

Challenge 2 : Dataset Collection

Sumber Dataset yang Digunakan

Untuk membangun sistem deteksi pelanggaran penggunaan Personal Protective Equipment (PPE) dari footage CCTV, saya mengumpulkan total 110 gambar dari tiga sumber berikut:

Sumber Dataset	Jumlah Gambar	Link Dataset
SH17 Dataset (Kaggle)	57	SH17 PPE Dataset
PPE Dataset YOLOv8 (Kaggle)	28	PPE YOLOv8
Google Images	25	Dicari secara manual dengan kata kunci spesifik di Google Image Search

Alasan Pemilihan Masing-masing Sumber

1. SH17 Dataset (57 gambar)

- **Kualitas Visual Tinggi:** Sebagian besar gambar berasal dari platform seperti *Pexels*, menghasilkan gambar yang sangat tajam dan bersih. Ini bermanfaat untuk membantu model mengenali bentuk PPE secara presisi.

- Variasi Objek & Ambiguitas: Banyak gambar menunjukkan objek yang *mirip* dengan PPE, seperti topi biasa, hijab, atau kacamata mode. Ini berguna untuk melatih model agar mampu membedakan PPE dengan non-PPE.
- Distribusi Kategori Baik: Terdapat distribusi yang relatif seimbang antara hard hat, vest, gloves, dan boots, meskipun safety glasses masih terbatas.

2. PPE Dataset YOLOv8 (28 gambar)

- Simulasi CCTV Nyata: Gambar diambil dari sudut pandang tinggi (top-view) dan kualitas rendah, sangat cocok untuk meniru hasil rekaman CCTV di lapangan kerja.
- Variasi Lokasi dan Posisi Tubuh: Banyak gambar memperlihatkan orang dari belakang, samping, dan kondisi separuh badan, bagus untuk menguji robustness model.

3. Google Images (25 gambar)

- Melengkapi Label Minoritas: Gambar dengan safety glasses secara jelas banyak ditemukan melalui pencarian manual, mengisi gap dari dua dataset sebelumnya.
- Gambar-gambar dari Google memiliki kondisi pencahayaan dan resolusi yang sangat bervariasi, termasuk foto dengan efek blur atau noise. Guna meningkatkan ketahanan model terhadap input yang tidak ideal.