

Kumpulan Dataset Citra Sampah (Anorganik) untuk YOLO

Beberapa *dataset* citra sampah (garbage) tersedia secara publik dan cocok untuk melatih model deteksi/ klasifikasi sampah anorganik menggunakan YOLO. Dataset ini umumnya mencakup kategori seperti plastik, kaca, logam, kertas/karton, dan lainnya. Berikut beberapa dataset yang dapat diunduh beserta tautan sumbernya:

- **TACO (Trash Annotations in Context):** Dataset terbuka berisi ~1500 gambar sampah "in the wild" (diambil di berbagai lingkungan: hutan, jalanan, pantai, dll.) terbagi dalam 28 kategori ¹. Gambar-gambar tersebut diberi anotasi objek (bounding box) dan segmentasi (format COCO) ² ¹. Data TACO dapat diunduh dari situs resminya (tacodataset.org) ¹ atau melalui skrip `download.py` di GitHub TACO ³.
- **TrashNet (Stanford CS229):** Dataset dari proyek kuliah Stanford berisi 2.527 gambar sampah yang dibagi dalam 6 kelas (glass, paper, cardboard, plastic, metal, trash) ⁴. Data asli (~3,5 GB) tersedia melalui repo GitHub pengembang atau HuggingFace. Misalnya, dataset ini dapat diunduh dari [HuggingFace TrashNet](#) ⁵.
- **Garbage Classification (Kaggle):** Dataset berisi 2.467 gambar sampah rumah tangga, terbagi dalam 6 kategori (cardboard, glass, metal, paper, plastic, trash) ⁶. Tersedia di Kaggle pada tautan [Garbage Classification Dataset](#) ⁶.
- **WaDaBa (Plastic Waste Database):** Koleksi 4.000 gambar berbagai jenis plastik (PET, PP, PE-HD, dsb.) dengan anotasi detail tentang tipe plastik, warna, tingkat kerusakan, dsb. ⁷. Gambar-gambar ini bebas diunduh dari situs resmi WaDaBa (walaupun beberapa anotasi memerlukan persetujuan lisensi) ⁷ ⁸. [WaDaBa](#) menyediakan file gambar (dengan registrasi sederhana).
- **Waste Classification (Kaggle, TechSash):** Dataset >25.000 gambar (22.564 train + 2.513 test) sampah *organik* dan daur ulang (recyclable) ⁹. Setiap gambar dikategorikan sebagai "Organic" atau "Recyclable". Data ini dapat diunduh dari [Kaggle Waste Classification \(techsash\)](#) ⁹.
- **Waste Classification Data v2 (Kaggle, Sapal6):** Perpanjangan dataset di atas dengan menambah kategori "Non-Recyclable". Total dataset ≈28.000 gambar (termasuk 2.508 gambar kelas N pada train dan 397 pada test) ¹⁰. Tersedia di [Kaggle Waste Classification Data v2](#) ¹⁰.
- **Open Litter Map:** Kumpulan terbesar dengan >100.000 gambar sampah (11 kategori utama, 187 sub-kategori) ¹¹. Data dikumpulkan dari kontribusi masyarakat dan dapat diakses melalui situs [OpenLitterMap](#) ¹¹.

Catatan: Beberapa dataset seperti TACO sudah menyediakan anotasi bounding-box siap pakai (format COCO) ¹² sehingga mudah untuk digunakan dengan YOLO, sedangkan dataset seperti TrashNet atau Waste Classification di atas bersifat klasifikasi per-gambar (tanpa anotasi kotak). Anda dapat menggabungkan atau mengonversi anotasi agar sesuai format YOLO saat melatih model.

Sumber: Informasi dan tautan dataset di atas dirangkum dari repositori dan publikasi terkait ¹ ², serta sumber resmi masing-masing (misalnya Kaggle atau situs pengembang) ⁴ ⁷. Silakan kunjungi tautan yang tercantum untuk mengunduh dataset.

[1](#) [4](#) [5](#) [6](#) [7](#) [8](#) [9](#) [10](#) [11](#) GitHub - AgaMiko/waste-datasets-review: List of image datasets with any kind of litter, garbage, waste and trash

<https://github.com/AgaMiko/waste-datasets-review>

[2](#) [3](#) [12](#) GitHub - pedropro/TACO: Trash Annotations in Context Dataset Toolkit

<https://github.com/pedropro/TACO>