

基于姿势引导对齐的循环迭代三维人手姿态估计方法

作者: 邓小明, 左德鑫, 张寅达, 崔兆鹏, 程坚, 谭平, 常亮, Marc Pollefeys, Sean Fanello, 王宏安

Recurrent 3D Hand Pose Estimation using Cascaded Pose-guided 3D Alignments
IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, 2023

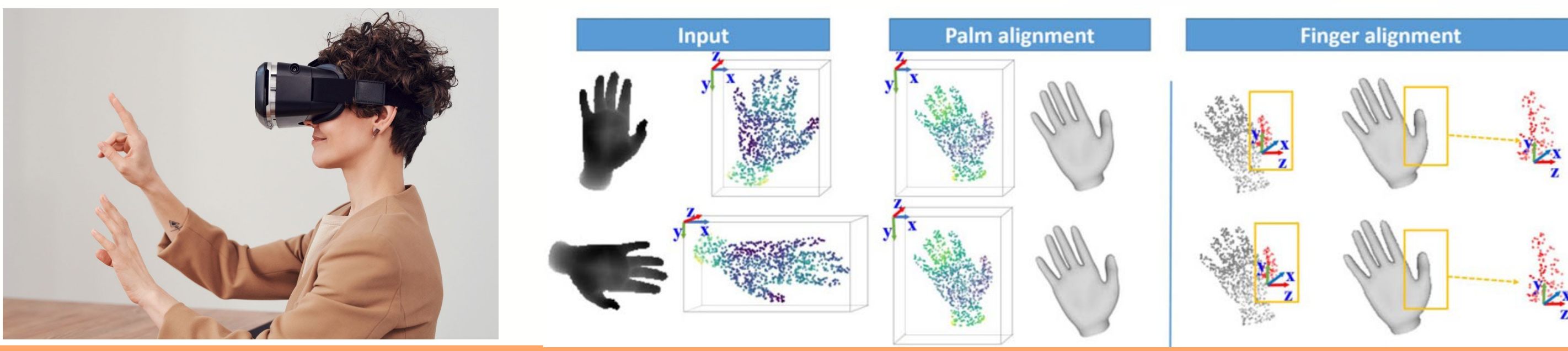
联系方式: 邓小明, 13717981135, xiaoming@iscas.ac.cn



视频 Video Demo

Motivation

- Hand pose estimation is one of the fundamental problems in computer vision and computer graphics, and has many applications in human-computer interactions and augmented reality.
- Hand pose estimation is challenging due to high degree of freedom. Visually different hands may have the same pose under proper alignments.
- Aligning data is an effective way to reduce the viewpoint variations of the input data. However, previous alignments for hand pose estimation are restricted in alignment transformation in 2D image plane.



Introduction

We design a hand pose estimation model to achieve the full 3D hand pose, perform both data alignment and feature extraction in 3D space. Our main contributions:

- We propose the first pose-guided data alignment of 3D point clouds for 3D hand pose estimation.
- We present a new recurrent hand pose module for aligned 3D representation that can extract recurrent pose-aware feature and iteratively refine the estimated hand pose.

Network Architecture

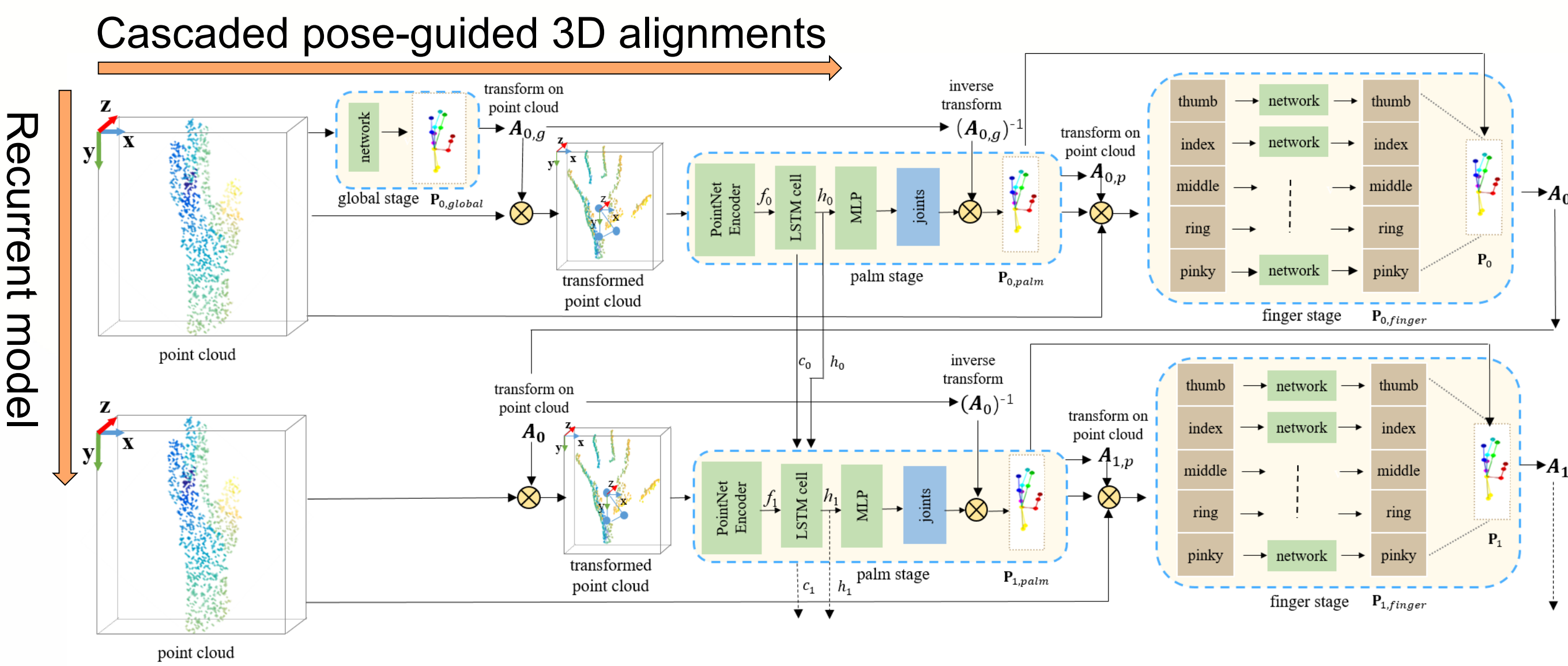


Illustration of recurrent hand pose network using cascaded pose-guided alignments

We first convert the input hand foreground depth to point cloud. Then we adopt multiple recurrent iterations to estimate the 3D hand pose. Specifically, we introduce several LSTM modules among multiple palm stages to refine the hand pose. In each recurrent iteration, we adopt a multi-stage network to predict hand joints by iterative pose regression and cascaded pose-guided 3D alignment, and we adopt the hand pose of the previous iteration to align the input point cloud of the current iteration.

Cascaded Pose-guided 3D Alignments

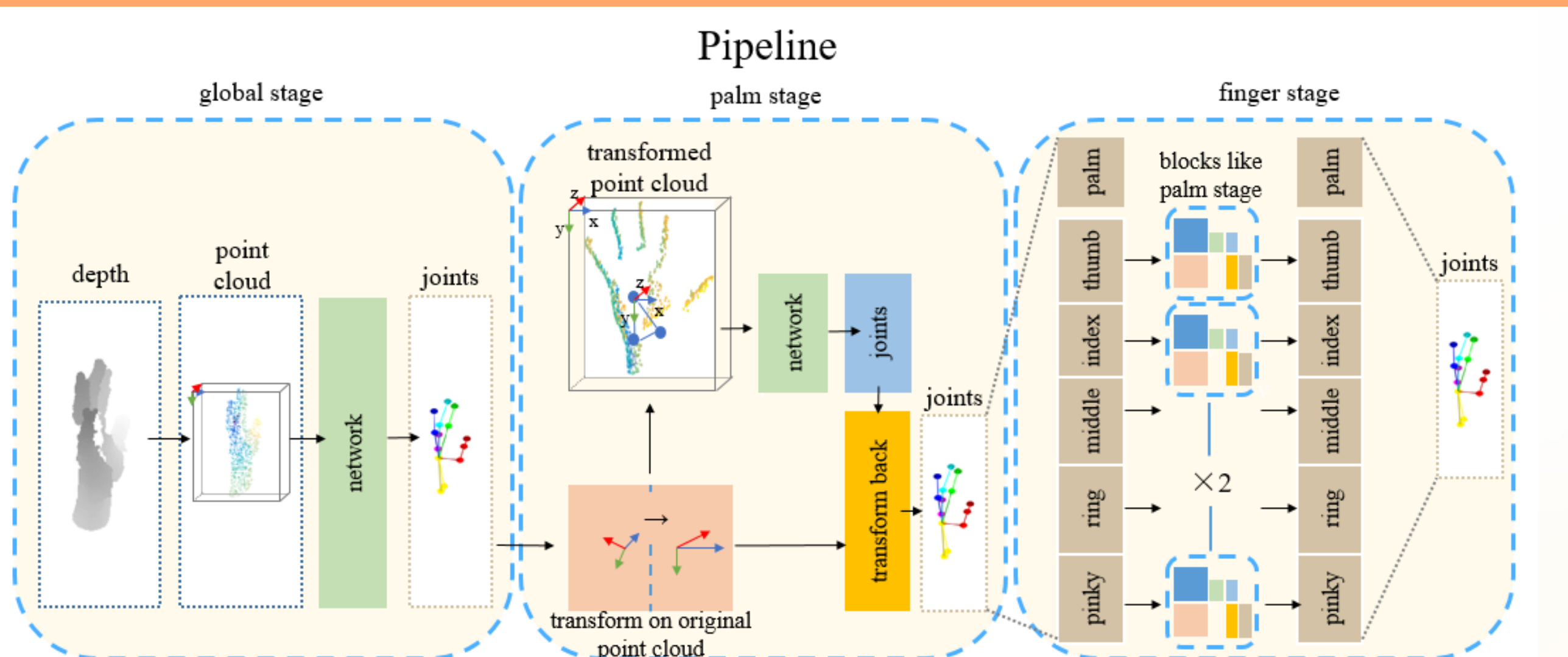
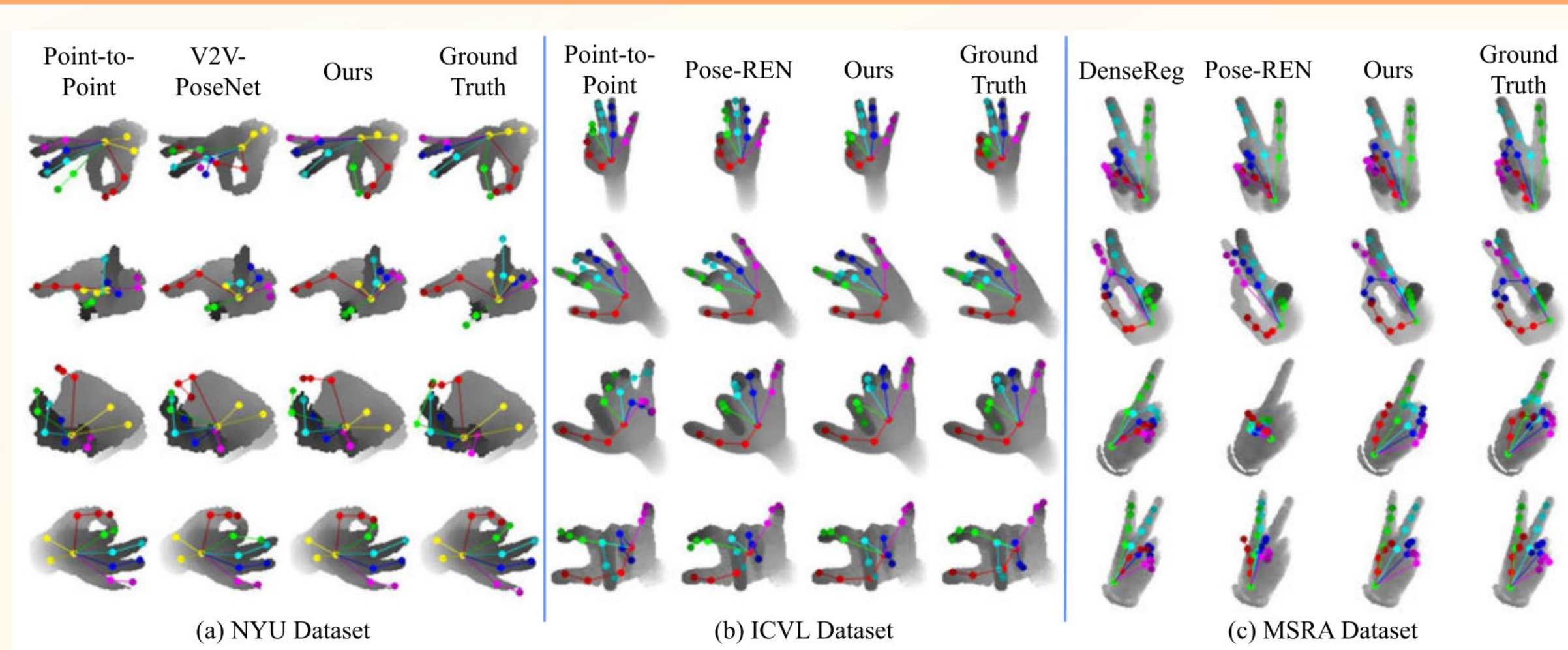


Illustration of a cascaded architecture guided by pose-guided 3D alignments

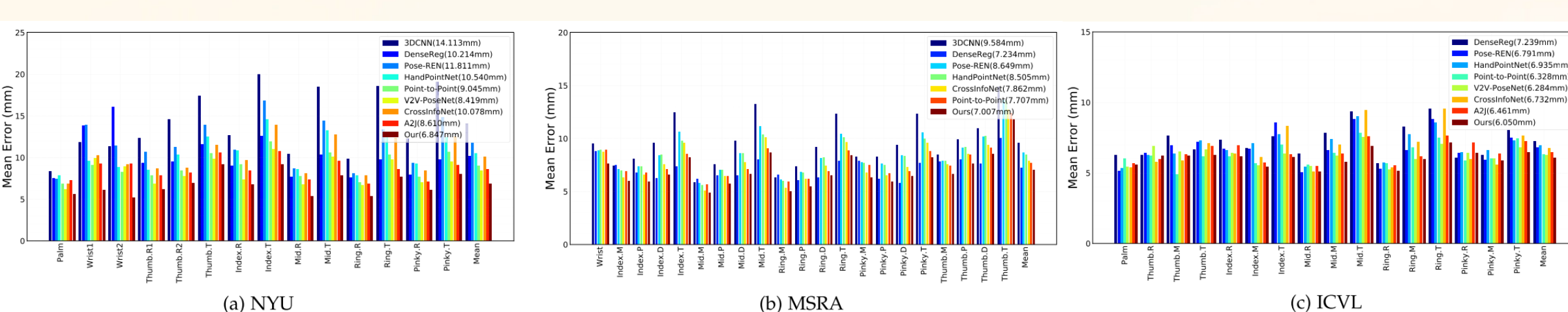
For each recurrent iteration, we design a cascaded architecture guided by pose-guided 3D alignments.

- Global stage produces a coarse hand pose in camera coordinate.
- Palm stage produces refined palm joints using aligned 3D point cloud with palm coordinate.
- Finger stage produces refined finger joints using aligned 3D point cloud with finger coordinate.
- Final hand pose is composed by the estimated hand pose from palm stage and finger stage of the last recurrent iteration.

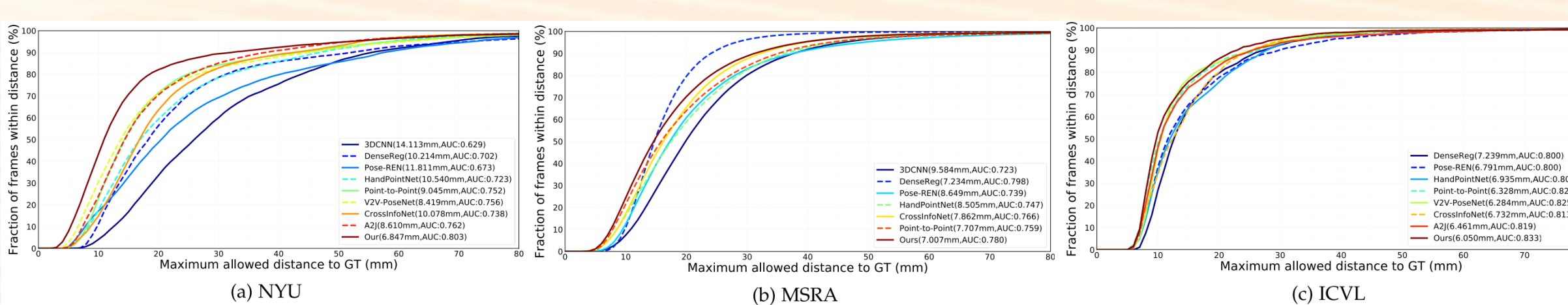
Comparison with SOTA



Qualitative results for NYU, ICVL and MSRA datasets with SoTA methods.



Comparison to SoTA methods on NYU, MSRA, ICVL datasets. We show mean joint errors for all the test examples.



Comparison to SoTA methods on NYU, MSRA, ICVL datasets. We show percentage of frames in the testing examples under different error thresholds

Coordinate Systems for the Alignments

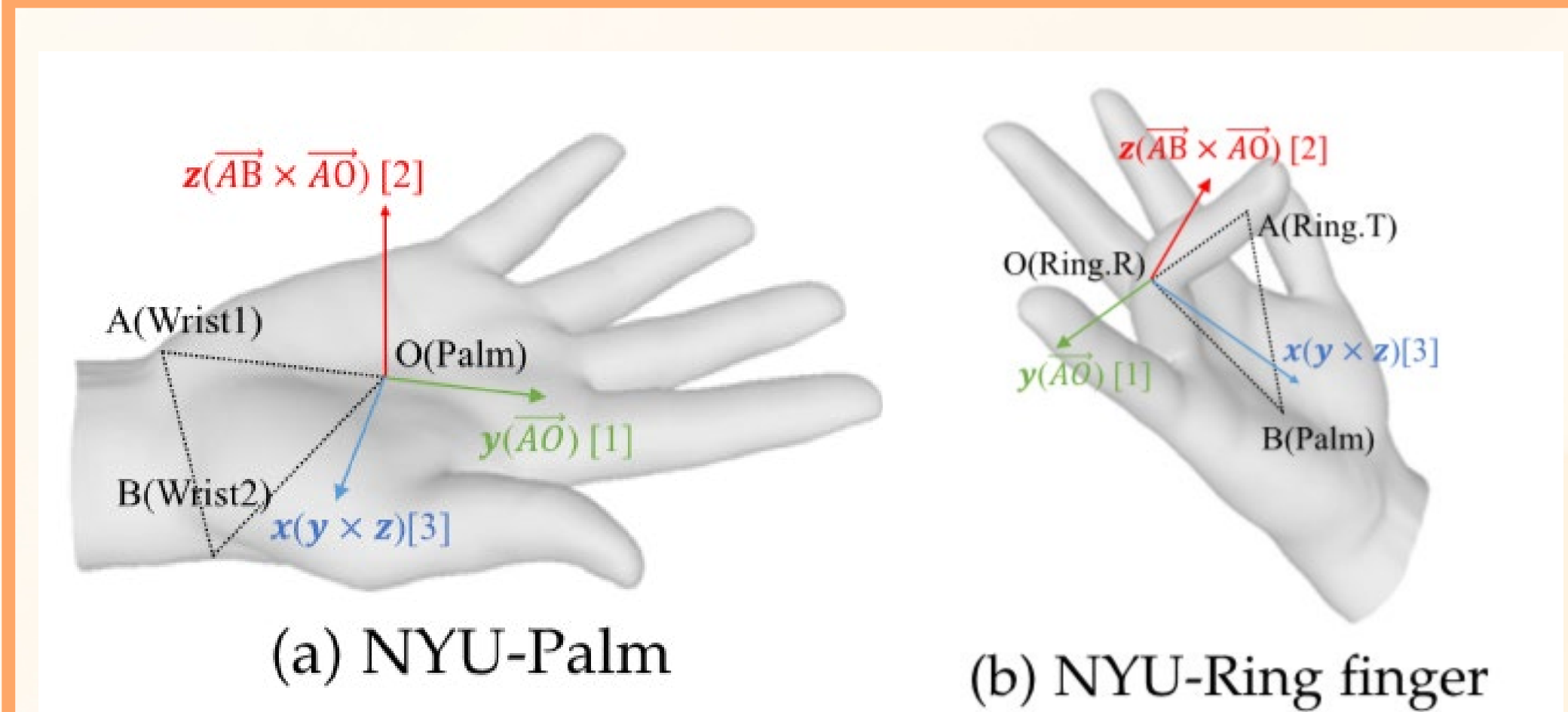
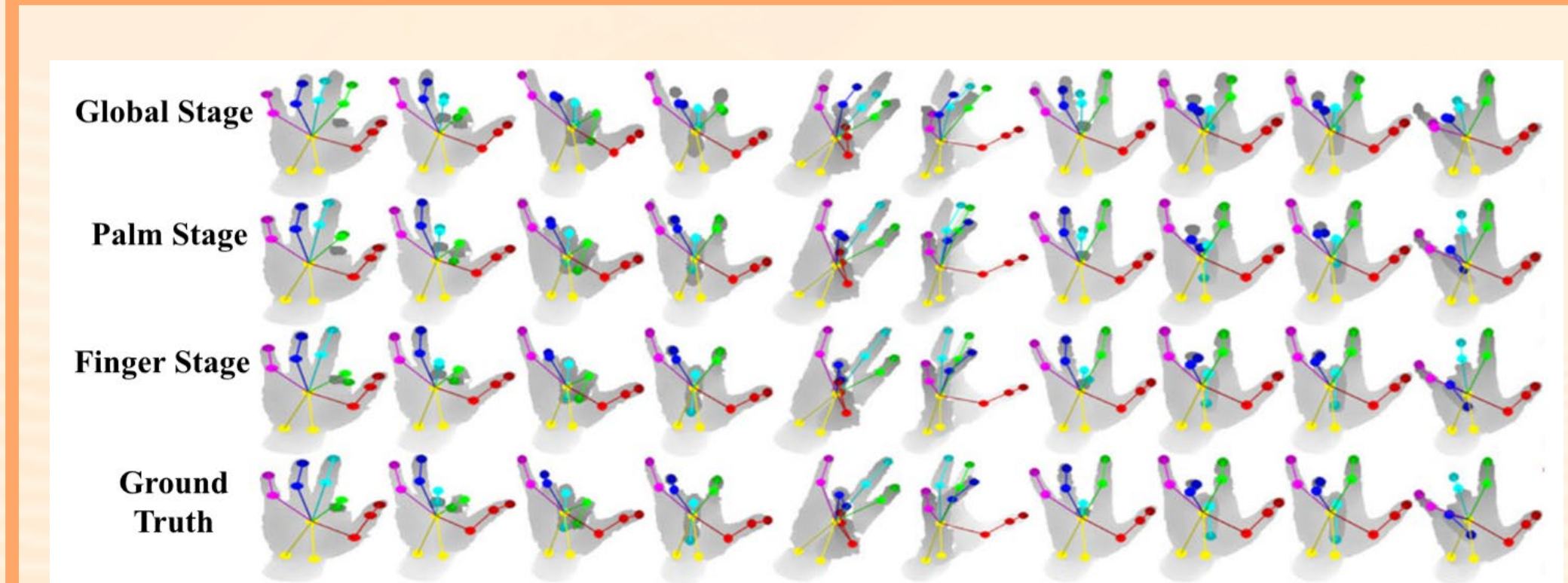


Illustration of the coordinate systems for the pose-guided alignments. (a)(b) show palm and ring finger coordinate systems of NYU datasets.

Qualitative Results



Qualitative results of different stages on NYU dataset. We show hand pose estimation results of global stage, palm stage and finger stage