Image classification - Data layer

Datasets — Benchmarks & Sources

CIFAR-10 / CIFAR-100

- **What it is:** 60k color images at 32×32 px. CIFAR-10 has 10 coarse classes, CIFAR-100 has 100 fine classes.
- **Why it matters:** Lightweight dataset for fast prototyping, debugging, and teaching.
- **Quirks:** Very low resolution; models overfit quickly; augmentations matter disproportionately.
- **Where:** Hugging Face (`cifar10`, `cifar100`).

ImageNet-1k

- **What it is:** 1.2M images across 1,000 categories. Large-scale benchmark for visual recognition.
- **Why it matters:** Gold standard for pretraining; many pretrained backbones expect ImageNet normalization.
- **Quirks:** Noisy labels, class imbalance, non-curated images.
- **Where:** Hugging Face (`imagenet-1k`) or via Google Cloud bucket (restricted).

COCO

- **What it is:** 330k images with object annotations (captions, bounding boxes, segmentations).
- **Why it matters:** Multi-purpose benchmark for detection, captioning, and VQA.
- **Quirks:** Heavily biased toward everyday objects; captions are short and colloquial.
- **Where:** Hugging Face (`coco_captions`, `coco_detection`).

Preprocessing (what to do and why)

Normalization

We adjust pixel values so they're centered and scaled, making training stable.

- **ImageNet stats:** `Normalize(mean=[0.485, 0.456, 0.406], std=[0.229, 0.224, 0.225])` → Matches the preprocessing expected by most pretrained models.
- **From scratch:** `Standardize per dataset` → If no pretrained weights, just scale to dataset-specific zero mean/unit variance.

Resizing

We make all images the same size so they fit into batches and pretrained backbones.

- **Train:** `RandomResizedCrop(224)` → Ensures the network learns from different object scales.
- **Eval:** `Resize(256) + CenterCrop(224)` → Stable, deterministic input size for validation/testing.

Augmentation

We add random variation to prevent overfitting and make the model generalize better.

- `RandomHorizontalFlip(p=0.5)` → Common for natural images.
- `ColorJitter/AutoAugment` (optional) → Adds robustness to lighting/color variations.
- `CutMix/MixUp` (advanced) → Encourages smoother decision boundaries, especially on small datasets.
