

## 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 <sup>1</sup>. Gambar-gambar tersebut diberi anotasi objek (bounding box) dan segmentasi (format COCO) <sup>2</sup>.  
<sup>1</sup> . Data TACO dapat diunduh dari situs resminya ([tacodataset.org](https://tacodataset.org)) <sup>1</sup> atau melalui skrip `download.py` di GitHub TACO <sup>3</sup>.
- **TrashNet (Stanford CS229):** Dataset dari proyek kuliah Stanford berisi 2.527 gambar sampah yang dibagi dalam 6 kelas (glass, paper, cardboard, plastic, metal, trash) <sup>4</sup>. Data asli (~3,5 GB) tersedia melalui repo GitHub pengembang atau HuggingFace. Misalnya, dataset ini dapat diunduh dari [HuggingFace TrashNet](#) <sup>5</sup>.
- **Garbage Classification (Kaggle):** Dataset berisi 2.467 gambar sampah rumah tangga, terbagi dalam 6 kategori (cardboard, glass, metal, paper, plastic, trash) <sup>6</sup>. Tersedia di Kaggle pada tautan [Garbage Classification Dataset](#) <sup>6</sup>.
- **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. <sup>7</sup>. Gambar-gambar ini bebas diunduh dari situs resmi WaDaBa (walaupun beberapa anotasi memerlukan persetujuan lisensi) <sup>7</sup> <sup>8</sup>. [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) <sup>9</sup>. Setiap gambar dikategori sebagai “Organic” atau “Recyclable”. Data ini dapat diunduh dari [Kaggle Waste Classification \(techsash\)](#) <sup>9</sup>.
- **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) <sup>10</sup>. Tersedia di [Kaggle Waste Classification Data v2](#) <sup>10</sup>.
- **Open Litter Map:** Kumpulan terbesar dengan >100.000 gambar sampah (11 kategori utama, 187 sub-kategori) <sup>11</sup>. Data dikumpulkan dari kontribusi masyarakat dan dapat diakses melalui situs [OpenLitterMap](#) <sup>11</sup>.

Catatan: Beberapa dataset seperti TACO sudah menyediakan anotasi bounding-box siap pakai (format COCO) <sup>12</sup> 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 <sup>1</sup> <sup>2</sup>, serta sumber resmi masing-masing (misalnya Kaggle atau situs pengembang) <sup>4</sup> <sup>7</sup>. 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>