Supplemental Material

6. Learning effect

As explained in the study design 4.2.1, the order of scenarios is randomized for each annotator to prevent potential familiarization bias in the statistics. This learning effect is measured during the study and visualized in Figure 10. Annotators who are assigned a certain scenario as their first task require 36.2% more time than the mean for this scenario. In contrast, annotators who are assigned a scenario as their fourth task are 20.7% faster than the scenario mean.

Therefore, large annotation projects should not underestimate the importance of interactive learning material for their annotator. Our experiences show that specifications alone are insufficient to provide optimal efficiency of the annotators. Future works should focus on pedagogical and interactive material to efficiently prepare human annotators to their task.

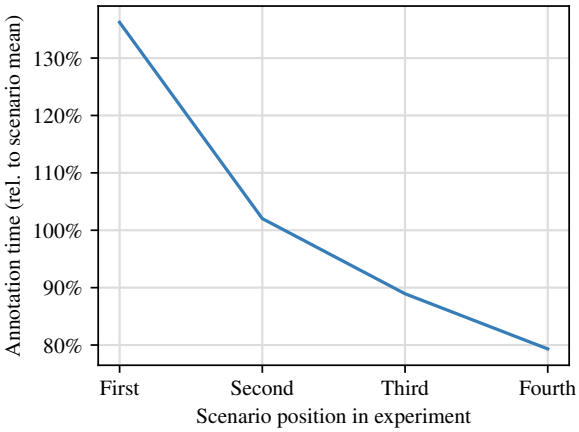


Figure 10: Relative annotation time for each scenario, depending on its random position during the experiment, compared to the mean time across all positions. 10 samples per position.

7. Annotation Quality

Figure 11 shows another potential problem with manual annotation. Since keypoints are placed separately on every frame, slight inaccuracies in the placement lead to jittery tracking across frames. In contrast, inter- or extrapolated keypoints naturally follow a smooth path. Even so, this does not necessarily mean that inter-/extrapolation is more accurate than manual annotation and the necessity of smooth tracking depends on the particular case of use.

Figure 12 demonstrates situations in which tool assistance works well, as well as frames in which tool assistance provides no benefit.

Annotation Method	AP							
	Head	Shou	Elb	Wri	Hip	Knee	Ankl	Total
Mean	55.3	82.0	76.9	65.9	75.0	80.1	74.9	71.7
Most accurate (Manual w. occ.)	63.3	90.0	86.9	77.7	90.2	85.1	87.5	81.7
Least accurate (Manual w/o. occ.)	49.6	63.4	62.0	57.2	57.7	69.7	57.6	59.0
Mean Extrapolation	52.9	82.4	75.0	62.6	69.1	80.9	71.0	69.4
Mean HPE + Interpolation	55.4	81.6	76.8	63.8	78.9	80.8	76.3	72.2
Mean Manual	60.8	84.8	82.8	74.6	76.4	79.8	82.3	76.3
Mean Manual w/o occlusion	51.2	77.0	69.5	57.1	70.6	78.1	68.8	66.4
HPE + copy	54.3	82.0	76.5	67.0	77.6	80.0	73.1	71.7

Table 3: Per-joint AP. Due to helmets and noise in the GT (see Figure 6) the AP for the three head keypoints remains quite low.

In Table 3 we further investigate the annotation quality from Section 4.1 against the ground truth provided by the PoseTrack18 dataset. As shown, best average results are achieved through manual annotation. Large quality differences are shown for *Head*, *Elbow* and *Ankle*.

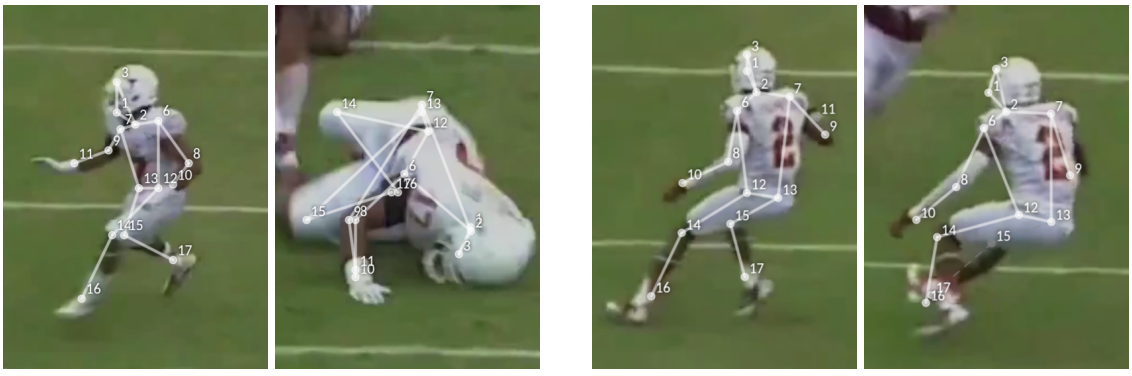
8. Annotator Feedbacks

Observing current workflows provides the opportunity to gain insight in which functionalities annotators utilize or desire. For this purpose, annotators with experience in annotating keypoints using Sloth [32] were interviewed before the implementation of this paper. In summary, annotators were irritated not being able to undo the placement of a keypoint or being able to cut, copy, or paste individual annotations. When annotating complex sequences with blurry images, occlusions, and fast movements, annotators switch between frames often and wish to be able to temporarily hide annotations or keypoint labels.

During the study, annotators provide feedback and ideas on how to improve annotating processes. Annotators find that manual annotation should be customizable independently from the dataset topology. For instance, several annotators would prefer to annotate the left and right side of a person consecutively. Minor improvements might also be made to keypoint labels and colors for a enhanced pose understanding. The annotation suggestions of the pose estimation processing tool can be optimized. Potential means are the application of continuous or active learning. Likewise, interpolation suggestions can be improved. Related work has shown promising results with visual interpolation [22], which could be a server-side alternative for difficult scenarios in which geometric interpolation does not produce satisfactory results.

(a) Manual annotation with correct occlusion (b) Accepted pose suggestion with missed occlusion (c) Manual annotation on frame X (d) Manual annotation on frame X+1

Figure 11: Two prominent problems in annotation quality: In (a), occluded keypoints are correctly annotated. The annotator generates suggestions for the keypoint positions in (b), but occlusions are not amended. The consecutive frames (c) and (d) are manually annotated and show slightly different positions of keypoints on the body, which results in jittery tracking. This is especially noticeable for keypoints 12 and 13 in this example. (Best seen in color)



(a) Accurate pose estimation (b) Inaccurate pose estimation (c) Accurate Interpolation (d) Inaccurate Interpolation

Figure 12: Situations in which tool assistance is applicable and in which the tool does not provide a useful suggestion. The estimated pose in (a) is reasonably accurate, whereas the pose estimator fails to estimate the pose in (b). In (c), the interpolated annotation is relatively close to the true pose; in (d) the player movement is difficult to predict.