (PSNR) e frames por segundo (FPS), conforme a Figura 4

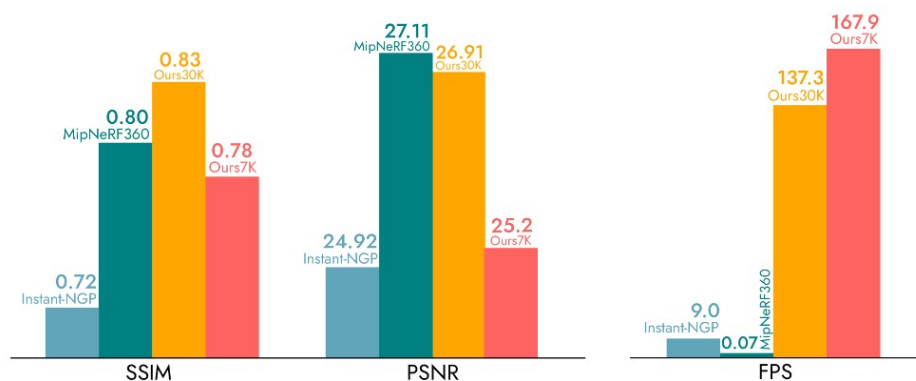


Figura 4. Performance do método Gaussian Splatting com 7 mil iterações (Ours7K) e 30 mil iterações (Ours30K) Comparado a outros métodos.

É possível ver como o método no geral performa melhores que outros, incluindo um tipo de metodologia Nerf “MipNeRF360”, contudo não há comparação com o “nerfacto” (método que ainda não possui publicação).

Utilizando base de dados própria, coletada anteriormente foram treinados os métodos explorados nesse trabalho para entender suas diferenças e buscar o melhor resultado, conforme mostrado nas Figuras 5 e 6, que exibem imagens da cena gerada no treinamento do Nerfacto, e nas Figuras 7 e 8, que mostram imagens da cena gerada no treinamento do Gaussian Splatting.



Figuras 5 e 6. Imagens da cena gerada no treinamento do Nerfacto



Figuras 7 e 8. Imagens da cena gerada no treinamento do Gaussian Splatting

Conforme esperado pela avaliação feita com o Gaussian Splatting (Kerbl, et al. 2023), as imagens geradas apresentaram maior nitidez e menos ruído do que as produzidas pelo “nerfacto”. No entanto, o método de Gaussian Splatting não conseguiu gerar informações detalhadas sobre os arredores dos objetos.

Nas Figuras 5 e 6, observamos que o método Nerfacto tenta gerar a cena como um todo, expandindo seus arredores, embora com menor nitidez. Já nas Figuras 7 e 8, a resolução é maior, mas a cena é reduzida. Essa diferença é resultado da poda realizada pelo método Gaussian Splatting, visando reduzir o custo computacional e gerar apenas regiões com informação suficiente.

4. CONCLUSÃO

Este estudo propõe uma abordagem inovadora para otimizar o processo de renderização de cenas utilizadas no treinamento de redes neurais convolucionais para detecção de frutos. Exploramos duas técnicas avançadas de renderização, o Nerf (Neural Radiance Fields) e o Gaussian Splatting, com o objetivo de expandir a base de dados disponível e aprimorar os resultados de detecção de objetos em imagens agrícolas.

Ao comparar a performance do Gaussian Splatting com outros métodos, observamos que ele se destaca em relação à nitidez e à redução de ruído, porém, não conseguiu gerar informações detalhadas sobre os arredores dos objetos. Por outro lado, o método Nerf tenta representar a cena como um todo, expandindo seus arredores, mas com menor nitidez.

Os resultados obtidos demonstram a viabilidade e a eficácia das técnicas de renderização estudadas para a geração de dados sintéticos a partir de dados reais, contribuindo para otimizar o processo de anotação de bases de dados em contextos agrícolas.

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.

6. REFERÊNCIAS

Base de dados MOrangeT disponível em:

<https://www.redape.dados.embrapa.br/dataset.xhtml?persistentId=doi:10.48432/OI7BFG>

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.

CHEN, Guikun; WANG, Wenguan. A survey on 3d gaussian splatting. arXiv preprint arXiv:2401.03890, 2024.

Documentação do método Nerfacto disponível em:

<https://docs.nerf.studio/nerfology/methods/nerfacto.html>



DOS REIS, Sara Andrade Veronese; SANTOS, T. T. Bases de dados sintéticas com anotação automática para detecção de frutos.

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.

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

KERBL, Bernhard et al. 3D Gaussian Splatting for Real-Time Radiance Field Rendering. ACM Trans. Graph., v. 42, n. 4, p. 139:1-139:14, 2023.

MILDENHALL, Ben et al. Nerf: Representing scenes as neural radiance fields for view synthesis. Communications of the ACM, v. 65, n. 1, p. 99-106, 2021.