Model-based Deep Hand Pose Estimation

Xingyi Zhou\(^1\), Qingfu Wan\(^1\), Wei Zhang\(^1\),
Xiangyang Xue\(^1\), Yichen Wei\(^2\)

\(^1\)Fudan University, \(^2\)Microsoft Research

\{zhouxy13, qfwan13, weizh, xyxue\}@fudan.edu.cn, \{yichenw@microsoft.com\}

We propose a model based deep learning approach that fully exploits the hand model geometry. We develop a new layer that realizes the non-linear forward kinematics, that is, mapping from the joint angles to joint locations. The layer is efficient, differentiable, parameter-free and serves as an intermediate representation in the network.

**Goal**

Given a depth image of human hand, estimate accurate 3D joint locations.

**Challenges**

- Highly articulated structure
- Significant self-occlusion
- Various viewpoint changes

**Previous Approaches**

Model based (Generative)

- Synthesize observation from hand geometry.
- Optimize the discrepancy to obtain the pose.
- Accurate but slow.

Learning based (Discriminative)

- Learn a direct regression function that maps the image appearance to hand pose.
- Efficient but suffer from invalid poses.

Hybrid Discriminative and Generative

- Discriminative method for initialization.
- Model based refinement.
- Separated multi-stages.

**Our Approach**

We propose a model based deep learning approach that fully exploits the hand model geometry. We develop a new layer that realizes the non-linear forward kinematics, that is, mapping from the joint angles to joint locations. The layer is efficient, differentiable, parameter-free and serves as an intermediate representation in the network.

**Contribution**

- For the first time, we show that the end-to-end learning using the non-linear forward kinematics layer in a deep neural network is feasible for hand pose estimation.
- We show that using joint location loss and adding an additional regularization loss on the intermediate pose representation are important for accuracy and pose validity.

**Code is available at**

https://github.com/tenstep/DeepModel

---

**Goal**

Given a depth image of human hand, estimate accurate 3D joint locations.

**Challenges**

- Highly articulated structure
- Significant self-occlusion
- Various viewpoint changes

**Previous Approaches**

Model based (Generative)

- Synthesize observation from hand geometry.
- Optimize the discrepancy to obtain the pose.
- Accurate but slow.

Learning based (Discriminative)

- Learn a direct regression function that maps the image appearance to hand pose.
- Efficient but suffer from invalid poses.

Hybrid Discriminative and Generative

- Discriminative method for initialization.
- Model based refinement.
- Separated multi-stages.

**Our Approach**

We propose a model based deep learning approach that fully exploits the hand model geometry. We develop a new layer that realizes the non-linear forward kinematics, that is, mapping from the joint angles to joint locations. The layer is efficient, differentiable, parameter-free and servers as an intermediate representation in the network.

**Contribution**

- For the first time, we show that the end-to-end learning using the non-linear forward kinematics layer in a deep neural network is feasible for hand pose estimation.
- We show that using joint location loss and adding an additional regularization loss on the intermediate pose representation are important for accuracy and pose validity.

**Code is available at**

https://github.com/tenstep/DeepModel