The Invisible EgoHand: 3D Hand Forecasting through EgoBody Pose Estimation

Supplementary Material

Table 7. **Balancing Hyperparameters**. We conduct an ablation study on the loss weights for the reprojection loss and visibility loss, and report the hand trajectory and pose forecasting accuracy in terms of ADE and MPJPE, respectively.

(a) λ_{reproj}

λ_{reproj}	ADE	MPJPE
0.5	0.262	0.125
0.05	**0.261**	**0.115**
0.01	0.262	0.123

(b) λ_{vis}

λ_{vis}	ADE	MPJPE
1.0	0.275	0.121
0.1	**0.261**	**0.115**
0.01	0.273	0.124

A. Additional Implementation Details

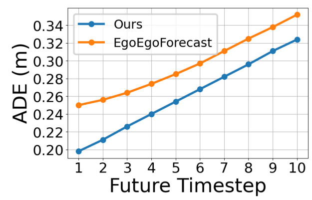
We use the ViT-Small as the visual encoder, so the dimension of image features d_{img} is 384. As for the dimension of tokens for each timestep d_z, we set it to 512, and the number of layers and the number of heads of the Transformer are set to 4 and 8.

B. Additional Evaluation Results

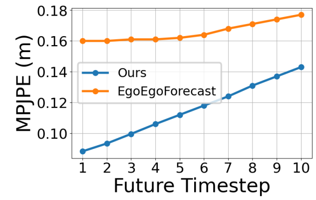
Balancing Hyperparameters. We conduct an ablation study to analyze the impact of balancing hyperparameters for the reprojection loss $\mathcal{L}_{\text{reproj}}$ and the visibility loss $\mathcal{L}_{\text{vis}}$ on the hand forecasting accuracy. We systematically vary the values of each balancing hyperparameter across predefined ranges: λ_{reproj} from 0.01 to 0.5, and λ_{vis} from 0.01 to 1.0. We vary one hyperparameter at a time while keeping the others fixed at their default values: 0.05 and 0.1 for $\mathcal{L}_{\text{reproj}}$ and $\mathcal{L}_{\text{vis}}$, respectively.

We report the hand trajectory and pose forecasting accuracy of the proposed model with varied hyperparameters in terms of ADE and MPJPE in Tab. 7. This ablation analysis reveals the importance of balancing these hyperparameters for optimal hand forecasting accuracy. Specifically, a lower value of λ_{reproj} reduces the model's reliance on consistency between 2D input and 3D output, leading to poor spatial alignment with visible 2D hand input. Conversely, a high value of λ_{reproj} overemphasizes reprojection accuracy, causing the model to neglect the correct 3D depth (i.e. distance from camera) estimation and resulting in suboptimal predictions. Regarding the weight for visibility loss, a higher value degrades the forecasting performance as it is not directly related to hand forecasting, while a lower value reduces the model's in- or out-of-view awareness, leading to a performance drop.

Per-timestep Hand Forecasting Accuracy. We report the hand trajectory and pose forecasting accuracy for each future timestep in Fig. 5. Overall, EgoH4 outperforms the EgoEgoForecast baseline on every future timestep for both hand trajectory and pose forecasting tasks. Specifically, the improvements over the baseline are most pronounced at earlier future timesteps in the hand pose forecasting, as EgoH4 achieves more accurate hand pose estimation by leveraging visible 2D hand locations. In the hand trajectory forecasting task, our model consistently outperforms the baseline by effectively accounting for in-view or out-of-view during the observation period.



(a) Per-timestep hand trajectory forecasting accuracy.



(b) Per-timestep hand pose forecasting accuracy.

Figure 5. **Per-timestep Hand Forecasting Accuracy**. We report the hand trajectory forecasting accuracy in ADE and hand pose forecasting accuracy in MPJPE for every future timestep. Lines in blue and orange represent the performance of our model and the EgoEgoForecast baseline, respectively.