Human Pose Estimation

Kuan-Lin Chen
kuc029@ucsd.edu

Hsiao-Chen Huang
hsh030@eng.ucsd.edu

Eddie Tseng
edtseng@eng.ucsd.edu

1 Introduction

The problem of human pose estimation, defined as the **problem of localization of human joints or parts** [8]. In this project, we aim to design and implement a real-time single-person pose detector by formulating the pose estimation as a joint regression problem.

Recently, there has been much research ongoing related to pose estimation using **convolutional architectures**. Toshev and Szegedy [8] have proposed a method of directly regressing the Cartesian coordinates by a series of refining regressors. The convolutional architecture proposed by Krizhevsky *et al.* [5] has shown a great potential on object detection and localization [7].

In our work, we are going to exploit the convolutional architectures and learn the underlying principle of designing a learning machine for pose estimation. Fig. 1 is our preliminary result of pose estimation using convolutional pose mahcines [9] and part affinity fields [3].

Figure 1: Example of 2D pose estimation

2 Datasets and Quantitative Analysis

The model will be trained and evaluated on a novel benchmark "**MPII Human Pose**" [2]. The overall performance will be verified on the "PCKh" measure which uses the matching threshold as 50% of the head segment length [2]. Table. 1 presents some selected works in terms of overall performance. Note that the investigation and implementation will primarily base on those selected works.

Method	Head	Sho.	Elb.	Wri.	Hip	Knee	Ank.	PCKh
Wei *et al.* [9]	97.8	95.0	88.7	84.0	88.4	82.8	79.4	88.5
Chu *et al.* [4]	98.5	96.3	91.9	88.1	90.6	88.0	85.0	91.5
Yang *et al.* [10]	98.5	96.7	92.5	88.7	91.1	88.6	86.0	92.0

Table 1: Overall performance of some selected works

3 Proposed Framework

Recently, there has been a surge of designing networks with branches, *e.g.*, convolutional pose machines [9] and pyramid residual modules [10]. In this project, we will take the advantages of those networks and observe different responses from tuning the hyperparameters such as number of layers and the size of receptive fields.

To address the difficulties in occlusion and invisible joints, the configurations of different parts has been proved essential for detecting correct parts [5, 8, 6, 9]. Therefore, the way how we encode the large spatial contextual information by using branches in convolutional architectures will be crucial to the performance. In other words, the part-to-part association needs to be investigated to adjust our architectures.

In our final work, we will implement **a sequential deep convolutional neural network (DCNN) with branches** that encode contextual information to demonstrate its tractability for human pose estimation.

4 Implementation and Expected Results

In this project, we will analyze the pros and cons of several architectures [8, 9, 4, 10] and implement our system using **TensorFlow** [1] on **Google Cloud Platform** (GCP) for pose estimation. The goal here is **NOT** to outperform state-of-the-art [10] performance on standard benchmarks but gain the underlying design principles and intuitions by implementing DCNN.

Since articulated human pose estimation is a fundamental challenging task due to its scale variations and invisible joints, it will be beneficial for us to go through this kind of problem. Finally, we hope to relate the experience of implementing DCNNs for human pose estimation with a strong design reasoning to explain how learning works. We expect the results are similar to Table. 1.

References

[1] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org.

[2] M. Andriluka, L. Pishchulin, P. Gehler, and B. Schiele. 2d human pose estimation: New benchmark and state of the art analysis. In *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pages 3686–3693, June 2014.

[3] Z. Cao, T. Simon, S. E. Wei, and Y. Sheikh. Realtime multi-person 2d pose estimation using part affinity fields. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1302–1310, July 2017.

[4] X. Chu, W. Yang, W. Ouyang, C. Ma, A. L. Yuille, and X. Wang. Multi-context attention for human pose estimation. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5669–5678, July 2017.

[5] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1*, NIPS'12, pages 1097–1105, USA, 2012. Curran Associates Inc.

[6] V. Ramakrishna, D. Munoz, M. Hebert, J. Andrew Bagnell, and Y. Sheikh. Pose machines: Articulated pose estimation via inference machines. In D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, editors, *Computer Vision – ECCV 2014*, pages 33–47, Cham, 2014. Springer International Publishing.

[7] C. Szegedy, A. Toshev, and D. Erhan. Deep neural networks for object detection. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 26*, pages 2553–2561. Curran Associates, Inc., 2013.

[8] A. Toshev and C. Szegedy. Deeppose: Human pose estimation via deep neural networks. In *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1653–1660, June 2014.

[9] S. E. Wei, V. Ramakrishna, T. Kanade, and Y. Sheikh. Convolutional pose machines. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4724–4732, June 2016.

[10] W. Yang, S. Li, W. Ouyang, H. Li, and X. Wang. Learning feature pyramids for human pose estimation. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 1290–1299, Oct 2017.