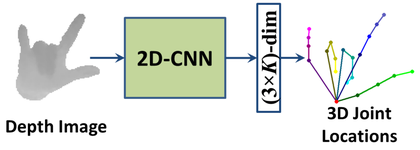
3D Convolutional Neural Networks for Efficient and Robust Hand Pose Estimation from Single Depth Images

Liuhao Ge, Hui Liang, Junsong Yuan, and Daniel Thalmann
Institute for Media Innovation, Nanyang Technological University, Singapore

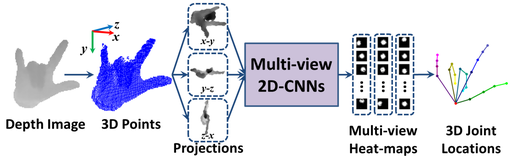
IEEE 2017 Conference on Computer Vision and Pattern Recognition
CVPR July 21-26 2017 HONOLULU

Introduction

- **Objective**: to estimate 3D hand pose in real-time from single depth images.
- **Motivation**: image based features extracted by 2D CNNs are not directly suitable for 3D hand pose estimation due to the lack of 3D spatial information.
- **2D CNN**:



- **Multi-view CNNs [1]**: still cannot effectively exploit 3D spatial information in the depth image; the computational complexity will be increased when using more views.



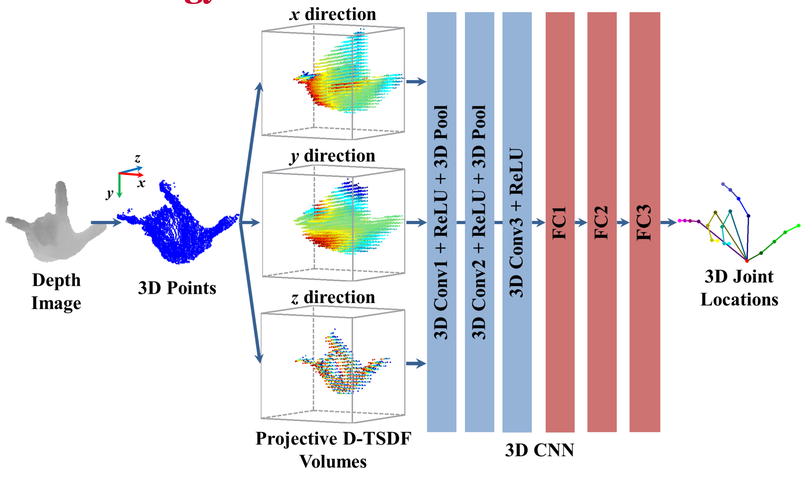
- **Our Approach**: we propose a 3D CNN-based approach that can capture the 3D spatial structure of the input and accurately regress full 3D hand pose in a single pass.

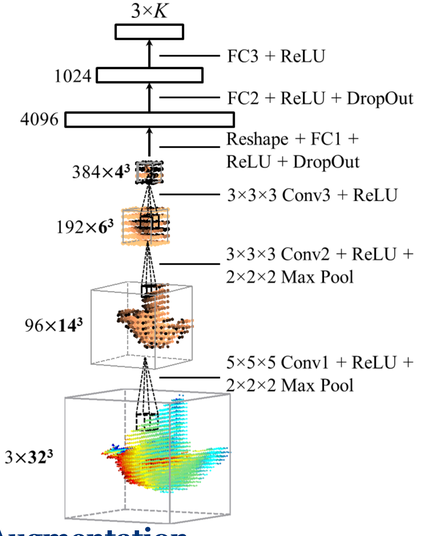
References

[1] L. Ge, H. Liang, J. Yuan, and D. Thalmann. Robust 3D hand pose estimation in single depth images: from single-view CNN to multi-view CNNs. In *CVPR*, 2016.

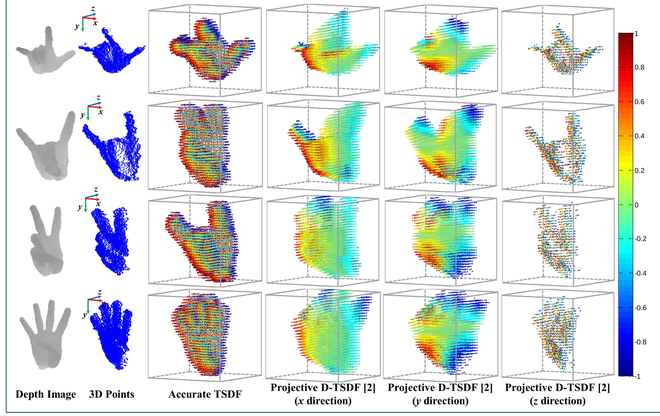
[2] S. Song and J. Xiao. Deep Sliding Shapes for amodal 3D object detection in RGB-D images. In *CVPR*, 2016.

Methodology

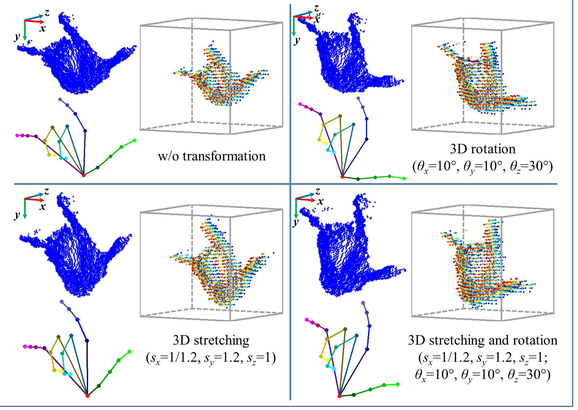




Volumetric Representation



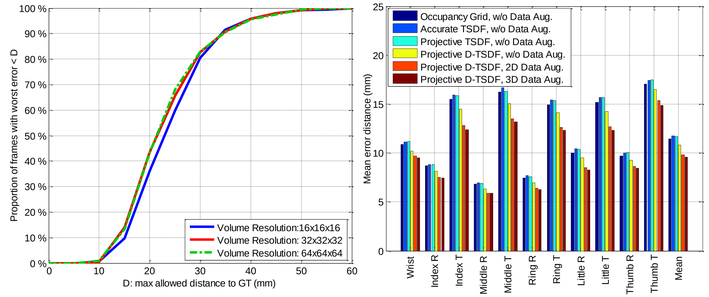
3D Data Augmentation



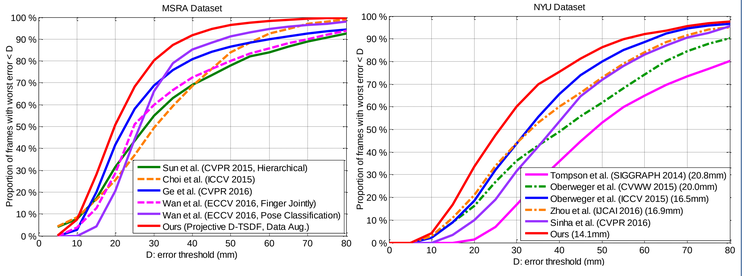
Acknowledgement This research is supported by the BeingTogether Centre, a collaboration between Nanyang Technological University and University of North Carolina at Chapel Hill. This work is also supported in part by Singapore Ministry of Education Academic Research Fund.

Experiments

Self-comparisons



Comparisons with State-of-the-art



Qualitative Results