Adaptive Graphical Model Network for 2D Handpose Estimation

Deying Kong [1]; Yifei Chen [2]; Haoyu Ma [3]; Xiangyi Yan [4]; Xiaohui Xie [1]
[1] University of California, Irvine; [2] Tencent; [3] Southeast University; [4] Southern University of Sci. & Tech.

Abstract

We propose a new framework of combining Deep Convolutional Neural Networks (DCNNs) and graphical models for 2D hand pose estimation from a monocular RGB image.

Method

The proposed Adaptive Graphical Model Network (AGMN) contains two branches of DCNNs and a probabilistic graphical model. The ***unary branch*** outputs preliminary confidence maps of positions of the keypoints, while the ***pairwise branch*** generates the pairwise potential functions between neighboring keypoints. The final confidence maps are then inferred via sum-product algorithm on a tree-structured ***graphical model***.

The key novelty lies in that the pairwise potential functions (or the parameters of the graphical model) are fully adaptive to and conditioned on the individual input image.

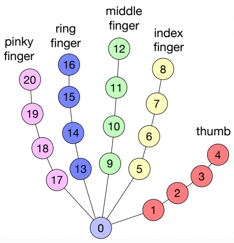


Fig. 3 Tree structured hand model.

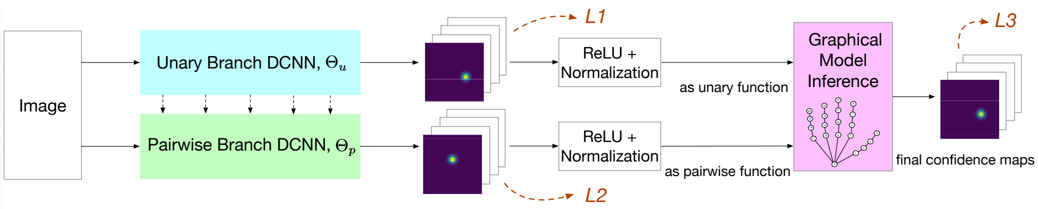


Fig. 1 Overview of the Adaptive Graphical Model Network (AGMN), with three loss functions indicated.

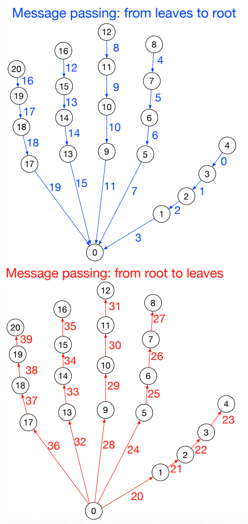

Message passing: from leaves to root
Message passing: from root to leaves

Fig. 4 Message passing strategy.

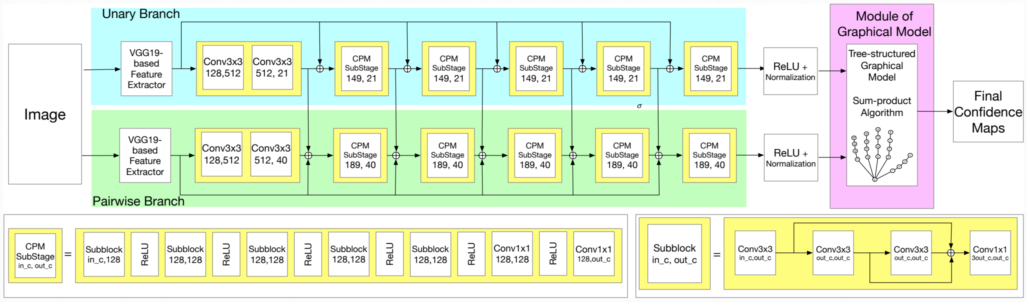


Fig. 2 Detailed structure of the Adaptive Graphical Model Network (AGMN).

Training Procedure and Results

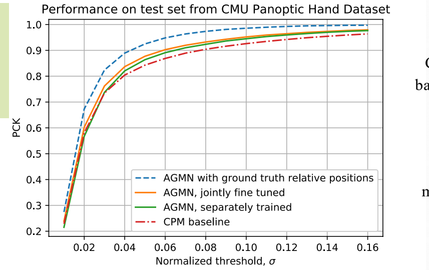
3-stage training:

1) Train unary branch with loss $L1$
2) Keep unary branch frozen, train pairwise branch with loss $L2$
3) Fine tune the whole network jointly with loss

$$L = L3 + 0.1 * L2 + 0.1 * L1$$

Normalized threshold of PCK	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.90	0.10
CPM Baseline (%)	22.88	58.10	73.48	80.45	84.27	86.88	88.91	90.42	91.61	92.61
AGMN Sep. Trained	21.52	56.73	73.75	82.06	86.39	89.10	91.00	92.35	93.63	94.50
AGMN Fine Tuned	23.90	60.26	76.21	83.70	87.72	90.27	91.97	93.23	94.30	95.20
Improvement	1.02	2.16	2.73	**3.25**	**3.45**	**3.39**	**3.06**	2.81	2.69	2.59

Table 1. Numerical results on CMU Hand Dataset



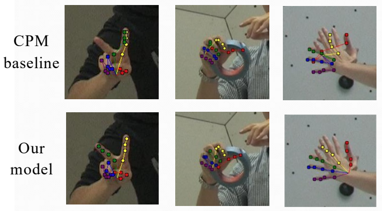
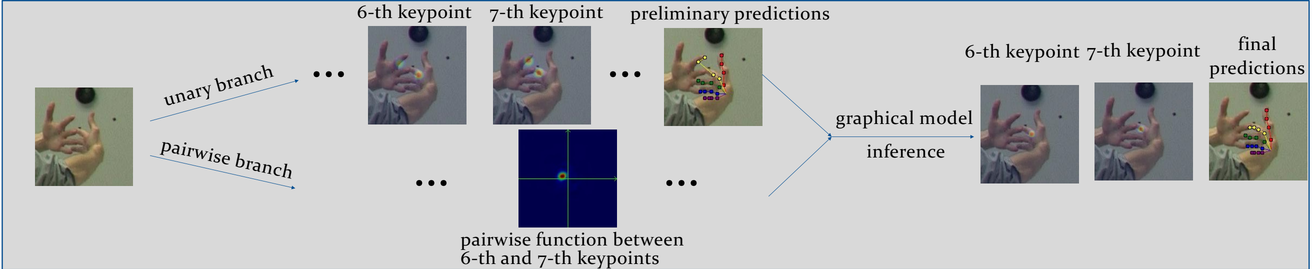


Fig. 4 Qualitative results.



Contact

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

References

1. Jonathan J Tompson, Arjun Jain, Yann LeCun, and Christoph Bregler. “Joint training of a convolutional network and a graphical model for human pose estimation," In *Advances in neural information processing systems*, pages 1799–1807, 2014.
2. Shih-En Wei, Varun Ramakrishna, Takeo Kanade, and Yaser Sheikh. “Convolutional pose machines,” In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4724–4732, 2016.
3. Tomas Simon, Hanbyul Joo, Iain Matthews, and Yaser Sheikh. “Hand keypoint detection in single images using Multiview bootstrapping,” In *The IEEE Conference on Computer Vision and Pattern Recognition*, volume 2, 2017.