

# 1 Dataset

To perform the pose estimation in Section ??, we need some data on which to train, validate and test our models. The following section describes the datasets that will be used, as well as the preprocessing of these datasets.

Keypoints	ClimbAlong 25 keypoints	ClimbAlong 17 keypoints	BRACE	Penn Action
Head	No	No	No	Yes
Nose	Yes	Yes	Yes	No
Left ear	Yes	Yes	Yes	No
Right ear	Yes	Yes	Yes	No
Left eye	No	Yes	Yes	No
Right eye	No	Yes	Yes	No
Left shoulder	Yes	Yes	Yes	Yes
Right shoulder	Yes	Yes	Yes	Yes
Left elbow	Yes	Yes	Yes	Yes
Right elbow	Yes	Yes	Yes	Yes
Left wrist	Yes	Yes	Yes	Yes
Right wrist	Yes	Yes	Yes	Yes
Left pinky	Yes	No	No	No
Right pinky	Yes	No	No	No
Left index	Yes	No	No	No
Right index	Yes	No	No	No
Left thumb	Yes	No	No	No
Right thumb	Yes	No	No	No
Left hip	Yes	Yes	Yes	Yes
Right hip	Yes	Yes	Yes	Yes
Left knee	Yes	Yes	Yes	Yes
Right knee	Yes	Yes	Yes	Yes
Left ankle	Yes	Yes	Yes	Yes
Right ankle	Yes	Yes	Yes	Yes
Left heel	Yes	No	No	No
Right heel	Yes	No	No	No
Left toes	Yes	No	No	No
Right toes	Yes	No	No	No

Table 1: Overview of the annotated keypoints of the four used datasets

## 1.1 The ClimbAlong Dataset

As the aim of our models is to perform well on climbers, we will be using some annotated data of climbers. For this, ClimbAlong ApS has developed a dataset that we will be using. The dataset consists of videos various climbers on bouldering walls, where each video contains just a single climber. Figure 1 illustrates two windows of consecutive frames of a single video from the ClimbAlong dataset. As shown in the figure, the videos in the dataset both contains static positions, where the climber holds a position for a while, as well as quick movements.

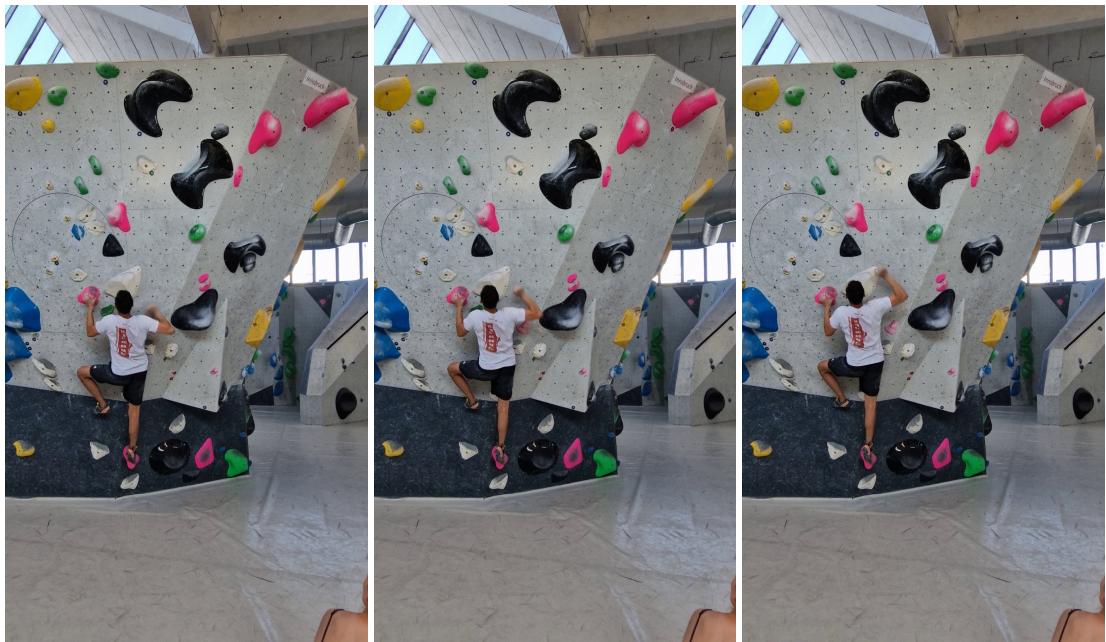
The ClimbAlong-dataset is split into two subsets. The first subset consists of 31 fully annotated videos, where each annotation consists of 25 keypoints. The second subset consists of



(a) Frame 262

(b) Frame 263

(c) Frame 264



(d) Frame 288

(e) Frame 289

(f) Frame 290

Figure 1: Example of two windows of three consecutive frames of a video from the ClimbAlong dataset

baseball_pitch	baseball_swing	bench_press
bowling	clean_and_jerk	golf_swing
jumping_jacks	jump_rope	pull_ups
push_ups	sit_ups	squats
strumming_guitar	tennis_forehand	tennis.Serve

Table 2: The original 15 action-types in the Penn Action dataset.

110 fully annotated videos, where each annotation consists of 17 keypoints. The two subsets have 13 overlapping videos.

Each of the videos is in the RGB-format, is filmed in portrait mode with a resolution of  $1080 \times 1920$  and 30 frames per second. Table 1 gives an overview of which keypoints are annotated in the two subsets.

## 1.2 The BRACE Dataset

The second dataset we will be using is the *BRACE* dataset [1]. We chose to use this dataset, as breakdancers tend to swap between static and acrobatic poses, similarly to the ones in the *ClimbAlong* dataset, making the poses relevant for our experiments in Section ??.

This dataset consists of 1,352 video sequences and a total of 334,538 frames with keypoints annotations of breakdancers. The frames of the video sequences are in RGB and have a resolution of  $1920 \times 1080$  [1].

The frames of the video sequences have been annotated by initially using state-of-the-art human pose estimators to extract automatic poses. This was then followed by manually annotating bad keypoints, corresponding to difficult poses, as well as pose outliers. Finally, the automatic and manual annotations were merged, interpolating the keypoint sequence with Bézier curves. The keypoints is a list of 17-elements, following the COCO-format [1].

HVIS EKSEMPEL OG SKRIV NOGET OM HVILKE KEYPOITNS DER ER ANNOTERET

## 1.3 The Penn Action Dataset

The final dataset we will be using is the *Penn Action* dataset [2]. This dataset consists of 2326 video sequences of 15 different action-types. Table 2 lists these 15 action-types [2].

Each sequence has been manually annotated with human joint annotation, consisting of 13 joints as well as a corresponding binary visibility-flag for each joint. The frames of each sequence are in the RGB-format and has a resolution within the size of  $640 \times 480$  [2].

Unlike the *BRACE* dataset, most of the poses in the *Penn Action* dataset are not very acrobatic and thus are not very relevant for the poses of climbers. For that reason, we have decided to focus on the action-types that may contain more acrobatic poses. Thus, we only keep the sequences that have *baseball\_pitch*, *bench\_press* or *sit\_ups* as their corresponding action-type [2].

HVIS EKSEMPEL OG SKRIV NOGET OM HVILKE KEYPOITNS DER ER ANNOTERET

## **1.4 Preprocessing of the Data**

## **1.5 The ClimbAlong Dataset**