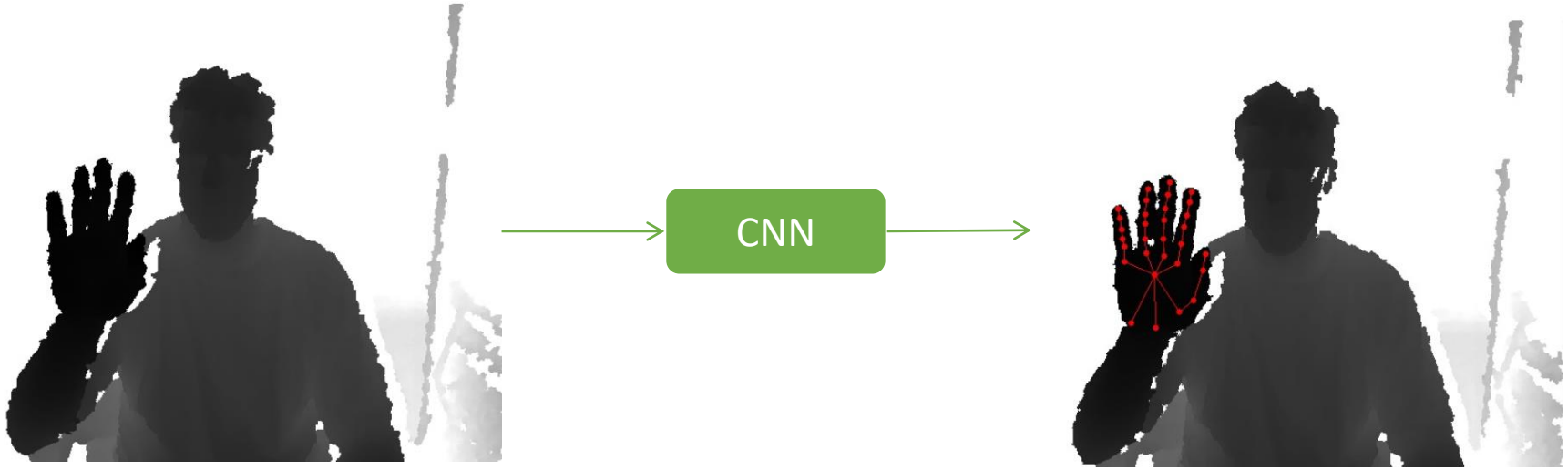
ICPR2020
25th INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION
Milan, Italy 10 | 15 January 2021

PEAN: 3D Hand Pose Estimation Adversarial Network

Linhui Sun(Presenter), Yifan Zhang, Jian Cheng, Hanqing Lu

Preliminaries

- Hand pose estimation is to predict the 3D coordinates in the real world of each joint from the input depth image.
- Accurate hand pose estimation is important for augmented reality and human-computer interaction.



Motivation

- Most of the previous methods took the 2D depth images as inputs and directly estimated 3D coordinates of joints through 2D CNN. However, such framework can not make full use of the depth information.

- The predicted pose should conform to kinematic constraints. However, such constrains are not employed to most of deep learning based methods.

Contributions

- We propose a novel 3D hand pose estimation adversarial network (PEAN) for hand pose estimation, which can implicitly utilize physical constraints to regularize the prediction in an adversarial learning framework.

- A new hierarchical model 3DHNet is designed for hand pose estimation by decoupling the task into multiple subtasks.

- We propose a PDNet for judging the reasonableness of the predicted poses. By introducing an adversarial learning, the accuracy and reasonableness of the predicted results can be significantly improved.

- The proposed PEAN achieves the state-of-the-art results on three challenging hand pose estimation datasets: NYU, MSRA , and ICVL.

Method

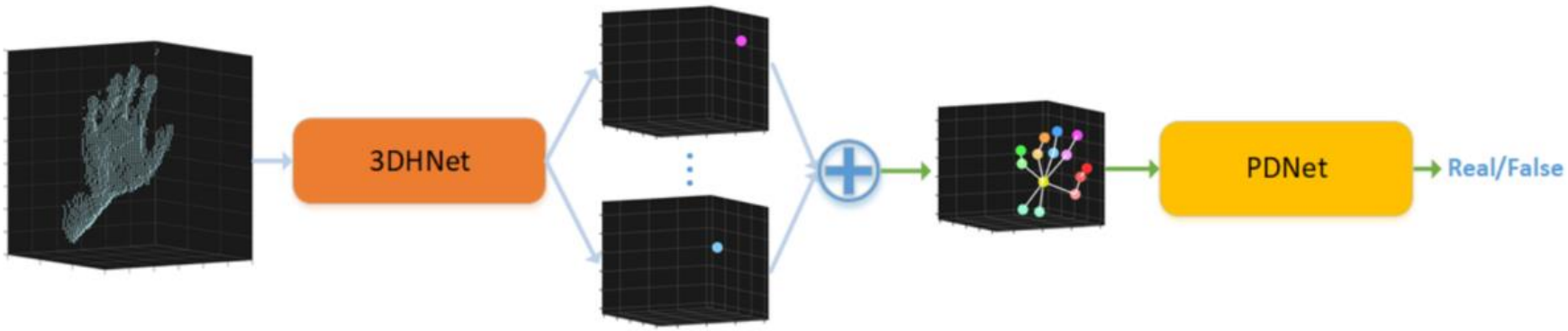


Fig. 1. Overall structure of PEAN. PEAN is comprised of two components: a hierarchical hand pose estimation network, named 3DHNet, and a pose discrimination network, named PDNet.

Method

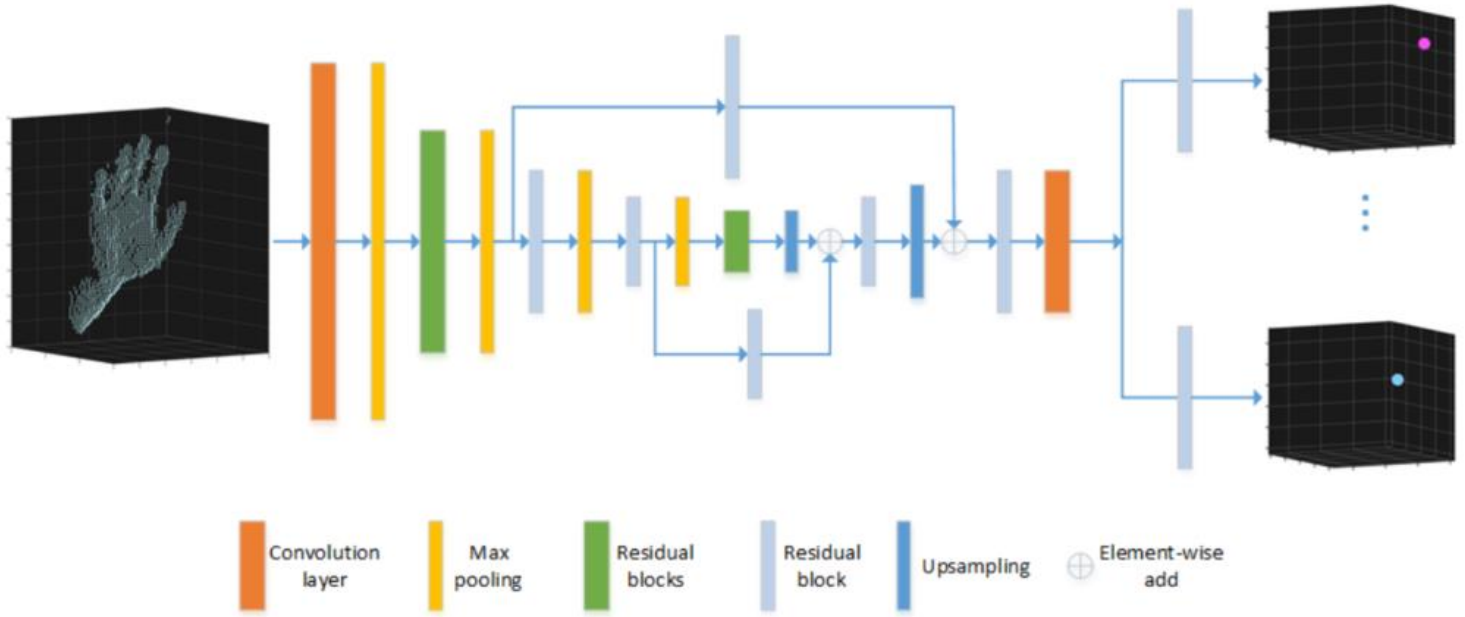


Fig. 2. Network structure of 3DHNet. Based on 3D CNN, the input of 3DHNet is the volumetric form of hand and the output is 3D heatmap of each joint.

Method

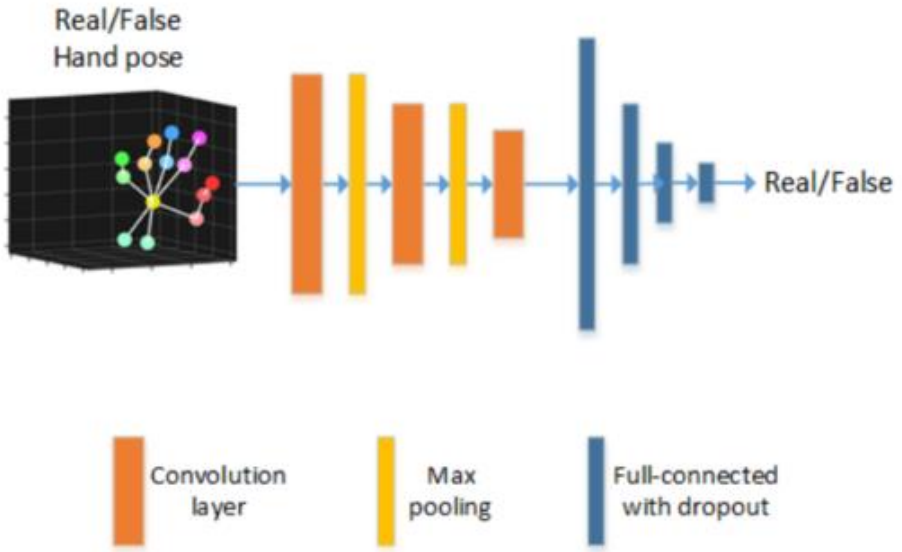


Fig. 3. Network structure of PDNet. The input of PDNet is 3D hand pose formed by ground truth or formed by the predicted results of 3DHNet, which will be judged as real or false pose by PDNet.

Method

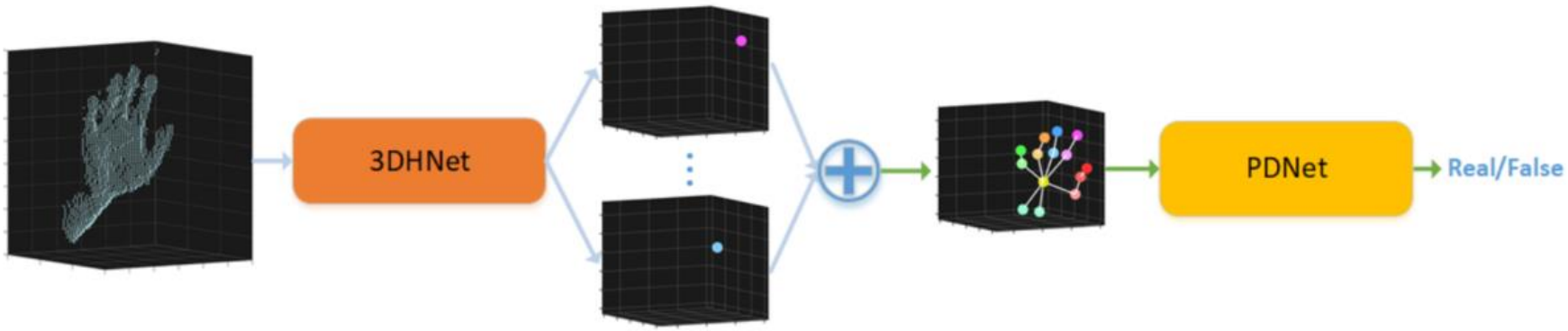


Fig. 1. Overall structure of PEAN. PEAN is comprised of two components: a hierarchical hand pose estimation network, named 3DHNet, and a pose discrimination network, named PDNet.

Experiments

- Ablation study:

TABLE I
ABLATION STUDY RESULTS ON AVERAGE 3D DISTANCE ERROR(mm). BASE: THE ESTIMATED RESULTS OF 3DHNET WITHOUT JOINTLY TRAINING WITH PDNET; BASE+PDNET: THE ESTIMATED RESULTS OF 3DHNET JOINTLY TRAINING WITH PDNET, WHICH CAN ALSO BE REGARDED AS THE RESULTS OF PEAN.

Structures	Average 3D distance error / mm		
	NYU	MSRA	ICVL
Base	8.73	7.58	6.28
Base+PDNet	8.59	6.91	5.85

Experiments

- Compared with state-of-the-art methods:

TABLE II

COMPARISONS WITH STATE-OF-THE-ART METHODS ON ICVL, NYU AND MSRA DATASETS. RED, BLUE AND GREEN REPRESENT THE BEST, THE SECOND BEST AND THE THIRD BEST RESULT ON THE CORRESPONDING DATASET, RESPECTIVELY. * INDICATES THAT THIS METHOD USED EPOCH ENSEMBLE,WHICH WILL IMPROVE THE ESTIMATION ACCURACY.

Methods	Average 3D distance error / mm		
	ICVL	NYU	MSRA
Feedback	-	-	15.97
Lie-X	-	-	14.51
DeepPrior	10.4	19.73	-
DeepPrior++	8.1	12.24	9.5
DeepModel	11.56	17.04	-
Pose-Ren	6.79	11.81	8.65
Ren-4x6x6	7.63	13.39	-
Ren-9x6x6	7.31	12.69	9.7
DensReg	7.3	10.27	**7.2**
CrossInfoNet	6.73	10.08	7.86
3DCNN	-	14.1	9.6
SHPR-Net	7.22	10.78	7.86
HandPointNet	6.94	10.54	8.5
Point-to-Point	6.3	9.1	7.7
V2V-PoseNet*	**6.28**	**8.42**	7.59
3DHNet(ours)	**6.28**	**8.73**	**7.58**
PEAN(ours)	**5.85**	**8.59**	**6.91**

Thanks for your attention!

PEAN: 3D Hand Pose Estimation Adversarial Network