Coarse-to-fine Animal Pose and Shape Estimation: Supplementary Material

Chen Li   **Gim Hee Lee**
Department of Computer Science
National University of Singapore
{lic, gimhee.lee}@comp.nus.edu.sg

We conduct further ablation studies for our approach in this supplementary material, including comparison with test-time optimization and sensitivity analysis of the refinement stage. Additional qualitative results are also provided.

Comparison with test-time optimization.   We compare our coarse-to-fine approach with the test-time optimization approach. As has been done in our coarse-to-fine pipeline, we also use the output from our coarse estimation stage as an initialization. Instead of apply the mesh refinement GCN, we further optimize the SMAL parameters based on the keypoints and silhouettes for 10, 50, 100, 200 iterations, respectively. We show the average PCK and IOU in the Table 1, as well as the inference time for the test set. We can see that the performance of the test-time optimization gets better with more optimization iterations. However, the inference time also increases linearly with the number of optimization iterations. In comparison, our regression based refinement achieves better performance with faster inference time. Moreover, the test-time optimization requires 2D keypoints and silhouettes, which are not always available in practice.

Table 1: Comparison with test-time optimization.

Num Iters	time (s)	Avg PCK	IOU
10	570	78.3	74.2
50	2561	79.2	76.1
100	5040	79.9	77.4
200	10018	81.7	79.0
Ours	64.5	83.4	81.6

Ablation study on the sensitivity of the second stage to the first stage results.   The refinement stage of our approach relies on the output of the coarse estimation stage as an initial point. We test the sensitivity of our model to the first stage results by adding Gaussian noise to the SMAL and camera parameters estimated from the coarse estimation stage, respectively. We compute the standard deviation σ of the estimated SMAL and camera parameters over the whole dataset, and set the standard deviation of the Gaussian noise to 10%, 20%, 30% and 50% of σ. We show the results under SMAL noise and camera noise in Table 2a and 2b, respectively. We can see that our model is robust to the noise adding to the SAML parameters, and relatively sensitive to the noise adding to the camera parameter. We expect the sensitivity to the camera parameter noise because we only refine the mesh vertices in the second stage.

Additional qualitative results.   We show more qualitative results in Figure 1. We can see that we can get reasonable results from the SMAL-based estimations in the first stage. The coarse estimations are further improved in the refinement stage, where both keypoints and silhouettes from the rendered 3D meshes align better with ground truth annotations.

Table 2: Adding Gaussian noise to the estimated SMAL parameters (a) and camera parameter (b).

<table>
<tr><td colspan="3" align="center">(a)</td><td colspan="3" align="center">(b)</td></tr>
<tr><td>SMAL Noise</td><td>Avg PCK</td><td>IOU</td><td>CAM Noise</td><td>Avg PCK</td><td>IOU</td></tr>
<tr><td>0.1</td><td>81.3</td><td>78.3</td><td>0.1</td><td>82.2</td><td>78.3</td></tr>
<tr><td>0.2</td><td>79.7</td><td>76.7</td><td>0.2</td><td>78.6</td><td>75.2</td></tr>
<tr><td>0.3</td><td>79.0</td><td>75.9</td><td>0.3</td><td>75.4</td><td>72.9</td></tr>
<tr><td>0.5</td><td>78.1</td><td>75.4</td><td>0.5</td><td>69.7</td><td>68.3</td></tr>
</table>

Figure 1: Some results of our coarse-to-fine approach. The first column is the input RGB images. The second to fourth columns are the estimated meshes in the camera view, the projected 2D keypoints and silhouettes from the coarse estimations. The fifth column to seventh column are the estimated meshes in the camera view, the projected 2D keypoints and silhouettes from the refined estimations. The last column is the 3D mesh in another view.

Checklist

1. For all authors...
   (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? [Yes]
   (b) Did you describe the limitations of your work? [Yes] See failure cases in Section 4.3 and discussion in limitations.
   (c) Did you discuss any potential negative societal impacts of your work? [Yes] See discussion in limitations.
   (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? [Yes]

2. If you are including theoretical results...
   (a) Did you state the full set of assumptions of all theoretical results? [N/A]
   (b) Did you include complete proofs of all theoretical results? [N/A]

3. If you ran experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [Yes] See Section 4 and our code repository.
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [Yes] See Section 4 and our code repository.
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [No] We have run our network several times to make sure that the results are reproducible.
   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [Yes] See training details in Section 4.

4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...
   (a) If your work uses existing assets, did you cite the creators? [Yes] See Section 4.1.
   (b) Did you mention the license of the assets? [No] License of some datasets is not available
   (c) Did you include any new assets either in the supplemental material or as a URL? [Yes] We include the code as a URL.
   (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [No] The datasets we are using mainly contain animal images
   (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [No] The datasets we are using mainly contain animal images

5. If you used crowdsourcing or conducted research with human subjects...
   (a) Did you include the full text of instructions given to participants and screenshots, if applicable? [N/A]
   (b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [N/A]
   (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [N/A]