# Supplementary Materials

## 1. Notation Table

The symbols used in this paper are summarized in Table 1.

| Symbol | Description |
|---|---|
| $I$ | The input image |
| $f_{enc}$ | The extracted feature from CNN and transformer encoder |
| $f_{2D}$ | The feature for generating 2D hypotheses |
| $f_{3D}$ | The feature for generating 3D hypotheses |
| $f_{hmr}$ | The feature for generating SMPL parameters |
| $z_{2D}$ | The sampled latent feature in 2D HGM |
| $z_{3D}$ | The sampled latent feature in 3D HGM |
| $\mu_{2D}$ | Predicted mean for the 2D pose hypotheses distribution |
| $\sigma_{2D}$ | Predicted standard deviation for the 2D pose hypotheses distribution |
| $\mu_{3D}$ | Predicted mean for the 3D pose hypotheses distribution |
| $\sigma_{3D}$ | Predicted standard deviation for the 3D pose hypotheses distribution |
| $H_{2D\_heat}$ | 2D pose hypotheses in the form of 2D heatmap |
| $H_{3D}$ | 3D pose hypotheses in the form of 3D coordinates |
| $H_{3D->2D}$ | Projected 2D coordinates from $H_{3D}$ |
| $H'_{2D\_heat}$ | 2D heatmap after making Gaussian blobs on $H_{3D->2D}$ |
| $f_{2D\_Sampled}$ | The sampled feature by $H_{2D\_heat}$ |
| $f_{3D\_Sdampled}$ | The sampled feature by $H'_{2D\_heat}$ |
| $\theta$ | Pose parameters of SMPL [4] |
| $\beta$ | Shape parameters of SMPL [4] |
| $M$ | The final mesh |
| $J_{3D}$ | The body joints regressed from $M$ |
| $K$ | Number of hypotheses |
| $N_J$ | Number of body joints |

Table 1. A list of the symbols used in the main manuscript.

## 2. Hyperparameters

The hyperparameters adopted for training our framework are summarized in Table 2.

Table 2. A summary of the hyperparameters used in our framework.

| Hyperparameter Settings | |
|---|---|
| Learning Rate | $10^{-4}$ |
| Weight Decay Factor | $10^{-4}$ |
| Betas | $(0.9, 0.999)$ |
| Batch Size | 64 |
| Image Crop Size | $224 \times 224$ |
| Epochs | 60 |
| $N_J$ | 24 |
| $k$ | 81 |
| $\lambda_{hmr}$ | 60 |
| $\lambda_{pose}$ | 5 |
| $\lambda_{smpl}$ | 1 |
| $\lambda_{3D}$ | 300 |
| $\lambda_{2D}$ | 200 |
| $\lambda_{reg}$ | 10 |
| **Hardware Settings** | |
| CPU | AMD Ryzen™ 9 5950X |
| Main Memory | 128GB |
| GPU | NVIDIA RTX 3090 |
| GPU Memory | 24GB |

## 3. Visualization of Hypotheses

In this section, we present additional visualizations of 2D and 3D pose hypotheses, as depicted in Fig. 1. For the 2D poses, we visualize the keypoints for the Right Ankle, Right Knee, Right Hip, Left Hip, Left Knee, and Left Ankle. For the 3D poses, we focus on the left and right wrists and the left and right ankles.

## 4. Visualization of Additional Qualitative Results

Fig. 2 presents the additional qualitative results of both FastMETRO [1] and our method on the Human3.6M and 3DPW datasets.

## 5. A Guideline for Reproduction

Our implementation follows the previous Transformer-based methods [1–3]. For detailed information and code, please refer to our anonymous repository: https://anonymous.4open.science/r/Progressive-Hypothesis-Transformer-for-3D-Human-Mesh-Recovery.