

CATS: Classification Aided Transport Splatting for Energy-Efficient 3D Reconstruction

Paulo Abner A Mesquita¹, Leonardo Rocha¹

¹Instituto Atlântico – Fortaleza – CE – Brazil
ALIA – Laboratório de Inteligência Artificial

{abner, leonardo_rocha}@atlantico.com.br

Abstract. *The demand for 3D reconstruction imposes high energy costs, necessitating Green AI strategies. This work proposes the CATS method, which utilizes Optimal Transport (OT) over semantic CLIP embeddings to select the most informative images for 3D Gaussian Splatting (3DGS). Monitoring consumption (Wh) and CO₂ emissions via the CodeCarbon library, we compared OT against random selection across different data volumes. OT reduces consumption in dense datasets by eliminating redundancies, with a trade-off of approximately 5 dB PSNR. In sparse scenarios (10–30%), low geometric connectivity makes the method less robust, generating emissions up to 40% higher than random selection.*

Resumo. *A demanda por reconstrução 3D impõe altos custos energéticos, exigindo estratégias de Green AI. Este trabalho propõe o método CATS, que utiliza Transporte Ótimo (OT) sobre embeddings semânticos do modelo CLIP para selecionar as imagens mais informativas para o 3D Gaussian Splatting (3DGS). Monitorando o consumo (Wh) e a emissão de CO₂ via biblioteca CodeCarbon, comparamos o OT contra a seleção aleatória em diferentes volumes de dados. O OT reduz o consumo em datasets densos eliminando redundâncias, com trade-off de aproximadamente 5 dB de PSNR. Em cenários de escassez (10–30%), a baixa conectividade geométrica torna o método menos robusto, gerando emissões até 40% superiores à seleção aleatória.*

1. Introdução

A transformação digital tem redefinido a maneira como interagimos com o mundo físico, impulsionando demandas crescentes por representações tridimensionais fotorrealistas para aplicações em Gêmeos Digitais, preservação do patrimônio cultural, Realidade Virtual (VR) e Cidades Inteligentes. Neste contexto, técnicas de reconstrução 3D, como a fotogrametria e, mais recentemente, o *3D Gaussian Splatting* (3DGS) [Kerbl et al. 2023], tornaram-se fundamentais. No entanto, a execução desses algoritmos em larga escala impõe um custo computacional severo, agravando o consumo energético de *Data Centers* e contribuindo diretamente para a pegada de carbono do setor tecnológico.

Em um cenário global de emergência climática, a sustentabilidade da Inteligência Artificial (*Green AI*) deixa de ser um diferencial para se tornar um requisito mandatório. O treinamento de modelos profundos e o processamento de nuvens de pontos massivas [Thengane et al. 2025] demandam horas de uso intensivo de GPUs (Graphics

Processing Unit) de alto desempenho resultando em uma pegada de carbono considerável[Strubell et al. 2019]. Frequentemente, a coleta de dados para essas reconstruções gera uma redundância excessiva de imagens que pouco contribuem para a qualidade final da geometria. Conforme demonstrado no trabalho Towards Internet-scale multi-view stereo [Furukawa et al. 2010], algoritmos de reconstrução multi-vistas atingem eficiência e precisão utilizando apenas um subconjunto otimizado de imagens para cada ponto da superfície. Dessa forma, a manutenção de dados redundantes escala o esforço computacional e o impacto ambiental sem ganhos proporcionais em fidelidade.

Embora métodos recentes como o LAVA (Locality-Aware Variable Associations)[Just et al. 2023] e o SAVA (Scalable Learning-Agnostic Data Valuation)[Kessler et al. 2024] tenham explorado a seleção de dados para otimizar a qualidade ou a eficiência de treinamento em Neural Radiance Fields (NeRFs)[Mildenhall et al. 2020] e modelos de difusão, ainda há uma lacuna na literatura sobre o impacto direto dessas escolhas na eficiência energética de pipelines baseados em 3DGS. A premissa deste trabalho é que processar todas as imagens disponíveis é ambientalmente insustentável e computacionalmente ineficiente.

Este trabalho propõe o **CATS** (*Classification Aided Transport Splatting*), uma abordagem que analisa o *trade-off* entre fidelidade visual e custo ambiental na reconstrução de cenas 3D. Diferente de abordagens tradicionais que utilizam todo o conjunto de dados ou realizam descartes aleatórios, investigamos o uso de Transporte Ótimo (*OT*) para classificar e selecionar as imagens mais informativas, reduzindo a redundância dos dados de entrada.

As principais contribuições deste artigo são:

1. Um comparativo detalhado do consumo energético (Wh) e da emissão de CO₂ equivalente (gCO₂eq) para a execução completa do pipeline de reconstrução, Structure-from-Motion (SfM) + 3DGS, monitorado via *CodeCarbon* [Fischer 2025, Courty et al. 2024];
2. Uma análise da eficiência do método de seleção via *OT* versus seleção aleatória em diferentes cardinalidades de dados;
3. A demonstração de que a curadoria inteligente de dados pode viabilizar a transformação digital de ambientes físicos com uma pegada de carbono significativamente reduzida, alinhando-se aos princípios de computação sustentável.

O restante deste artigo está organizado da seguinte forma: a Seção 2 revisa os trabalhos relacionados em SfM, 3DGS e métricas de sustentabilidade. A Seção 3 detalha a metodologia proposta e o ambiente de experimentação. A Seção 4 apresenta os resultados quantitativos de energia e qualidade visual, seguidos pelas conclusões e trabalhos futuros na Seção 5.

2. Trabalhos Relacionados

2.1. Reconstrução 3D: SfM e 3D Gaussian Splatting

A reconstrução 3D multimagem baseia-se em *Structure-from-Motion* (SfM) [Schönberger and Frahm 2016], cujo fluxo incremental envolve extração de *features*, *matching* de pares e triangulação refinada por *Bundle Adjustment*. O sistema COLMAP [Schönberger and Frahm 2016] é o padrão para essa tarefa, provendo poses de

câmera precisas e nuvens de pontos esparsas. Na etapa de renderização, o *3D Gaussian Splatting* (3DGS) [Kerbl et al. 2023] introduz uma abordagem explícita em alternativa aos NeRFs [Mildenhall et al. 2020], modelando a cena por gaussianas 3D definidas por posição, opacidade, harmônicos esféricos (cor) e covariância fatorada em rotação e escala. Sua renderização em tempo real e treino rápido decorrem de um pipeline diferenciável baseado em *splatting* com ordenação paralela e *alpha-blending*. O treinamento do 3DGS otimiza esses parâmetros e aplica o Controle de Densidade Adaptativo (*cloning/splitting*). Como a convergência e o custo computacional do 3DGS dependem estritamente da inicialização geométrica gerada pelo SfM, a seleção inteligente de imagens torna-se crítica, fundamentando o método CATS no equilíbrio entre fidelidade visual e eficiência energética.

2.2. Optimal Transport (OT)

A teoria do *OT* fornece uma estrutura matemática rigorosa para comparar medidas de probabilidade, permitindo quantificar o esforço mínimo necessário para transformar uma distribuição de dados em outra [Peyré et al. 2019]. Na visão computacional, o OT tem sido amplamente adotado devido à sua capacidade de capturar a geometria subjacente dos dados, superando métricas tradicionais (como a Divergência de Kullback-Leibler) que muitas vezes falham quando as distribuições não têm suporte sobreposto [Arjovsky et al. 2017].

A métrica central derivada desta teoria é a Distância de Wasserstein [Peyré et al. 2019] (ou *Earth Mover's Distance* [Rubner et al. 2000]). Intuitivamente, se considerarmos as características visuais das imagens como "massas" em um espaço de alta dimensão, a Distância de Wasserstein mede o custo para mover a massa de uma distribuição empírica (o conjunto de dados completo) para outra (o subconjunto selecionado).

Recentemente, abordagens como SAVA [Kessler et al. 2024] e LAVA [Just et al. 2023] demonstraram que o OT pode ser utilizado efetivamente para a curadoria e valorização de dados (data valuation). Ao formular a seleção de imagens como um problema de transporte, é possível identificar amostras que afetam negativamente a qualidade do modelo, comparando um conjunto de dados amplo com um pequeno subconjunto de referência curado (limpo).

Neste trabalho, criamos um ranking de importância das imagens via transporte ótimo, avaliando o impacto de cada amostra frente a uma distribuição de referência ideal. Isso garante que, mesmo sob redução de dados, o subconjunto preserve a qualidade, diversidade angular e semântica essenciais para a reconstrução 3D, impedindo o descarte de vistas críticas ao SfM ao contrário de uma seleção aleatória.

3. Metodologia

A metodologia proposta investiga o equilíbrio entre a eficiência energética (emissões de CO₂) e a qualidade visual na reconstrução via 3DGS, analisando como a cardinalidade e a seleção de imagens impactam o custo energético. O fluxo de trabalho, sintetizado na Figura 1, integra a seleção inteligente de dados, o pré-processamento SfM via COLMAP e o treinamento via 3DGS. Para os testes, utilizamos um conjunto S de 288 imagens (4320×7680 px), particionado em subconjuntos $S_p = \lceil n \cdot p \rceil$ com taxas de amostragem

$p \in \{0.1, \dots, 1.0\}$, comparando a seleção aleatória (D_s) com a ordenação baseada em qualidade (D_{ot}).

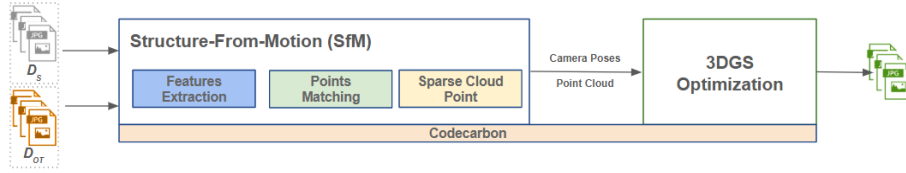


Figura 1. Pipeline de Reconstrução 3D para Gaussian Splatting

A curadoria de D_{ot} utiliza a metodologia de ranqueamento SAVA [Kessler et al. 2024], na qual extraímos características semânticas das imagens via modelo CLIP (ViT-B/32) [Radford et al. 2021] para gerar *embeddings* vetoriais. O cálculo do custo de transporte é estruturado de forma hierárquica, utilizando o algoritmo de Sinkhorn [Cuturi 2013] e *Earth Mover's Distance* [Rubner et al. 2000] para definir planos de transporte entre distribuições de imagens ruidosas e de referência. Desse processo, extraem-se os *SAVA scores*, que permitem ordenar as imagens de forma que os subconjuntos reduzidos priorizem amostras com maior integridade visual e semântica, otimizando a convergência do modelo 3DGS.

Os experimentos foram conduzidos em um ambiente de alto desempenho equipado com CPU AMD EPYC-Genoa e GPU NVIDIA H100L (96 GiB). O perfilamento temporal foi realizado de forma modular, cronometrando isoladamente: (i) o cálculo dos *SAVA scores*, (ii) o processamento SfM e (iii) o ciclo de treinamento do 3DGS.

Utilizou-se a biblioteca *CodeCarbon* [Courty et al. 2024] para a monitoração energético e de emissão de CO_2 integrado à interface *nvidia-smi*, permitindo medir a potência e o consumo (Wh) em tempo real. A estimativa de gCO_{2eq} foi calculada aplicando fatores de regionalidade baseados na matriz energética local (identificada via IP/configuração da ferramenta). A fidelidade visual foi aferida pelas métricas PSNR e SSIM.

4. Resultados e Discussão

A análise experimental revela que o perfil de consumo de potência mantém um comportamento qualitativo semelhante em todos os subconjuntos $S_i \in \{D_s, D_{ot}\}$, com estabilidade térmica da GPU operando estritamente entre 62°C e 84°C. Contudo, o tempo de execução e a eficiência energética são sensivelmente influenciados pela densidade de seleção das imagens.

Notadamente, as etapas de pré-processamento de *Structure-from-Motion* (SfM), executadas via COLMAP [Schönberger and Frahm 2016], possuem um impacto ambiental significativamente menor em comparação ao treinamento volumétrico do 3DGS, que representa a maior parcela da pegada de carbono do pipeline. O tempo de processamento do COLMAP correlaciona-se fortemente com a redução do número de imagens; embora a complexidade teórica do *exhaustive matching* seja quadrática $O(n^2)$, a utilização de estratégias de correspondência seletiva e a redução drástica no volume de dados permitiram uma economia de tempo substancial. Em contrapartida, a fase de treinamento do 3DGS demonstra maior variabilidade, sugerindo que a convergência do modelo depende não apenas da quantidade, mas da relevância das amostras selecionadas via D_{ot} .

A emissão de CO_2 mostra-se proporcional à densidade de dados de entrada, dado que o volume de imagens dita o tempo de ocupação da GPU no pipeline geométrico e na otimização radiométrica. O processo fixo de seleção por Transporte Ótimo (OT) totalizou 463,44 segundos, com emissão acumulada de $3,59 \times 10^{-3}$ kg de CO_2eq . Esta etapa é amplamente dominada pela extração de características via rede neural (CLIP), enquanto o cálculo dos *scores* (SAVA) apresenta impacto desprezível ($< 0,1\%$ do tempo total). Isolando-se esse custo fixo inicial, os balanços negativos de ΔCO_2 na Tabela 1 indicam que o D_{ot} apresenta maior eficiência ecológica em subconjuntos de densidade alta. Isso ocorre porque a seleção criteriosa reduz o esforço computacional necessário para a convergência das propriedades esféricas harmônicas das *Gaussians*, demonstrando que a qualidade da informação é um fator tão determinante para a sustentabilidade do modelo quanto a sua quantidade bruta.

Tabela 1. Resumo Comparativo de Diferenças ($\Delta = D_{ot} - D_s$): Desempenho, Emissões e Qualidade

Subconjunto (S_i)	Δ Tempo Total (s)	ΔCO_2 Total (kg)	Δ PSNR (dB)	Δ SSIM
S_{10} (100%)	+94.0	+0.0001	-5.82	-0.287
S_9 (90%)	-102.5	-0.0015	-4.80	-0.174
S_8 (80%)	-291.0	-0.0027	-5.04	-0.209
S_3 (30%)	+83168.0	+0.0045	Falha	Falha
S_2 (20%)	-567.5	-0.0037	Falha	Falha
S_1 (10%)	+1098.5	+0.0125	-31.78	-0.701

Nota: Valores calculados como $D_{ot} - D_s$. Sinais negativos em Tempo e CO_2 indicam redução de gastos (vantagem para o D_{ot}), enquanto sinais negativos em PSNR e SSIM indicam a perda de qualidade visual inerente ao trade-off.

5. Conclusões

A investigação estabeleceu três regimes claros de viabilidade para o método proposto com base nos indicadores da Tabela 1. Em datasets de alta cardinalidade (80% a 100%), o método de OT demonstra sua maior utilidade para a *Green AI* ainda que assuma um *trade-off* visual de cerca de 5 dB de PSNR abaixo da seleção aleatória. Nos cenários de densidade intermediária (40% a 70% omitidos da Tabela 1 por restrição de espaço), observa-se um ponto de equilíbrio em ambos os métodos.

Por fim, em regimes de baixo número de imagens (10% a 30%), manifesta-se o ponto de ruptura crítico documentado na Tabela 1, onde o OT perde a sua vantagem competitiva. Nesses casos, a seleção aleatória mostra-se superior em todos os eixos: converge mais rapidamente, gera cerca de 40% menos poluição total e produz resultados de PSNR e SSIM significativamente superiores.

Em suma, a metodologia baseada em Transporte Ótimo é altamente recomendada para mitigar redundâncias e emissões em volumes massivos de dados, enquanto a amostragem aleatória permanece imperativa para cenários de extrema escassez. Para trabalhos futuros, propõe-se explorar métricas de qualidade sem referência (*no-reference*), como NIQE e BRISQUE, como critérios complementares de amostragem.

Referências

- Arjovsky, M., Chintala, S., and Bottou, L. (2017). Wasserstein generative adversarial networks. In *International conference on machine learning*, pages 214–223. PMLR.
- Courty, B., Schmidt, V., Goyal-Kamal, MarionCoutarel, Feld, B., Lecourt, J., LiamConnell, SabAmine, inimaz, supatomic, Léval, M., Blanche, L., Cruveiller, A., ouminasara, Zhao, F., Joshi, A., Bogroff, A., Saboni, A., de Lavoreille, H., Laskaris, N., Abati, E., Blank, D., Wang, Z., Catovic, A., alencon, Bauer, C., Lucas-Otavio, JPW, and MinervaBooks (2024). mlco2/codecarbon: v2.4.1.
- Cuturi, M. (2013). Sinkhorn distances: Lightspeed computation of optimal transport. In *Advances in Neural Information Processing Systems (NeurIPS)*, pages 2292–2300.
- Fischer, R. (2025). Ground-truthing ai energy consumption: Validating codecarbon against external measurements.
- Furukawa, Y., Curless, B., Seitz, S. M., and Szeliski, R. (2010). Towards internet-scale multi-view stereo. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 1434–1441.
- Just, H. A., Kang, F., Wang, T., Zeng, Y., Ko, M., Jin, M., and Jia, R. (2023). LAVA: Data valuation without pre-specified learning algorithms. In *The Eleventh International Conference on Learning Representations*.
- Kerbl, B., Kopanas, G., Leimkühler, T., and Drettakis, G. (2023). 3d gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics*, 42(4).
- Kessler, S., Le, T., and Nguyen, V. (2024). Sava: Scalable learning-agnostic data valuation.
- Mildenhall, B., Srinivasan, P. P., Tancik, M., Barron, J. T., Ramamoorthi, R., and Ng, R. (2020). Nerf: Representing scenes as neural radiance fields for view synthesis.
- Peyré, G., Cuturi, M., et al. (2019). Computational optimal transport: With applications to data science. *Foundations and Trends® in Machine Learning*, 11(5-6):355–607.
- Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., et al. (2021). Learning transferable visual models from natural language supervision. In *International Conference on Machine Learning (ICML)*, pages 8748–8763. PMLR.
- Rubner, Y., Tomasi, C., and Guibas, L. J. (2000). The earth mover’s distance as a metric for image retrieval. *International journal of computer vision*, 40(2):99–121.
- Schönberger, J. L. and Frahm, J.-M. (2016). Structure-from-motion revisited. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4104–4113.
- Strubell, E., Ganesh, A., and McCallum, A. (2019). Energy and policy considerations for deep learning in NLP. In Korhonen, A., Traum, D., and Márquez, L., editors, *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 3645–3650, Florence, Italy. Association for Computational Linguistics.
- Thengane, V., Zhu, X., Bouzerdoum, S., Phung, S. L., and Li, Y. (2025). Foundational models for 3d point clouds: A survey and outlook.