Rotation-invariant Mixed Graphical Model Network for 2D Hand Pose Estimation

Deying Kong[1], Haoyu Ma[1], Yifei Chen[2], and Xiaohui Xie[1]

[1] Department of Computer Science, University of California at Irvine

[2] Tencent Hippocrates Research Lab

Introduction

- **Objective**: 2D hand pose estimation (keypoint detection)
- **Application**: AR/VR, gesture recognition, basic for 3D task.
- **Challenge**: self-occlusion due to articulation, viewpoint and object.
- **Current Approach**:
  - *Deep convolutional neural network (DCNN)*: Convolutional Pose Machines (CPM) is commonly used in 2D hand pose estimation, however, it only captures pose structure information implicitly.
  - *Graphical Model (GM)*: Spatial constrains among body parts can be modeled explicitly, however, studies usually apply a GM with fixed parameters, which limits its ability to model a variety of pose.
- **Our Contributions**:
  - We propose a new model named R-MGMN which combines graphical model and deep convolutional neural network efficiently.
  - R-MGMN has several independent graphical models which can be selected softly based on image, instead of only one GM.

Mixed Graphical Model for Hand Pose

- **Basic graphical model**: $P^{basic}(X|I_{rt}) = \underbrace{\prod_{v_i \in V} \phi_i(xi|I_{rt})}_{\text{unary function}} \underbrace{\prod_{(j,k)\in \varepsilon} \psi(xj, xk|I_{rt})}_{\text{pairwise function}}$

$V = \{v_1, v_2, \ldots, vk\}$ denote the set of hand keypoints, each of which is associate with a variable x_i representing its 2D position

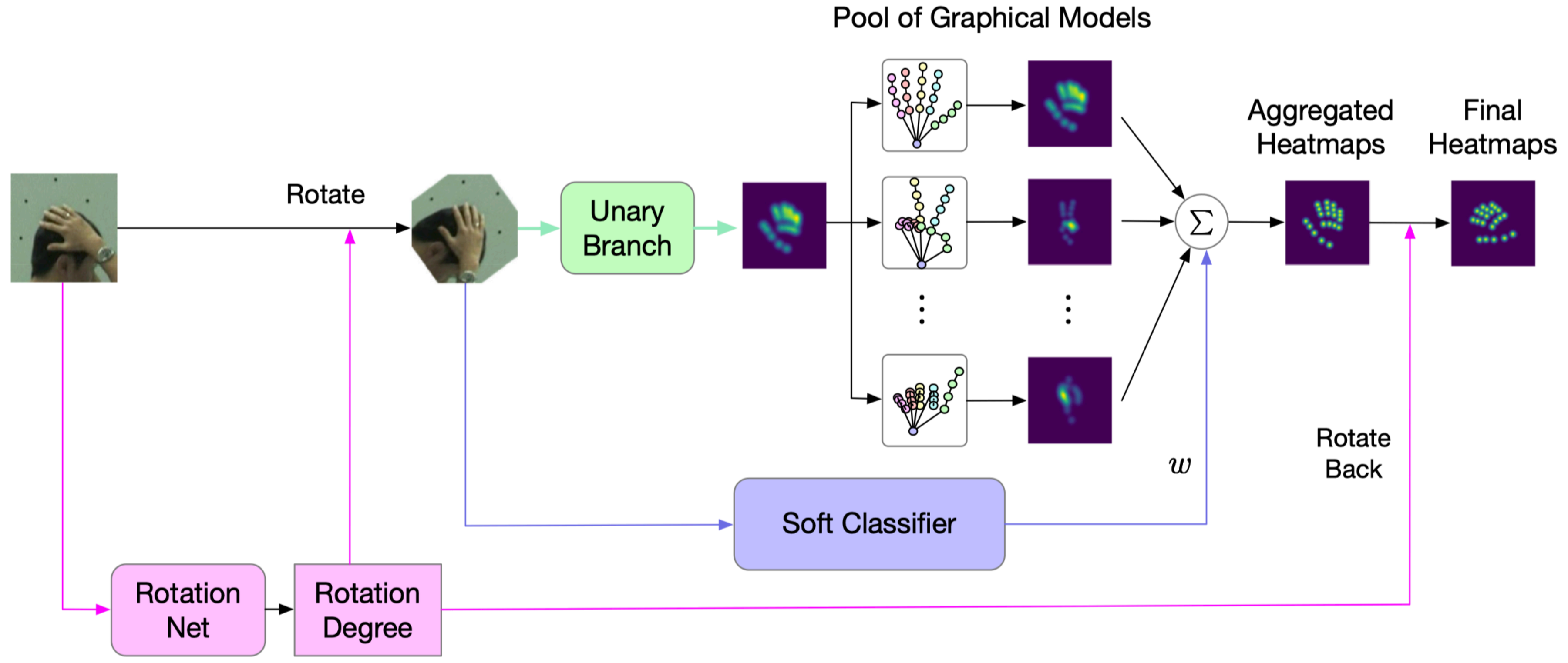
- **Mixed graphical model**: $P(X|I_{rt}) = \sum_{l=1}^{L} w_l \prod_{v_i \in V} \phi_i(xi|I_{rt}) \prod_{(j,k)\in \varepsilon} \psi(xj, xk|I_{rt})$

Marginal probability could be calculated by summing up the marginal probabilities of each individual graphical models, which could be calculated using message passing.

$$P(x_i|I_{rt}) = \sum_{l=1}^{L} w_l P_l(x_i|I_{rt})$$

Methodology

- **Network Architecture:**

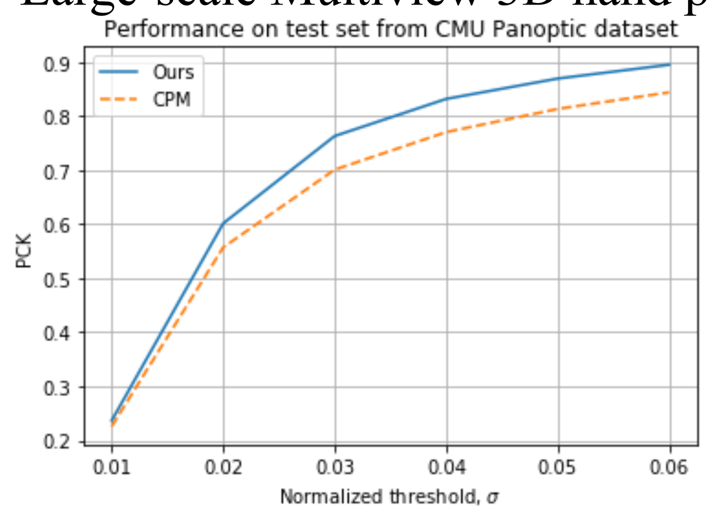


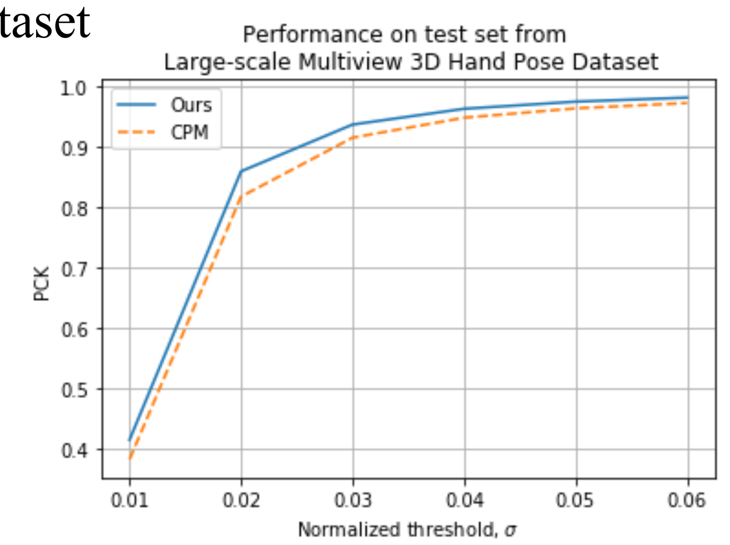
- Unary Branch: Apply deep neural network to the rotated image and output a set of heatmaps
- Rotation Net: Regress a rotation degree to make hands upwards
- Soft Classifier: Classify images by gestures and output a weight with Softmax
- Pool of Graphical Model: Each of the graphical model shares the same structure, but every single GM is associated with different parameters. Marginal probabilities are inferred on each GM, and then aggregated via a weight vector which comes from the soft classifier.

- **Learning:**
  - Generate the ground truth rotation degree from hand keypoints, then train the Rotation Net (ResNet-18).
  - Train unary branch (CPM) on rotated image
  - Generate hand gestures classification labels by applying K-means algorithm to rotated images based on relative position of keypoints, and train the Soft Classifier (ResNet-152)
  - Keep unary branch and soft classifier frozen, train the parameters of mixed graphical model
  - Jointly train all the parameters

Results

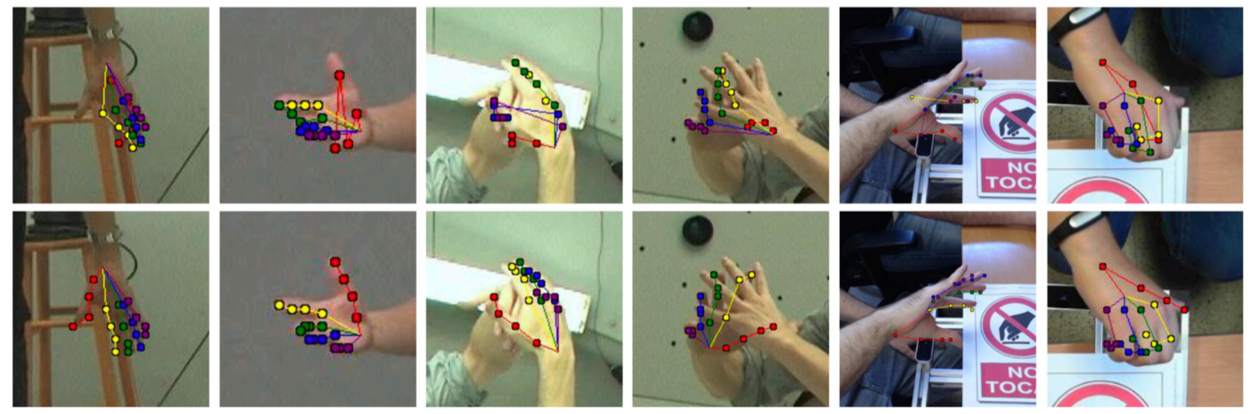
- **Quantitative Results:**
  - Probability of Correct Keypoint (PCK) curve on CMU panoptic dataset and Large-scale Multiview 3D hand pose dataset





Threshold of PCK, σ	0.01	0.02	0.03	0.04	0.05	0.06	mPCK
CMU Panoptic Hand Dataset							
CPM Baseline (%)	22.60	55.69	70.06	77.01	81.30	84.36	65.17
Ours	23.67	60.12	76.28	83.14	86.91	89.47	69.93
Improvement	1.07	4.43	**6.22**	**6.13**	**5.61**	**5.11**	**4.76**
Large-scale Multiview 3D Hand Pose Dataset							
CPM Baseline (%)	38.27	81.78	91.54	94.84	96.39	97.27	83.35
Ours	41.51	85.97	93.71	96.33	97.51	98.17	85.53
Improvement	**3.24**	**4.19**	2.17	1.49	1.12	0.90	2.18

- **Qualitative Results**



- **Ablation study**

Threshold of PCK, σ	0.01	0.02	0.03	0.04	0.05	0.06	mPCK	improvement
CPM Baseline (%)	22.60	55.69	70.06	77.01	81.30	84.36	65.17	-
CPM + Single GM	22.58	55.78	70.14	77.05	81.34	84.41	65.21	0.04
CPM + Mixture of GMs	23.39	57.53	71.95	78.49	82.28	85.02	66.44	1.27
Rotaion + CPM [1]	22.70	57.91	72.95	79.94	83.90	86.71	67.35	2.18
Rotaion + CPM [2]	21.97	57.59	74.53	81.98	86.21	88.83	68.52	3.35
R-MGMN	23.67	60.12	76.28	83.14	86.91	89.47	69.93	4.76

Contact: Deying Kong. University of California, Irvine. Email: deyingk@uci.edu