Nonparametric Structure Regularization Machine for 2D Hand Pose Estimation

Yifei Chen*[1], Haoyu Ma*[2], Deying Kong[2], Xiangyi Yan[2] , Jianbao Wu[1], Wei Fan[1] , and Xiaohui Xie[2]

[1] Tencent Hippocrates Research Lab

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

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*: Convolutional Pose Machines (CPM) and Stacked Hourglass, only capturing pose structure information implicitly.
  - *Multi-task learning*: unify hand pose estimation with hand mask segmentation, requiring a large amount of manually labelled mask for hand.
- **Our Contributions**:
  - We propose a novel cascade structure regularization methodology for 2D hand pose estimation, which utilizes synthetic hand masks to guide keypoints structure learning.
  - We propose a novel probabilistic representation of hand limbs and an anatomically inspired composition strategy for hand mask synthesis.

Learning

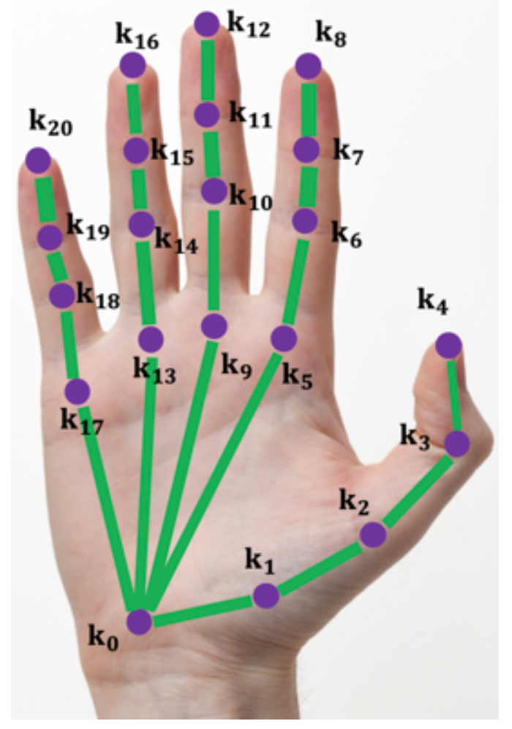
- **Loss:** $Loss = Loss_{keypoint} + \lambda_1 Loss^{G1}_{structure} + \lambda_2 Loss^{G6}_{Structure}$
- **Training Strategy:**

- End-to-End Training
- *Decayed loss schedule*: Structure learning is an auxiliary task, thus there is no need to get an accurate results, and our ultimate goal is keypoint. Let λ_1 and λ_2 decay by a ratio of 0.1 every 20 epochs during training.

Contact: haoyum3@uci.edu
Code: https://github.com/HowieMa/NSRMhand

Methodology

- **Limb Mask Representation:** Generate synthetic limb mask from labeled keypoints



- Hand model: 21 Keypoints + 20 Limbs L (Line Segment)
- **Limb Deterministic Mask (LDM)**: 0/1 mask around a limb

$$S_{LDM}(p|L) = \begin{cases} 1 & if\ p \in L \\ 0 & otherwise \end{cases}$$

- **Limb Probabilistic Mask (LPM)**: Gaussian heatmap around a limb

$$S_{LPM}(p|L) = \exp(-\frac{D(p,\overline{p_i p_j})}{2\sigma^2})$$

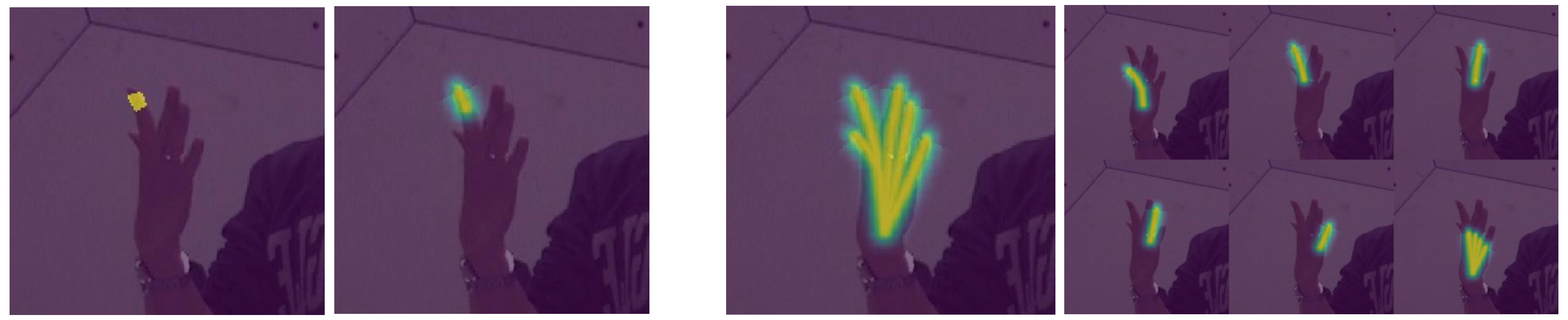
- **Limb Composition**

G1: coalesce 20 limbs together (whole hand mask)

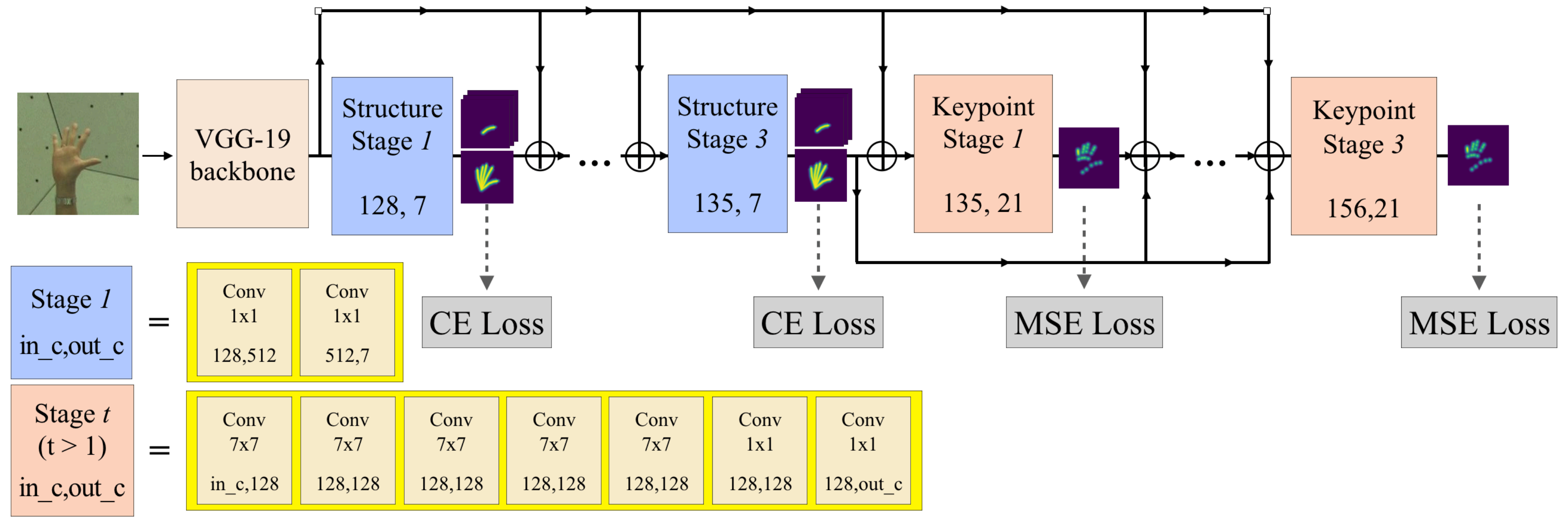
G6: coalesce 20 limbs into 6 groups (5 fingers + palm)

$$S*(p|g) = \max(S(p|L_1), S(p|L_2), \ldots, S(p|L_{|g|})$$

In practice, we mainly focus on utilizing G1 and G1&6 (the combination of G1 and G6).


LDM LPM G1 G6

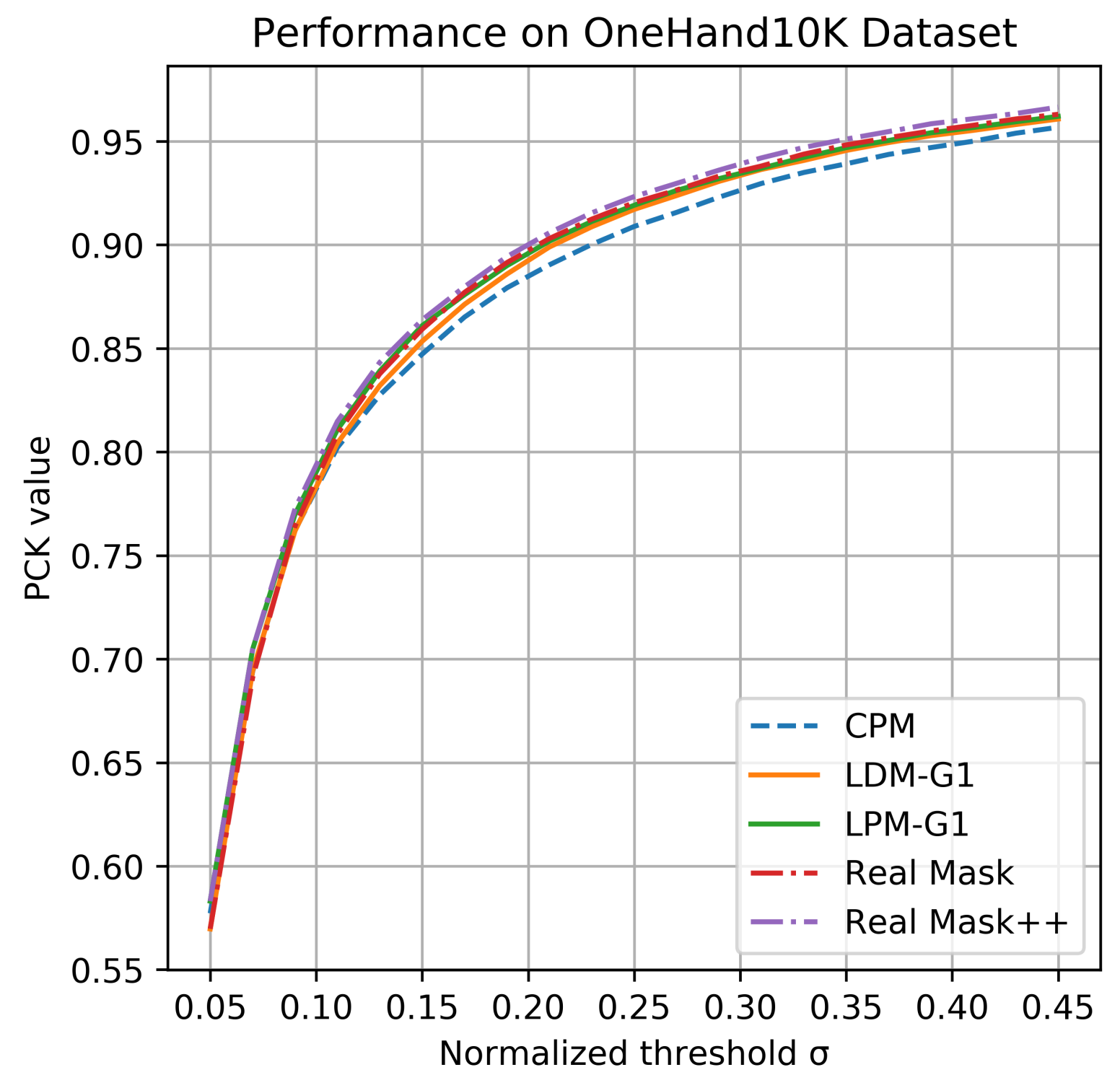
- **Network Architecture**: based on CPM

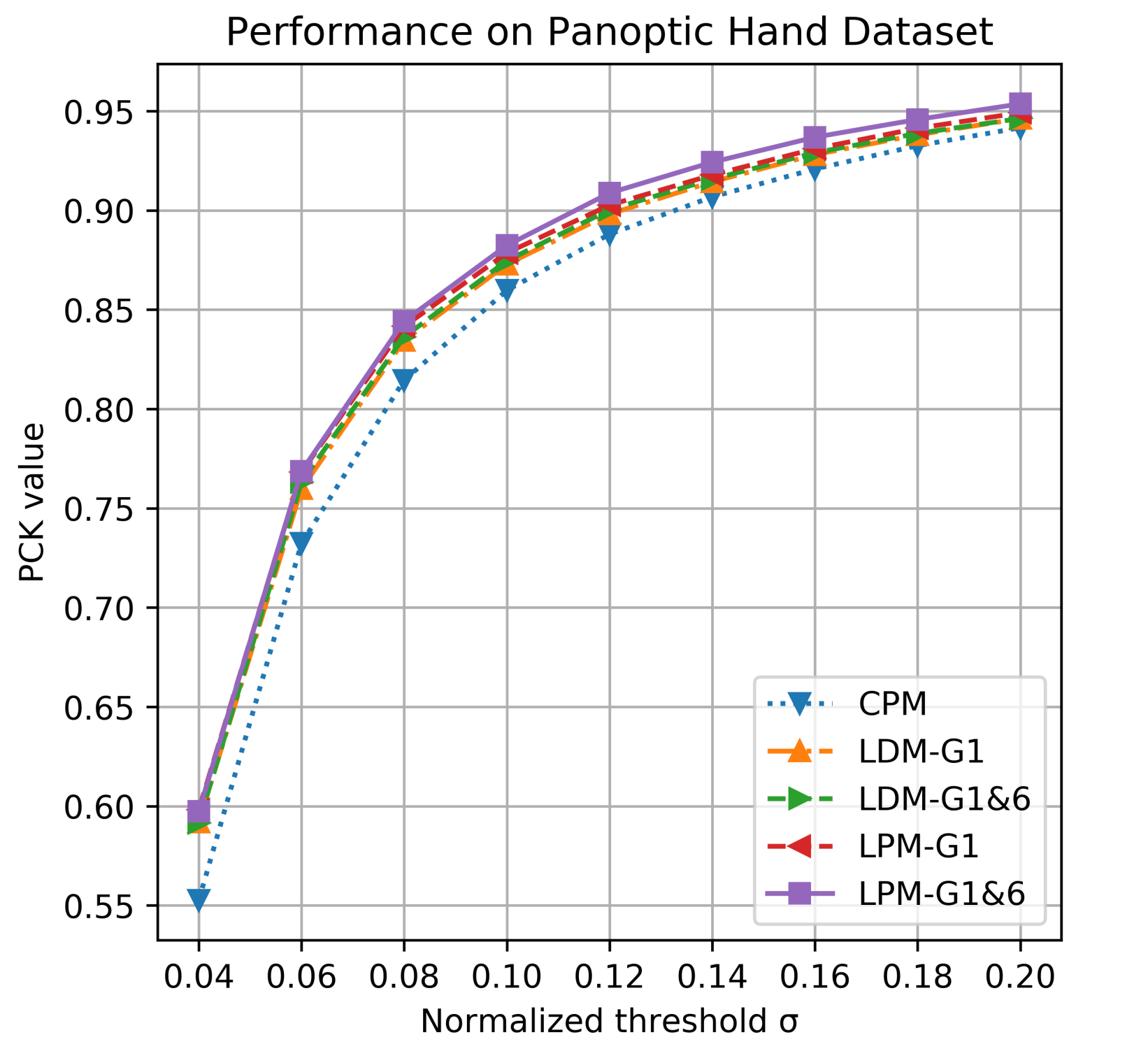


Results

- **Quantitative Results:**

- Probability of Correct Keypoint (PCK) curve on Onehand10k and panoptic hand dataset


Performance on OneHand10K Dataset


Performance on Panoptic Hand Dataset

- PCK value on Panoptic dataset

σ_{PCK}	0.04	0.06	0.08	0.10	0.12	ave	improvement
CPM	55.25	73.23	81.45	85.97	88.80	76.94	-
LDM-G1	59.20	75.98	83.45	87.28	89.81	79.14	+2.20 (+2.86%)
LDM-G1&6	59.16	76.32	83.63	87.46	90.03	79.32	+2.38 (+3.09%)
LPM-G1	59.81	76.82	84.16	87.86	90.26	79.78	+2.84 (+3.69%)
LPM-G1&6	59.73	76.86	84.43	88.23	90.87	80.03	**+3.09 (+4.01%)**

- **Qualitative Results**