

# VAST-ReID: A Low-Light Benchmark Dataset for Person Re-Identification with Visual and Attribute-Rich Semantic Tracking

## (Supplementary Document)

[VAST-ReID Team]

### 1. Attribute and Sub-Attribute Lists

To facilitate detailed analysis and fine-grained person characterization, we annotate each identity with a comprehensive set of visual and contextual attributes. We have total of 25 attributes which again divided into 181 sub-attributes. These attributes cover appearance (e.g., gender, age, hair, clothing), viewpoint and occlusion conditions, carried objects, accessories, activities, and environmental context. Each attribute is further divided into sub-attributes, capturing variations in style, color, or state.

The full list is provided below:

- **Gender:** Male, Female, Other, NA, NO#
- **Age:** Adult, Teenager, Middle Aged, Child, Old, NA, NO#
- **Hair:** Short, Long, Bald, NA, NO#
- **Hair Color:** Black, Blonde, White, NA, Other, NO#
- **Sleeve Length:** Short, Long, No Sleeve, NA, NO#
- **Viewpoint:** Front, Back, Left, Right, Top, NA
- **Is Occluded? (B):** No, Yes, NA
- **Activity:** Standing, Walking, Running, Seating, Cellphoning, Cycling, Riding kick scooter, Pooling, Dancing, Talking, NA, NO#
- **Body Posture:** Standing, Seating, NO#, NA
- **Which Part Occluded? (B):** NA, Head-Shoulder, Lower-body, Upper-Body, Left Half, Right Half, Full Body
- **Hat (B):** No, Yes, NO#, NA
- **Glasses (B):** No, Yes, NO#, NA
- **Bag (B):** No, Yes, NO#, NA
- **Bag Type:** NA, Handbag, Backpack, Plastic Bag, Suitcase, Trolley, Shoulder-Bag, Other, NO#
- **Body Shape:** Thin, Normal, Fat, NO#, NA
- **Upper Body Clothing:** T-shirt, Shirt, Dress, Jacket, Suit, Coat, Sweater, Formal, No, Burqa, Saree, Other, NO#, NA
- **Upper Body Clothing Color:** Black, White, Blue, Red, Green, Brown, Grey, Orange, Pink, Purple, Yellow, NA, Other, NO#
- **Lower Body Clothing:** Shorts, Jeans, Pants, Skirt, Burqa, Saree, No, NA, Other, NO#
- **Lower Body Clothing Color:** Black, White, Blue, Red, Green, Brown, Grey, Orange, Pink, Purple, Yellow, NA, Other, NO#
- **Shoes (B):** Slippers, Flip-flops, Formal Shoes, Sneakers, Sandals, Running Shoes, Other, No, NA, NO#
- **Shoes Color:** Black, White, Blue, Grey, Other, NA, NO#
- **FaceMask (B):** No, Yes, NA, NO#
- **Weather:** Sunny, Cloudy, Rainy, High Humidity, Snow, Night Time, NO#
- **Height:** Normal, Short, Tall, NA, NO#
- **Accessories:** Nothing, Cellphone, Trolley, Umbrella, BiCycle, Kick scooter, Bike, Other, NA, NO#

### 2. Demographic Distribution

The demographic attributes in VAST-ReID dataset include Gender, Age, Hair, Body Shape, and Height. For each attribute, values such as NO# indicate Not Observable (cases where the attribute cannot be visually identified), No means the absence

of that attribute, NA denotes Not Applicable, and values marked as Uncertain correspond to annotators not being sure of the exact class. These definitions ensure consistent labeling and reduce ambiguity in downstream evaluation.

### 2.1. Gender Distribution

Figure 1 suggests that the dataset primarily consists of Male and Female individuals. In some frames, the person of interest (POI) is not present, leading to NA (Not Applicable) or NO# (Not Observable) values when gender cannot be inferred. A small proportion of cases are labeled as Uncertain, though these remain minimal, indicating that annotators could typically assign gender labels with high confidence.

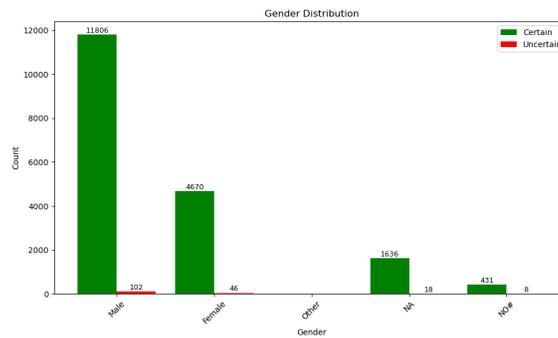


Figure 1. Gender distribution.

### 2.2. Body Shape Distribution

Figure 2 suggests that most individuals fall into the Normal body shape category, with smaller but notable proportions in the Thin and Fat categories. The NA (Not Applicable) and NO# (Not Observable) values represent cases where body shape could not be determined or person of interest is not in the frame. Uncertain cases are rare, indicating that body shape is generally annotated with high confidence.

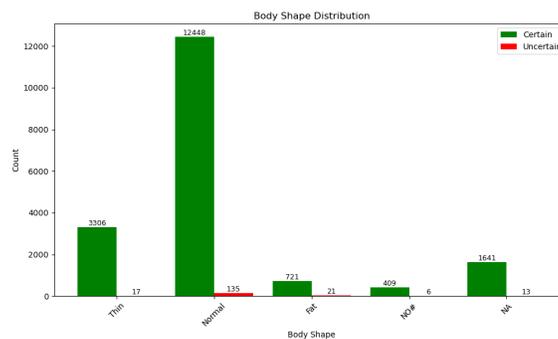


Figure 2. Body shape distribution of the persons.

### 2.3. Height Distribution

Figure 3 suggests that most individuals fall into the Normal height category, with smaller proportions categorized as Short or Tall. NA and NO# values occur in cases where height could not be estimated, such as severe occlusion or person of interest is not in the frame. The scarcity of Uncertain labels shows that height estimation is typically consistent across annotators.

### 2.4. Hair Distribution

Figure 4 suggests that most individuals have either Short or Long hair, with a smaller portion being Bald. NA and NO# cases reflect situations where the head is occluded, not visible, or not relevant for annotation. Uncertain cases remain rare, highlighting that hair type is generally straightforward to annotate.

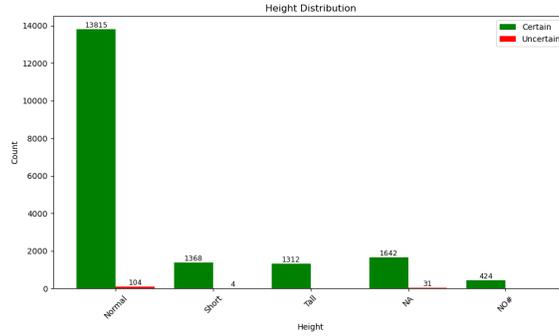


Figure 3. Height distribution of the persons.

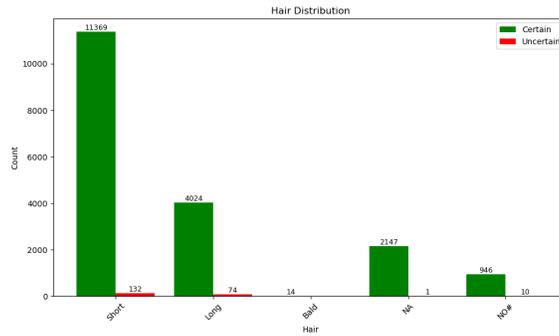


Figure 4. Hair distribution of the persons.

## 2.5. Age Distribution

Figure 5 suggests that the dataset is dominated by Adults, which is expected since the videos were primarily recorded within an institute campus environment. Smaller but meaningful representations exist for Teenagers, Middle-Aged, Children, and Older individuals. NA (Not Applicable) and NO# (Not Observable) values appear in cases where age could not be reasonably inferred from visual evidence or person of interest is not there. Uncertain annotations are very few, indicating reliable labeling of age categories.

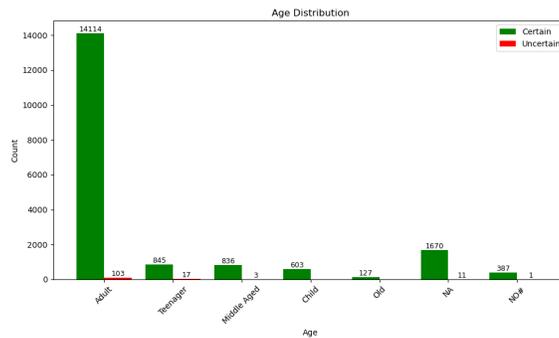


Figure 5. Age distribution of the persons.

## 3. Bounding-box Distribution

Figure 6 presents the distribution of bounding box statistics across the dataset. The bounding box widths are mostly concentrated below 250 pixels, while the heights are more spread out, with the majority under 500 pixels. The area distribution

shows that most bounding boxes are relatively small, reflecting that pedestrians typically occupy limited regions of the frames. The aspect ratio distribution is centered around 0.5–1.0, consistent with the upright human body structure, where height generally exceeds width. Overall, these statistics confirm that the dataset captures realistic person bounding boxes with natural variation in scale, pose, and viewing conditions.

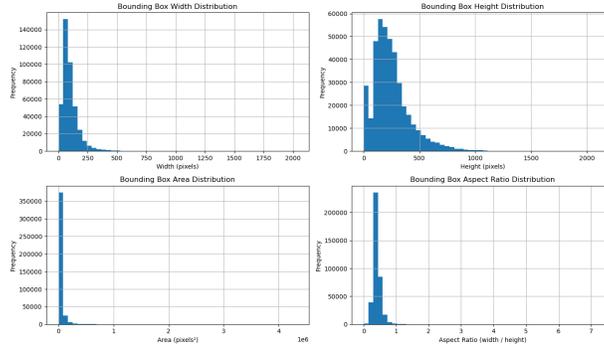


Figure 6. Bounding Box Distribution

## 4. Distribution of Views, Locations, and Subjects

Understanding the distribution of views(cameras), capture locations, and subject(person) counts is essential for analyzing dataset diversity and ensuring robust evaluation in person ReID tasks. Tables 1, 2, and 3 provide detailed breakdowns.

### 4.1. Subjects and Cameras

Table 1 reports the number of unique individuals captured by each camera. Camera C2 has observed the highest number of subjects (201), followed by C11 (121) and C5 (56). In contrast, cameras such as C3 (5) and C7 (6) have captured relatively few subjects. This variation highlights differences in field of view, placement, and activity density across cameras.

Camera	Num_Persons
C1	79
C2	201
C3	5
C4	42
C5	56
C6	32
C7	6
C8	9
C9	29
C10	46
C11	121
C12	40

Table 1. Number of unique persons captured by each camera in the dataset.

### 4.2. Subjects and Locations

Table 2 summarizes the number of unique individuals appearing in each location. Locations L22 (58), L23 (64), and L5 (71) are the most populated, whereas L7 (3) and L9 (7) have minimal coverage. Such diversity in subject density across locations supports evaluating models under different levels of crowding and scene complexity.

Table 2. Number of person and weather conditions observed across different locations.

Location	#Person	#Weather
L1	7	1
L2	11	1
L3	26	2
L4	11	1
L5	71	1
L6	13	1
L7	3	1
L8	9	1
L9	7	1
L11	32	2
L12	13	1
L13	15	1
L14	13	1
L15	11	1
L16	29	2
L17	13	2
L18	12	1
L19	20	1
L20	40	3
L21	8	2
L22	58	3
L23	64	3
L24	19	2

	C1	C11	C4	C8	C7	C5	C12	C10	C9	C3	C6	C2	Total
L1	2	0	0	0	0	0	0	0	0	0	0	6	8
L2	8	0	0	0	0	0	0	0	0	0	0	14	22
L3	12	32	0	4	0	0	0	32	0	0	0	3	83
L4	12	0	0	8	0	0	0	0	0	0	0	8	28
L5	16	14	114	0	0	140	0	0	0	0	0	71	355
L6	2	0	0	5	0	0	0	0	0	8	0	15	30
L7	6	0	0	0	0	0	0	0	0	0	0	0	6
L8	10	0	0	1	0	0	0	0	0	3	0	1	15
L9	0	0	0	0	11	0	0	0	0	0	0	2	13
L11	2	40	0	0	0	0	0	40	0	0	2	56	140
L12	13	0	0	0	0	0	0	0	0	0	12	8	33
L13	14	0	0	0	0	0	0	0	0	0	13	6	33
L14	22	0	0	0	0	0	0	0	0	0	4	26	52
L15	0	0	0	0	0	0	0	0	0	0	22	0	22
L16	5	6	0	0	0	0	0	0	77	0	6	69	163
L17	8	0	0	0	0	0	0	3	0	0	0	3	14
L18	10	0	0	0	0	0	0	4	0	0	0	0	14
L19	15	0	0	0	0	0	0	16	0	0	0	12	43
L20	0	0	0	0	0	0	163	0	0	0	0	0	163
L21	6	6	0	0	0	0	0	0	6	0	4	0	22
L22	0	59	0	0	0	0	0	0	0	0	0	0	59
L23	0	8	0	0	0	0	0	0	0	0	0	67	75
L24	0	22	0	0	0	0	0	0	0	0	0	26	48
Total	163	187	114	18	11	140	163	95	83	11	63	393	1441

Table 3. Distribution of camera view points across various geographical locations.

### 4.3. Camera-Location Distribution

Table 3 shows how subjects are distributed across different cameras (C1–C12) and physical locations (L1–L24). The coverage is heterogeneous: some locations, such as L5, L11, L16, and L20, contain a large number of samples, while others, such as L7 or L9, have relatively few. This imbalance reflects realistic surveillance conditions, where some areas have higher foot traffic or longer monitoring durations.