Bootstrapping Human Optical Flow and Pose

Supplementary Material

We provide additional results excluded from the main paper due to spatial constraints.

A Ablation study

To motivate the design choices of our method we conduct an ablation study.

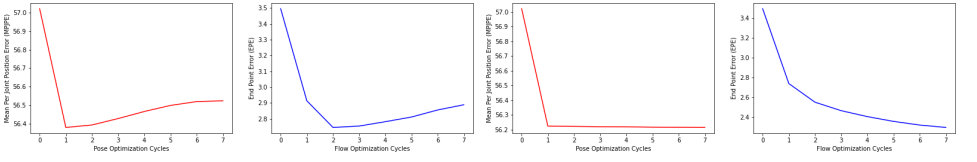
Loss terms in pose optimization. We first examine how the pose estimation performance of our method varies as we add new loss terms in Eq. (2). We report these results in Table 3, using the Human3.6M dataset. As shown, each loss term contributes to enhanced MPJPE, demonstrating that all loss components are important. We emphasize once more here that, while there are four loss terms in total, we use the *same* hyperparameter setting for all our experiments.

Method	MPJPE ↓
Initial pose estimates (METRO)	54.07
$\mathcal{L}_{3D}$	54.07
$\mathcal{L}_{3D} + \mathcal{L}_{2D}$	53.93
$\mathcal{L}_{3D} + \mathcal{L}_{2D} + \mathcal{L}_{temp}$ (without bone consistency)	53.45
$\mathcal{L}_{3D} + \mathcal{L}_{2D} + \mathcal{L}_{temp}$	53.29
$\mathcal{L}_{3D} + \mathcal{L}_{2D} + \mathcal{L}_{temp} + \mathcal{L}_{opt}$	**53.15**

Table 3: **Ablation study on pose-related loss terms –** Ablation study on the Human3.6M dataset showing the effects of adding different loss terms to our pose refinement pipeline. All loss terms contribute to the enhancement of human pose accuracy.

Number of optimization cycles With a representative video from the Human3.6M dataset, we report how pose estimation accuracy and human optical flow accuracy change as we perform more optimization cycles. As shown in Figure 6 (a–b), the best pose is achieved after the first pose optimization cycle, whereas the best flow is achieved after the second. This demonstrates that a single pose optimization cycle is enough to take optical flow into account when estimating the poses, which can then correct optical flow with the enhanced poses. The increase in error afterward indicates that there is a potential drift after more optimization cycles. For example, when too many optimization cycles are performed, the RAFT network can overfit to the rough flow estimates shown in Figure 3 (c), which *will* contain errors, leading to degradation.

As shown in Figure 6 (c–d), where these estimation errors are not present, this drifting does not happen. We note that this is a limitation of *any* pipeline that self-boot-straps and is not unique to our method. In addition, even when degradation happens, the performance still is much better than the initial MPJPE and EPE values. Based on these results, we perform one pose optimization cycle and two flow cycles for all our experiments.



(a) MPJPE vs #cycles (b) EPE vs #cycles (c) MPJPE vs #cycles /w GT (d) EPE vs #cycles /w GT

Figure 6: **Ablation study on the number of cycles –** MPJPE and EPE change with respect to the number of optimization cycles, on a video sequence of the Human3.6M dataset. (a–b) when the optimization is purely based on our method using estimated pose and flow, and (c–d) when we replace the optimization cycle with ground-truth measurements rather than estimated pose and flow. For the case when the estimated pose and flow are used, the best pose is already achieved at the first optimization cycle, whereas the flow at the second. The results then deteriorate, showing 'drifting'. When using the ground truth, this does not happen, further suggesting drifting. Nonetheless, optimized results with our method improve over initial estimates even with drifting. While these measurements are for a single video, the same trend can be observed in general.