

Step1: Collection & Screening

Source Data



Four Categories

- Identity Recognition
- Re-Identification
- Identity Grounding
- Identity Editing

Quality Control



Dataset1 ✓
Dataset2 ✗
Dataset3 ✓
.....

Step2: Refinement & Mining

Noise Filtering



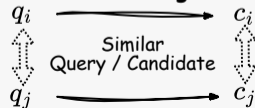
Sample1 ✓
Sample2 ✓
Sample3 ✗
.....

Identity Resampling



Addressing the long-tail distribution of IDs.

Hard Case Mining



Step3: Instruction & Splitting

Task-specific Instruction

Dataset1: Retrieve all images that the same outfit from different views
Dataset2: Find the same object as the one in the given image.
.....

ID-Aware Splitting

Train $ID_1 ID_2 \dots ID_m$

Eval $ID_{m+1} \dots ID_n$