

# DeepPrior++: Improving Fast and Accurate 3D Hand Pose Estimation

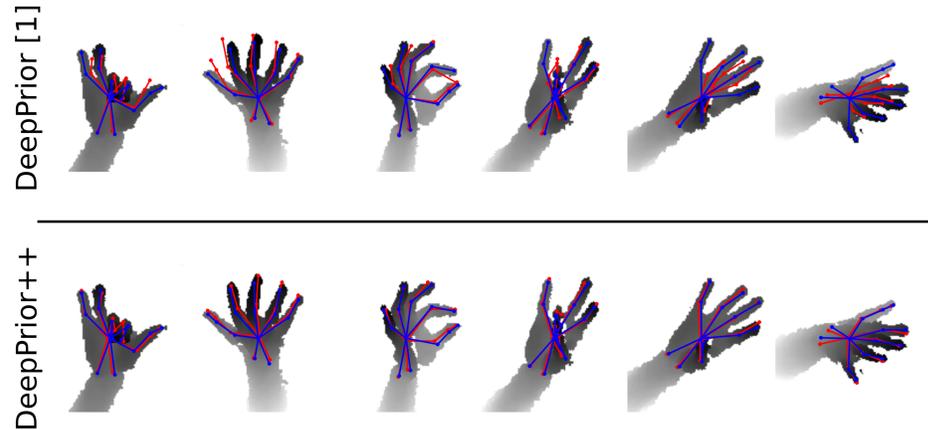
Markus Oberweger and Vincent Lepetit

Graz University of Technology, Institute for Computer Vision and Graphics, icg.tugraz.at



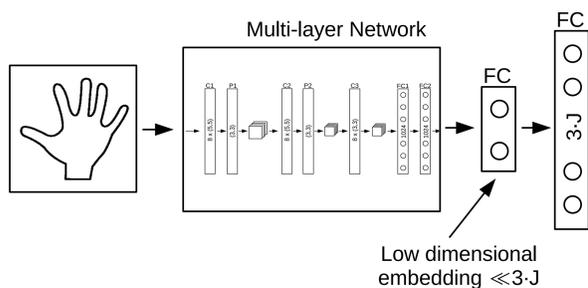
## Motivation

- We show that, with simple improvements, we achieve **SOTA performance** on three main benchmarks (NYU, ICVL, MSRA) while keeping the simplicity of the DeepPrior [1] method.
- Up to **38% smaller error** using same training data
- Code available** online  
[github.com/moberweger/deep-prior-pp](https://github.com/moberweger/deep-prior-pp)



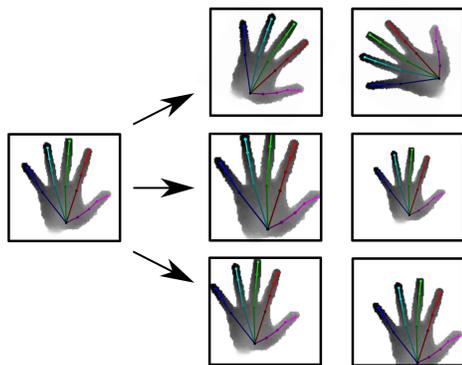
## Method

- Based on DeepPrior [1]
  - Integrate prior on 3D hand pose into CNN

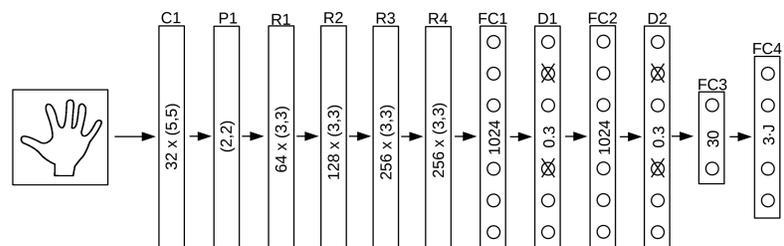


- Improved training data augmentation

- Rotation
- Scaling
- Translation
- Online during training:
  - >10M different samples
- Robust prior with augmented 3D poses

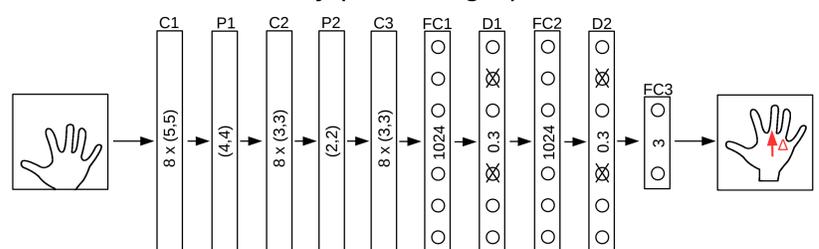


- More powerful network architecture
  - Based on Residual Network (ResNet)
  - Regularization using Dropout



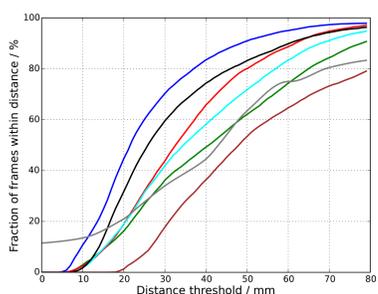
- Refined hand localization

- Detect hand using center-of-mass
- Refine localization by predicting update on 3D location

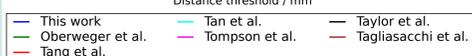
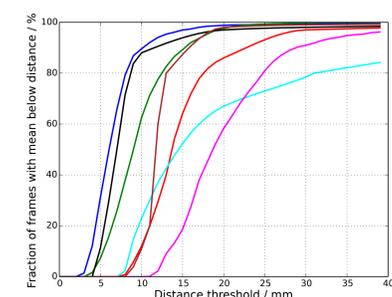
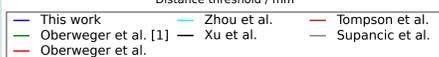


## Results

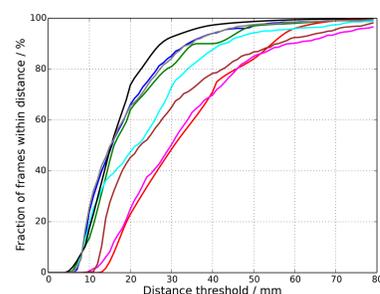
- Excellent results on main benchmark datasets



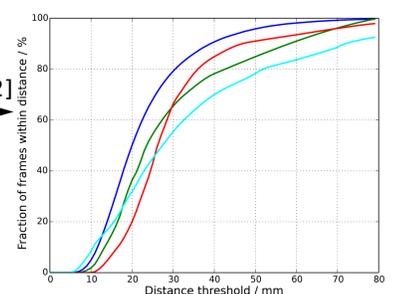
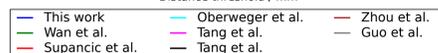
NYU [4]



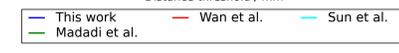
Method	Avg. 3D error
Oberweger et al. (DeepPrior)	19.8mm
Oberweger et al. (Feedback)	16.2mm
Deng et al. (Hand3D)	17.6mm
Guo et al. (REN)	13.4mm
Bouchacourt et al. (DISCO)	20.7mm
Zhou et al. (DeepModel)	16.9mm
Xu et al. (Lie-X)	14.5mm
Neverova et al.	14.9mm
Wan et al. (Crossing Nets)	15.5mm
Fourure et al. (JTSC)	16.8mm
Zhang et al.	18.3mm
Madadi et al.	15.6mm
This work (DeepPrior++)	<b>12.3mm</b>



MSRA [2]



ICVL [3]



- Self-comparison

Training data augmentation	
Augmentation	Avg. 3D error
No augmentation	19.9mm
Translation (T)	14.7mm
Rotation (R)	13.8mm
Scale (S)	17.1mm
All (R+T+S)	12.3mm
All (R+T+S) no prior aug.	21.7mm

Network architecture		
Localization	Avg. 3D error	fps
Original [1]	16.6mm	100
Original more filters	13.7mm	80
ResNet	12.3mm	30

### Localization accuracy

Localization	Avg. 3D pose error	Avg. 3D loc. error
CoM	13.8mm	28.1mm
Refined CoM	12.3mm	8.6mm
Ground truth	10.8mm	0.0mm

[1] M. Oberweger, P. Wohlhart, and V. Lepetit. Hands Deep in Deep Learning for Hand Pose Estimation. In Proc. of CVWW, 2015.  
 [2] X. Sun, Y. Wei, S. Liang, X. Tang, and J. Sun. Cascaded Hand Pose Regression. In CVPR, 2015.  
 [3] D. Tang, H. J. Chang, A. Tejani, and T.-K. Kim. Latent Regression Forest: Structured Estimation of 3D Articulated Hand Posture. In CVPR, 2014.  
 [4] J. Tompson, M. Stein, Y. LeCun, and K. Perlin. Real-Time Continuous Pose Recovery of Human Hands Using Convolutional Networks. ACM Transactions on Graphics, 33, 2014.