DenseRaC: Joint 3D Pose and Shape Estimation by Dense Render-and-Compare

Yunlu Xu1,2, Song-Chun Zhu1, Tony Tung1 1Facebook Reality Labs, Sausalito 2University of California, Los Angeles (UCLA) merayxu@gmail.com, sczhu@stat.ucla.edu, tony.tung@fb.com

**Motivation**

In this paper, we propose an end-to-end framework for jointly estimating 3D human pose and shape from a monocular RGB image.

- **A Two-Stage Framework**
  - Existing frameworks: direct regression, little information about body details

  - **Part Mask Loss** \( L_{mask} \): GT part masks vs. part masks from rendered output
  - **Adversarial Loss** \( L_{advers} \): real/fake 3D body configurations

**Experiments**

- **MOCA Dataset**: large-scale synthetic dataset (2M+ images) from covering various camera views, human actions and body shapes, with fully paired ground truth.
  - **Pose**: from web-crawled Mocap sequences and 3D animations (260K frames)
  - **Body Shape**: 3D human scans (CAESAR dataset, 2,781 samples)
  - Random camera view, cropping and scaling.

**Reference**


**A Render-and-Compare Framework**

- **Minimizing differences between input and rendered output**
  - **Regression Loss** \( L_{reg} \): GT parameters vs. regressed parameters
  - **Reconstruction Loss** \( L_{rec} \): GT mesh/skeleton vs. reconstructed mesh/skeleton
  - **Reprojection Loss** \( L_{reproj} \): GT 2d landmarks (joints, dense landmarks from IUV) vs landmarks from rendered output

- **Flexibility** (beyond reconstruction)
  - **Part Mask Loss** \( L_{mask} \): GT part masks vs part masks from rendered output
  - **Adversarial Loss** \( L_{advers} \): real/fake 3D body configurations

**3D Pose Estimation**

- **MOCA Dataset**: ground truth mesh
  - **BodyNet**: ground truth mesh
  - **3D Human Body**: synthetic data

**Semantic Segmentation**

- **UP3D**: 3D Human Body
  - **HMR**: Human3.6M, MPI
  - **HRMS**: HRMS, MPI

**Applications**

- **Virtual Dressing**
  - Under-clothes body reconstruction + clothes simulation
  - Volumetric reconstruction + under-clothes body mesh fitting
  - Fidelity (visually plausible vs. details)
  - Robustness (crowd, occlusions)
  - Flexibility (beyond reconstruction)