Robust 6DoF Pose Estimation Against Depth Noise and a Comprehensive Evaluation on a Mobile Dataset

Zixun Huang*[1] Keling Yao*[1 2] Seth Z. Zhao*[1] Chuanyu Pan*[1] Chengfeng Xu[1]
Kathy Zhuang[1] Tianjian Xu[1] Weiyu Feng[1] Allen Y. Yang[1]

Equal contribution [1] University of California, Berkeley [2] The Chinese University of Hong Kong, Shenzhen.
Correspondence to: Allen Y. Yang <yang@eecs.berkeley.edu>

https://openark-berkeley.github.io/DTTDNet/

Webpage

Background

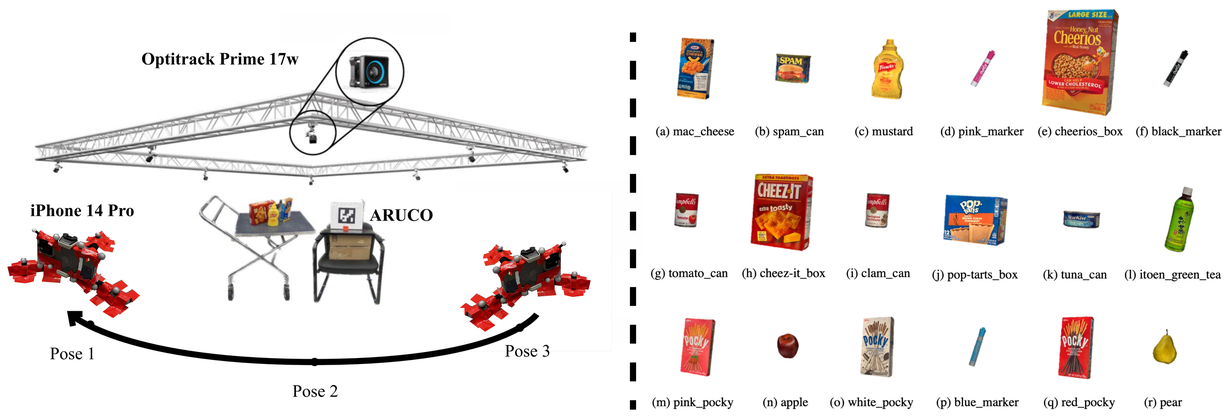
Six-degrees-of-freedom (6DoF) pose estimation determines an object's position and orientation in 3D space. This technology is crucial for applications like augmented reality (AR), where accurate object localization enhances user experience. Existing 6DoF pose estimation methods face challenges with depth sensor noise, leading to significant performance discrepancies.

Proposed Solution

We present a simple and effective transformer-based 6DoF pose estimation approach called DTTDNet, featuring a novel geometric feature filtering module and a Chamfer distance loss for training.

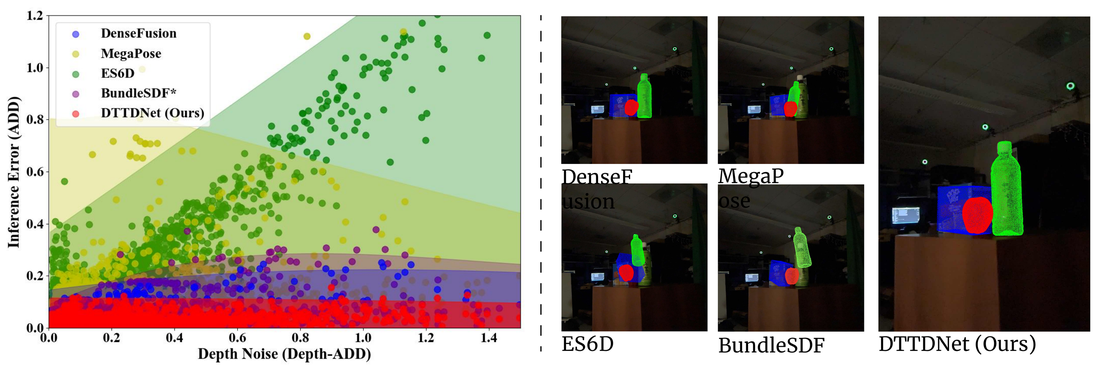
Moreover, we advance the field of robust 6DoF pose estimation and introduce a new dataset -- Digital Twin Tracking Dataset Mobile (DTTD-Mobile), tailored for digital twin object tracking with noisy depth data from the mobile RGBD sensor suite of the Apple iPhone 14 Pro.

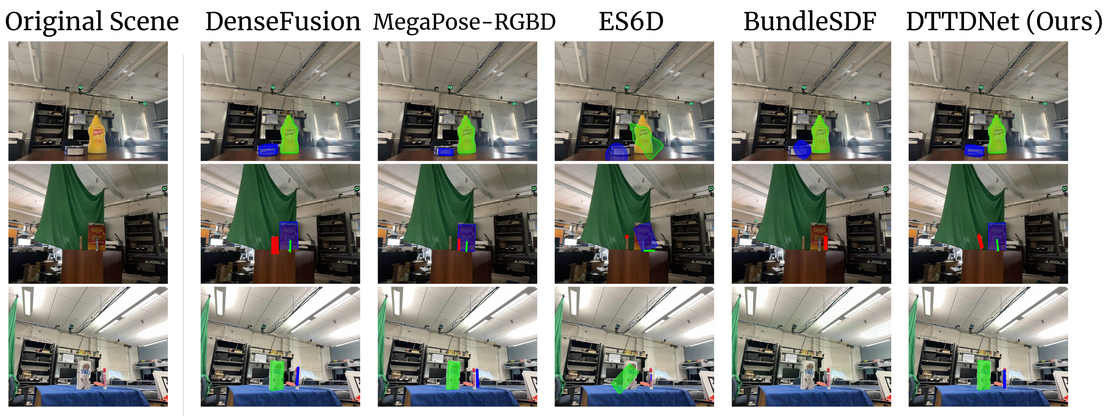
Digital Twin Tracking Dataset Mobile (DTTD-Mobile)



We introduce DTTD-Mobile, collected using the Apple iPhone 14 Pro. It includes RGBD data under varying conditions, such as different lighting and occlusion scenarios. This dataset is designed to test and benchmark pose estimation algorithms in mobile environments.

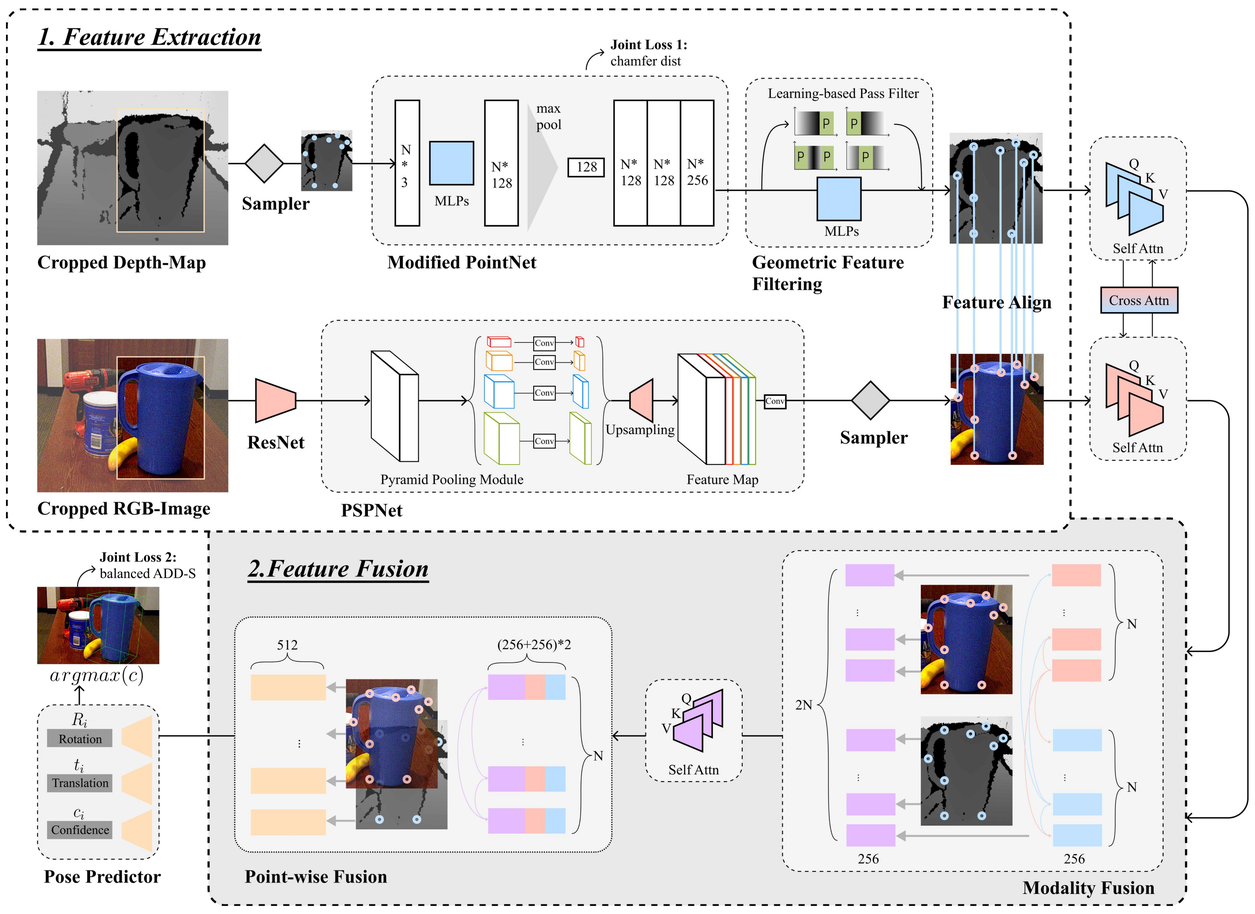
Results





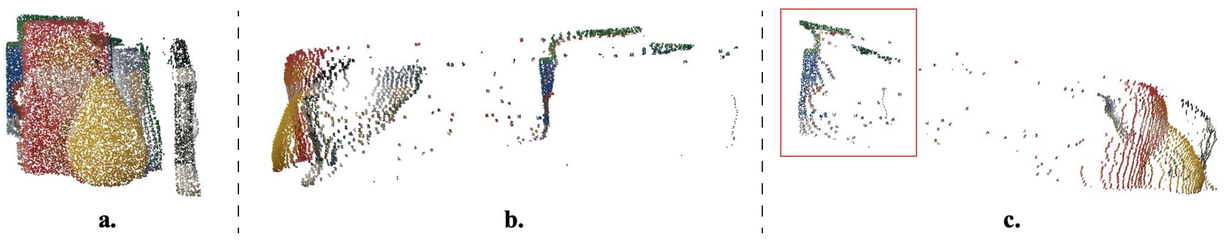
Extensive experiments demonstrate that DTTDNet significantly outperforms state-of-the-art methods by least 4.32, and up to 60.74 points in ADD metrics on DTTD-Mobile. More importantly, our approach exhibits superior robustness to varying levels of measurement noise, setting a new benchmark for robustness to noise measurements.

Model Architecture Overview



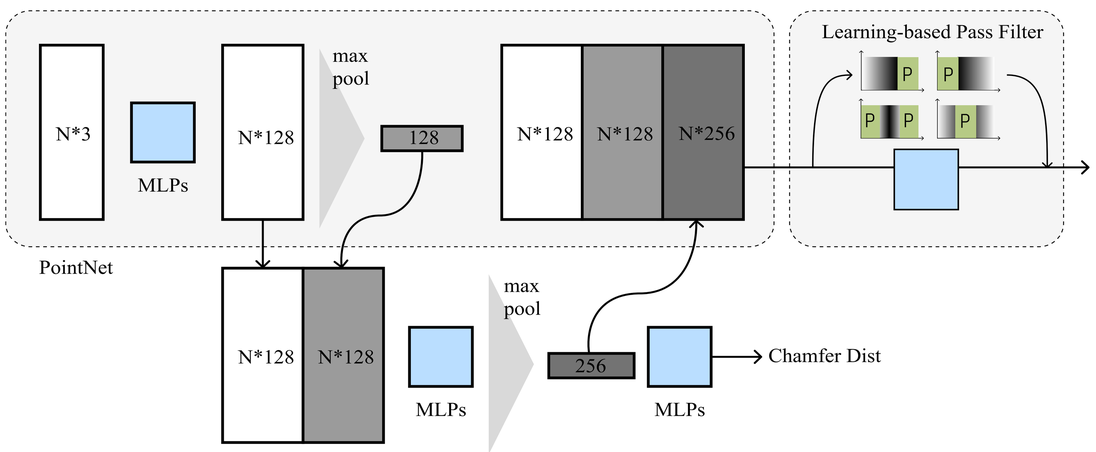
DTTDNet pipeline starts with segmented depth maps and cropped RGB images. The point cloud from the depth map and RGB colors are encoded and integrated point-wise. Extracted features are then fed into an attention-based two-stage fusion. Finally, the pose predictor produces point-wise predictions with both rotation and translation.

Challenges



One central challenge we faced was assessing the robustness of current pose estimation methods in handling distortion and interpolation noise. This challenge stemmed from the specific characteristics of depth sensors, especially in the scenarios where filling holes by interpolating depth maps lead to long-tailed noisy data.

Robustifying Depth Data: Chamfer Distance Loss and Geometric Feature Filtering



To handle noisy and low-resolution LiDAR data in DTTDNet, we introduce Chamfer Distance Loss (CDL) and Geometric Feature Filtering (GFF). CDL uses a point-cloud reconstruction task and Chamfer distance as a loss function to filter out noise from the depth data. GFF incorporates Fast Fourier Transform (FFT) into geometric feature encoding to select significant features from noisy input signals, providing a refined embedding resilient to non-Gaussian LiDAR noise.