Supplementary Material for Multi-Hypothesis 3D Hand Mesh Recovering from a Single Blurry Image

Anonymous ICME submission

I. Loss of Selection Model

The selection model is expected to select the relatively "good" hypotheses from a set of hypotheses.[1] Let $\mathbf{x}$ denotes the hypothesis set $\{x^1, \ldots, x^K\} = \{\mathbf{H}^1, \ldots, \mathbf{H}^K\}$ and $\mathbf{y} = \{y^1, \ldots, y^K\}$ denotes the corresponding labels. $y^k = 1$ if $\mathbf{H}^k$ is relatively "good", otherwise $y^k = 0$.

It can be regarded as a classification problem that classifies the hypotheses into two classes: whether the hypothesis is "good" or not. Thus, a point-wise loss [1] can be applied to select the hypotheses. It can be formulated as a binary cross-entropy loss with an energy-based model (EBM):

$$\mathcal{L}_{pointwise}(x^k) = y^k \log(\hat{s}^k) + (1 - y^k) \log(1 - \hat{s}^k) \tag{1}$$

where $\hat{s}^k = S_\phi(\mathbf{H}^k|\mathbf{I})$ is the estimated score of $\mathbf{H}^k$ given image $\mathbf{I}$. Then the probability for $\mathbf{H}^k$ to be "good" is

$$p(y^k = 1|x^k) = \frac{\exp(\hat{s}^k)}{Z(\phi)} \tag{2}$$

where $Z(\phi)$ is the unknown normalization constant.

However, the above probability does not only depend on the considered hypothesis $\mathbf{H}^k$ and the image. It also conditions on other hypotheses $\mathbf{H}^{-k}$, which are not involved in Eq.1. In other words, Eq.1 assumes that the groundtruth score is absolute and only conditions on the image. In the multi-hypotheses setting, such an assumption tends to result in suboptimal estimated scores as it ignores the order information among different hypotheses of the same image.

A widely used training loss in previous works [2] is the pair-wise loss [1], which can be formulated as a Bradley-Terry model [3]. It assumes that the estimated score $\hat{s}^k = S_\theta(x^k|\mathbf{I})$ satisfies

$$P(x^i \succ x^j|\mathbf{I}) = \sigma(\hat{s}^i - \hat{s}^j) \tag{3}$$

where LHS is the probability that x^i is better than x^j given image $\mathbf{I}$. The training loss maximizes the probabilities of all hypothesis pairs, that is

$$\mathcal{L}_{pairwise}(x^i, x^j) = -\frac{1}{1 + \exp(-(\hat{s}^i - \hat{s}^j))} \tag{4}$$

where x^i and x^j satisfies $x^i \succ x^j$.

Pair-wise loss leverages the order information and trains a relative score only used for comparison. However, pair-wise loss ranks all hypothesis pairs equally, while previous works [2] only consider the average result of the top-n hypotheses instead of utilising the *order* of them. The groundtruth order for pairs within the top-n ones and within the non-top-n ones are extra constraints for our targets.

To better fit our target, we combine the pointwise loss and listwise loss.

We aim at maximizing the joint probability of $P(\mathbf{x}, \mathbf{y}|\mathbf{I})$ as $\mathbf{x}$ and $\mathbf{y}$ are coupled. We use two heads to explicitly predict the logit value of $p(x, y = 0)$ and $p(x, y = 1)$, denoted as ν and μ, respectively. The joint probability is optimized from two sides.

One is $p(\mathbf{y}|\mathbf{x}, \mathbf{I})$, the conditional probability of $\mathbf{y}$ given a hypothesis $\mathbf{H}^k$. Following the definition of probability, we have

$$p(y^k, x^k|\mathbf{I}) = p(y^k|x^k, \mathbf{I}) p(x^k|\mathbf{I}) \tag{5}$$

Notice that our generation model generates K hypotheses with the same probabilities. Thus, $p(x^k|\mathbf{I})$ is a constant which can be dropped in gradient computation. The **classifying loss**can be formulated as

$$\begin{aligned}
\mathcal{L}_{class} &= -p(y^k|x^k, \mathbf{I}) && (6)\\
&= -\frac{p(y = y^k, x^k|\mathbf{I})}{p(y = 0, x^k|\mathbf{I}) + p(y = 1, x^k|\mathbf{I})} && (7)\\
&= -\frac{(y^k) \exp \mu^k + (1 - y^k) \exp \nu^k}{\exp \mu^k + \exp \nu^k} && (8)
\end{aligned}$$

The other side is $p(\mathbf{x}|\mathbf{y}, \mathbf{I})$, the conditional probability of $\mathbf{x}$. We have

$$p(x^k|y^k, \mathbf{I}) = p(y^k|x^k, \mathbf{I}) \frac{p(x^k|\mathbf{I})}{p(y^k|\mathbf{I})} \tag{9}$$

.

Notice that $p(y^k|\mathbf{I})$ does not equal to n/K or $(n - K)/K$ as "top-n in K hypotheses" is a heuristic approach to the relatively "good" hypotheses. Instead, the probability of making a real "good" hypothesis can be empirically estimated by

$$\begin{aligned}
p(y^k|\mathbf{I}) &= \sum_{x^i \in \mathbf{x}} p(y^i = y^k|x^i, \mathbf{I}) p(x^i|\mathbf{I}) && (10)\\
&\propto \sum_{x^i \in \mathbf{x}} p(y^i = y^k|x^i, \mathbf{I}) && (11)\\
&\propto \sum_{x^i \in \mathbf{x}} \left[(y^k) \exp \mu^i + (1 - y^k) \exp \nu^i\right] && (12)
\end{aligned}$$

[1] Heuristically, we consider hypotheses with the top-n scores to be relatively "good".

Then we have the **ranking loss**

$$\begin{aligned}\mathcal{L}_{rank} &= -p(x^k|y^k, \mathbf{I}) && (13)\\ &= -\frac{(y^k)\exp\mu^k + (1-y^k)\exp\nu^k}{\sum_{x^i\in\mathbf{x}}\left[(y^k)\exp\mu^i + (1-y^k)\exp\nu^i\right]} && (14)\end{aligned}$$

By minimizing Eq.14, the hypotheses with $y=1$ are ranked in front of those with $y=0$ and vice versa.

The **combined classifying and ranking** (CCR) loss is obtained by

$$\mathcal{L}_{ccr} = \alpha\mathcal{L}_{class} + (1-\alpha)\mathcal{L}_{rank} \quad (15)$$

where $\alpha \in [0,1]$ is the hyper-parameter for balance.

In [4], there is a discussion of the loss function with the same formulation. But its discussion is conducted from the perspective of combining point-wise ranking loss and list-wise ranking loss [1].

II. Additional Qualitative Results

Diversity of Hypotheses. In Figure 3, we visualize the hypotheses under different extend of blurriness. It shows that the hypotheses become more diverse as the image becomes blurry. Figure 1 shows more qualitative results of the hypotheses.

Failure Cases. Figure 2 show an example of failure cases. Under extreme motion conditions, the multi-hypothesis method is unable to cover the correct pose but still proposes plausible results.

References

[1] Chris Burges, Tal Shaked, Erin Renshaw, Ari Lazier, Matt Deeds, Nicole Hamilton, and Greg Hullender, "Learning to rank using gradient descent," in *Proceedings of the 22nd International Conference on Machine Learning*, New York, NY, USA, 2005, ICML '05, p. 89–96, Association for Computing Machinery.

[2] Yuan Xu, Xiaoxuan Ma, Jiajun Su, Wentao Zhu, Yu Qiao, and Yizhou Wang, "Scorehypo: Probabilistic human mesh estimation with hypothesis scoring," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2024, pp. 979–989.

[3] Ralph Allan Bradley and Milton E Terry, "Rank analysis of incomplete block designs: I. the method of paired comparisons," *Biometrika*, vol. 39, no. 3/4, pp. 324–345, 1952.

[4] Xiang-Rong Sheng, Jingyue Gao, Yueyao Cheng, Siran Yang, Shuguang Han, Hongbo Deng, Yuning Jiang, Jian Xu, and Bo Zheng, "Joint optimization of ranking and calibration with contextualized hybrid model," in *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, New York, NY, USA, 2023, KDD '23, p. 4813–4822, Association for Computing Machinery.

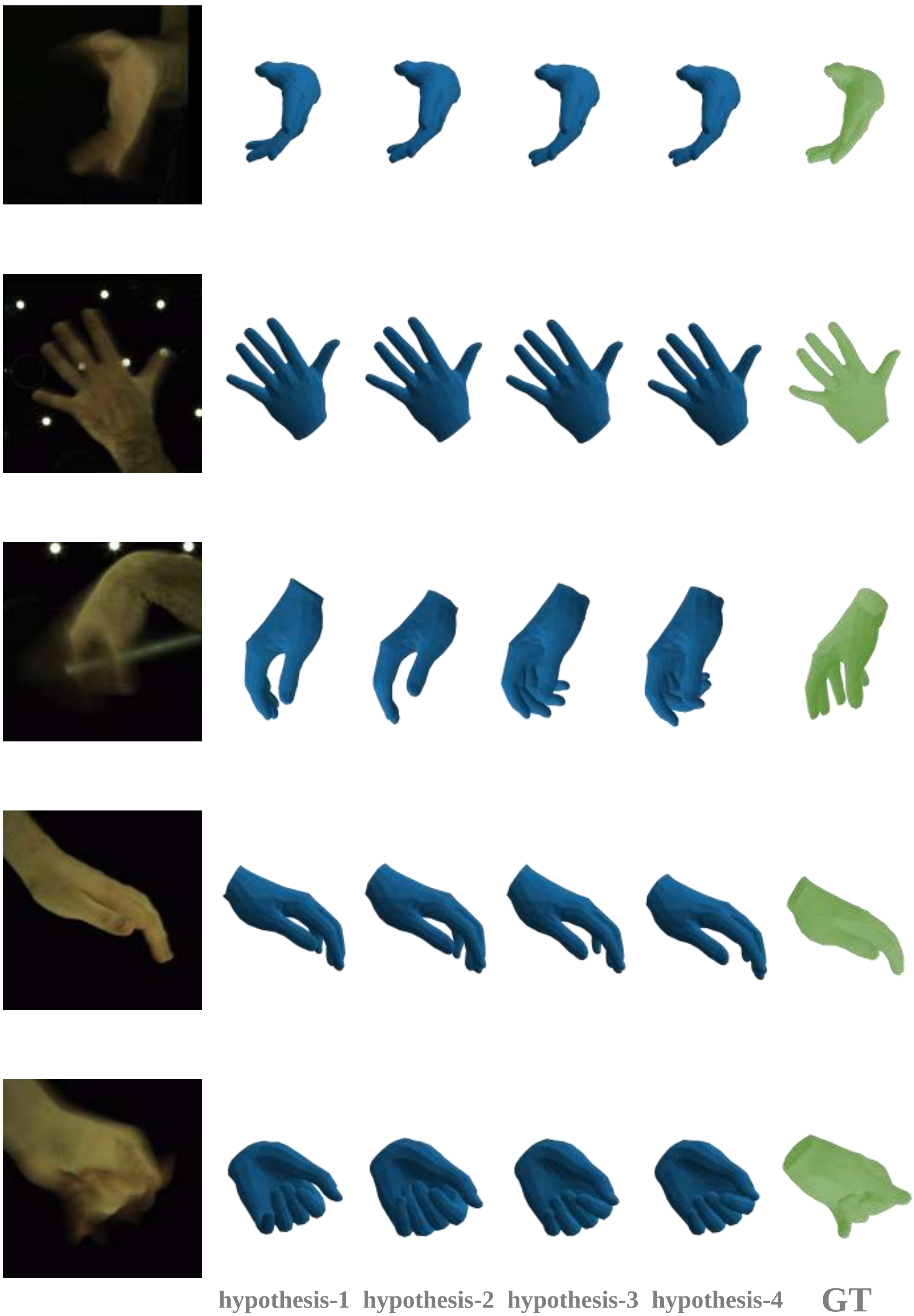

Fig. 1. More visualization cases at the middle frame of our method.

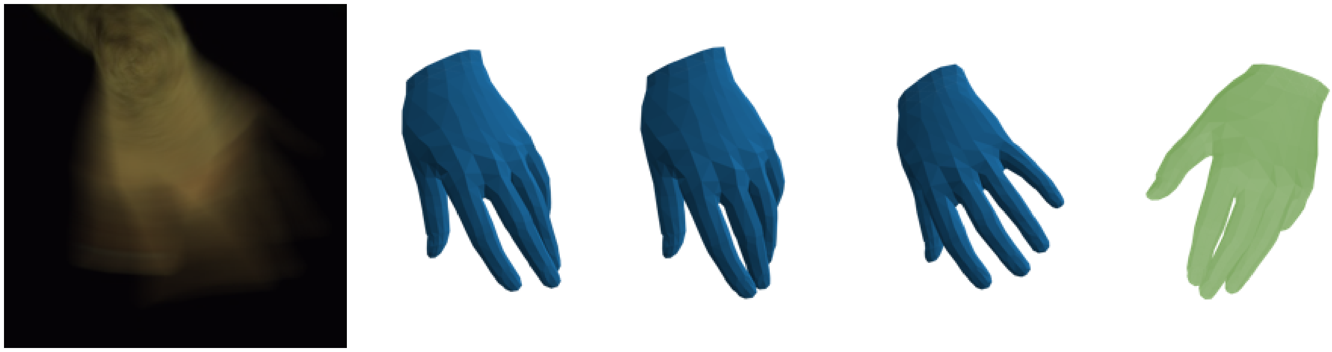

Fig. 2. One failure case that makes the wrong prediction at the middle frame due to the motion of hand.

Fig. 3. Visualization of mesh sequence in one image, past, middle, future represent the color of each timestep, respectively. It better shows how the hypothesis differs from each one.