

WasteInsight: Conjunto de Dados para Detecção e Estimativa de Volume de Resíduos Sólidos Urbanos

Julio Leite Azancort Neto¹, Romário da Costa Silva¹, Thalita Ayass de Souza¹,
Carlos André de Mattos Teixeira¹, Evelin Helena Silva Cardoso¹, Jasmine Priscyla
Leite de Araújo¹, Carlos Renato Lisboa Francês¹

Instituto de Tecnologia – Universidade Federal do Pará (UFPA)¹
Belém - PA - Brazil

{julio.azancort.neto, carlos.mattos}@itec.ufpa.br,
romario.silva@castanhal.ufpa.br, thalita_ayass@hotmail.com,
ehs.cardoso@gmail.com, {jasmine, rfrances}@ufpa.br

Resumo. *A gestão eficiente de resíduos sólidos é crucial para tornar a cidade um ambiente limpo e sustentável. Com base nessa afirmação, este artigo apresenta o conjunto de dados "WasteInsight", composto por imagens de sacolas de lixo em ambientes residenciais, acompanhadas de rótulos que indicam as coordenadas das sacolas. O objetivo é fornecer uma base de dados para treinar algoritmos de detecção de objetos e estimativa de volume, visando melhorar a gestão de resíduos sólidos urbanos. O conjunto de dados está disponível gratuitamente no Mendeley Data, promovendo transparência e colaboração na pesquisa científica.*

Abstract. *Efficient solid waste management is crucial to making the city a clean and sustainable environment. Based on this statement, this article presents the "WasteInsight" dataset, composed of images of garbage bags in residential environments, accompanied by labels indicating the bags' coordinates. The objective is to provide a database to train object detection and volume estimation algorithms, aiming to improve the management of urban solid waste. The dataset is freely available on Mendeley Data, promoting transparency and collaboration in scientific research.*

1. Introdução

A geração de resíduos é uma consequência do crescimento populacional, da urbanização e do desenvolvimento econômico. O problema da Gestão de Resíduos Sólidos (GRS) afeta tanto os indivíduos quanto os governos em todo o mundo. O manejo inadequado desses resíduos impacta diretamente a saúde pública e o meio ambiente (Arbeláez-Estrada et al. 2023).

A gestão eficaz dos resíduos sólidos é uma prioridade na agenda contemporânea, dada a crescente intensidade dos desafios ambientais. Nesse contexto, o compartilhamento de conjuntos de dados é crucial para projetos e iniciativas voltadas para essa questão. Ao disponibilizar esses dados de forma gratuita e acessível, não apenas avançamos no campo da gestão ambiental, mas também facilitamos o desenvolvimento de soluções inovadoras, melhorando assim a qualidade de vida das comunidades e promovendo resultados positivos em termos de sustentabilidade.

O compartilhamento de dados na gestão de resíduos sólidos é essencial porque oferece *insights* valiosos baseados em evidências concretas, que podem orientar políticas

e práticas. Além disso, o acesso transparente a dados confiáveis pode impulsionar a inovação e o desenvolvimento de novas tecnologias e abordagens na gestão de resíduos. Isso permite que pesquisadores, acadêmicos, empresas e organizações da sociedade civil explorem e analisem os dados disponíveis, estimulando a colaboração e o intercâmbio de conhecimentos para soluções mais eficazes e sustentáveis.

A distribuição de conjuntos de dados sobre resíduos sólidos não é uma ideia nova. O conjunto "*Taco*" de Proença e Simões (2020) tem 1.500 imagens em 28 categorias e 60 subcategorias. Já o conjunto "*Open Litter Map*" de Lynch (2018) contém mais de 100 mil imagens distribuídas em 11 categorias principais e 187 subcategorias. Até o lançamento do "*WasteInsight*", não foram identificados conjunto de dados que refletissem a frequente presença de resíduos em sacos plásticos, comumente utilizados para descarte no Brasil.

A disponibilização gratuita e acessível de conjuntos de dados é essencial para promover a participação cidadã e responsabilizar os tomadores de decisão na gestão de resíduos sólidos. Isso permite que as comunidades se envolvam mais ativamente, exijam transparência e contribuam para a implementação de medidas eficazes ao entenderem claramente as informações sobre a gestão de resíduos.

Um exemplo disso é a situação em Belém, capital do estado do Pará, onde apesar da Política Nacional de Resíduos Sólidos exigir o fim dos lixões em todo o Brasil, o aterro localizado no município de Marituba ainda recebe resíduos sólidos da capital e de sua região metropolitana, incluindo Ananindeua, Benevides, Santa Bárbara e Marituba. Esses municípios juntos produzem cerca de 2 mil toneladas de lixo por dia (Brito et al., 2020).

Apenas um quarto dos resíduos sólidos municipais coletados globalmente são adequadamente tratados, tornando essencial estabelecer métodos eficazes de gestão de resíduos para garantir a reciclagem e tratamento adequado, especialmente devido à variação nas práticas entre os países (Nanda e Berruti 2021).

É crucial ressaltar que o compartilhamento desses dados vai além de uma questão técnica, sendo também um imperativo ético e moral. Em face dos desafios urgentes da crise climática e da degradação ambiental, é essencial adotar uma abordagem colaborativa e inclusiva, valorizando a transparência, a cooperação e o engajamento público. Priya et al. (2018) argumentam em seu artigo que os departamentos ambientais têm negligenciado a devida atenção à abordagem matemática do problema da coleta e disposição de resíduos.

Em suma, o compartilhamento de conjuntos de dados na gestão de resíduos sólidos representa não apenas uma necessidade atual, mas também uma oportunidade única para promover a sustentabilidade ambiental, melhorar a qualidade de vida das comunidades e impulsionar resultados positivos em escala global. Ao trabalhar para tornar os dados mais acessíveis e utilizáveis, é possível avançar significativamente em direção a um futuro mais limpo, saudável e sustentável para todos.

2. Objetivo

O objetivo principal deste artigo é disponibilizar um conjunto de dados rotulados que possam ser utilizados para treinar algoritmos de detecção de objetos capazes de identificar sacolas de lixo em ambientes residenciais e estimar seus volumes com base em suas dimensões. Ao fornecer esse conjunto de dados, pretende-se facilitar o desenvolvimento e aprimoramento de sistemas automatizados para a gestão de resíduos sólidos urbanos. Espera-se que esse recurso contribua significativamente para o avanço da tecnologia

aplicada à gestão de resíduos, promovendo soluções mais eficientes e sustentáveis para os desafios enfrentados em ambientes urbanos.

3. Material e método

O *dataset* contém 447 imagens, divididas em 60% para treinamento, 20% para validação e 20% para testes. As imagens foram coletadas em residências, utilizando câmeras de dispositivos móveis, capturando sacolas de lixo em diferentes ambientes e condições de iluminação, Figura 1.



Figura 1. Exemplo de imagens presentes no *dataset*

Cada imagem no conjunto de dados está acompanhada de rótulos que indicam as coordenadas das caixas delimitadoras ao redor das sacolas de lixo, Figura 2. O processo de rotulação das imagens foi realizado manualmente através da plataforma *Roboflow*, a qual fornece uma ampla variedade de ferramentas otimizadas para esse tipo de tarefa (Dwyer et al., 2024). Essas caixas delimitadoras são usadas para identificar e segmentar as sacolas nas imagens. Os rótulos incluem as coordenadas do ponto superior esquerdo e do ponto inferior direito da caixa delimitadora, permitindo que os algoritmos de detecção de objetos localizem e isolem as sacolas de lixo.



Figura 2. Imagens do *dataset* com a presença de caixas delimitadoras

As imagens variam em termos de resolução, orientação e condições de iluminação. Elas foram capturadas em diferentes ambientes residenciais, como cozinhas, áreas de serviço, entre outros. Além disso, a rotulação manual foi imprescindível visto que as sacolas de lixo podem apresentar diferentes cores, texturas e tamanhos, refletindo a diversidade encontrada em ambientes domésticos reais.

4. Disponibilidade do Dataset

É fundamental ressaltar que o conjunto de dados "WasteInsight" está disponível no site Mendeley Data, sendo totalmente gratuito e de código aberto. Os pesquisadores

interessados podem acessar este recurso por meio do Mendeley Data e baixá-lo para utilização em seus projetos (Azancort Neto et al., 2024). Este conjunto de dados aberto e acessível visa promover a transparência, replicabilidade e colaboração na comunidade científica, permitindo o avanço do conhecimento e o desenvolvimento de soluções inovadoras na área da gestão de resíduos sólidos urbanos.

5. Conclusão

Em síntese, o conjunto de dados proposto oferece uma valiosa oportunidade para aprimorar algoritmos de detecção de objetos e estimativa de volume, contribuindo assim para soluções mais eficientes na gestão de resíduos sólidos urbanos. Sua disponibilidade gratuita e de código aberto no Mendeley Data não apenas promove a transparência e a replicabilidade na pesquisa científica, mas também estimula a colaboração entre os pesquisadores. Espera-se que o uso desse conjunto de dados leve a avanços significativos no desenvolvimento de tecnologias que abordam os desafios ambientais e urbanos enfrentados pelas comunidades em todo o mundo. Em estudos futuros, planeja-se expandir o dataset com mais imagens, obtidas via múltiplos dispositivos, e novas classes, como entulho e podas de árvore. Além disso, pretende-se disponibilizar modelos de detecção baseados no dataset, incentivando adaptações regionais e climáticas para melhorar a generalização das soluções.

Referências

Arbeláez-Estrada, J. C., Vallejo, P., Aguilar, J., Tabares-Betancur, M. S., Ríos-Zapata, D., Ruiz-Arenas, S., & Rendón-Vélez, E. (2023). A Systematic Literature Review of Waste Identification in Automatic Separation Systems. *Recycling*, 8(6), 86. <https://doi.org/10.3390/recycling8060086>.

Azancort Neto, Julio; Silva, Romário; Souza, Thalita; Teixeira, Carlos; Cardoso, Evelin; Araújo, Jasmine; Francês, Carlos (2024), “WasteInsight: Dataset for Detection and Volume Estimation of Municipal Solid Waste”, Mendeley Data, V1, doi: 10.17632/p8n7nbxyw3.1.

Brito, D. A. C., Seabra, L. C., Lima, P. D. M., & Souza, C. M. N. (2020). Manejo de Resíduos Sólidos e de Águas Pluviais: O (Des)Controle Social em Belém, Pará. *Revista Eletrônica de Gestão E Tecnologias Ambientais*, 8(2), 103. <https://doi.org/10.9771/gesta.v8i2.42221>.

Dwyer, B., Nelson, J., Hansen, T., et. al. (2024). Roboflow (Version 1.0) [Software]. Available from <https://roboflow.com>. [computer vision](https://roboflow.com).

Lynch, S. OpenLitterMap. com—open data on plastic pollution with blockchain rewards (littercoin). *Open Geospatial Data, Software and Standards*, v. 3, n. 1, p. 1-10, 2018.

Nanda, S.; Berruti, F. (2021). Municipal solid waste management and landfilling technologies: a review. *Environmental chemistry letters*, v. 19, n. 2, p. 1433-1456.

Priya, K.S., Burman, I., Tarafdar, A., & Sinha, A. (2018). Impact of Ammonia Nitrogen on COD Removal Efficiency in Anaerobic Hybrid Membrane Bioreactor Treating Synthetic Leachate. *International Journal of Environmental Research*, 13, 59-65.

Proença, P. F.; Simões, P. Taco: Trash annotations in context for litter detection. arXiv preprint arXiv:2003.06975, 2020.