Internship of Research Master 2 in Computer Science

Work on COCO 2018 Keypoint Detection Task

Author :
Qixiang PENG

Stage chief :
Dr. Gang YU

Host organization : Megvii Research

Secretariat - tel : 01 69 15 81 58
Email Address: alexandre.verrecchia@u-psud.fr

Contents

Contents i

1 Introduction to Company and Team 2
- 1.1 Company Introduction . . . 2
- 1.2 Team Introduction . . . 2

2 Presentation of the context of the task 3
- 2.1 Presentation of Human Pose Estimation . . . 3

3 Conclusion et perspectives 4

Bibliography 6

Abstract

This report summarizes my internship in Megvii Research: Work on COCO 2018 Keypoint Detection Task[1]. This challenge is designed to push the state of the art in multi-person pose estimation.

The topic of multi-person pose estimation has been largely improved recently, especially with the development of convolutional neural network. However, there still exist a lot of challenging cases, such as occluded keypoints, invisible keypoints and complex background.

Nowadays, two solutions are adopted widely: Bottom-Up approaches and Top-Down approaches. In this challenge, our team proposed a novel top-down method.

Keywords

COCO 2018 Keypoint Detection, human pose estimation, top-down.

[1] More detail about this challenge can be found in https://competitions.codalab.org/competitions/12061

CHAPTER 1

Introduction to Company and Team

1.1 Company Introduction

Founded in October 2011, Megvii is an Articial Intelligence company specialized in providing enterprises and developers with intelligent solutions and data services, and is dedicated to the mission of Create machines that can see and think. With the "cloud + end" system of Megvii Cloud and Megvii SensorNet as its core products, Mevgii has successfully offered solutions for over 800 enterprises in nance, security, office, real estate and other business sectors.

Megvii holds more than 350 domestic and international patents. Over seven years of development, Megvii has gathered a workforce of over 1000 people, among whom more than 70% are RD stas. The core team of Megvii is composed of top geeks who are alumni of universities like Tsinghua, Columbia, Oxford, etc., and adventurers formerly working for Google, Alibaba, Huawei and IBM. Over 80 people in Megvii have been awarded golden prizes of informatics at national and international levels. Research teams from Megvii have been and are holding the first places in more than ten international AI benchmarks.

The name Megvii is from mega vision, which means our work is concentrated on offering computer vision technologies that enable your applications to read and understand the world better. In fact, FACE++, the best product of megvii, now, is the biggest platform of face detection over the world.

Here is the link to official website: https://www.faceplusplus.com/

1.2 Team Introduction

During the internship, I worked in Detection Team in Megvii Research. The team leader is Gang YU[1].

In general, our team is in charge of 4 main issues:

1. **Detection:** Face Detection, Pedestrian/human Detection, Vehicle/Plate Detection, General Object Detection, Object Detection in Video, 3D Object detection (combined with Point Cloud)
2. **Segmentation:** Semantic Segmentation, Instance Segmentation, Panoptic Segmentation, Video Segmentation, 3D Segmentation
3. **Skeleton:** Human Pose Estimation, Hand Pose Estimation
4. **Action:** Action Recognition in Video

Our team has a solid technical accumulation, especially in the detection aspect. We have the winner solution of COCO2017 Detection: MegDet [2]. From a product perspective, we built a small repo of imagenet base model for training and exploring models with less than 100M FLOPs. In addition to Detection, our skeleton solution also took the first in the COCO2017 Human Pose competition: CPN [1]. In terms of Segmentation, we also have some better work published. In addition, we have sufficient GPU resources, as well as very large internal data sets for exploring the upper-bound of various research tasks.

[1]His google scholar is: https://scholar.google.com/citations?user=BJdigYsAAAAJhl=en

CHAPTER 2

Presentation of the context of the task

In this chapter, several brief presentations will be given to explain the context of the human pose estimation and describe the COCO dataset.

2.1 Presentation of Human Pose Estimation

Localizing body parts for human body is a fundamental yet challenging task in computer vision, and it serves as an important basis for high-level vision tasks, e.g., activity recognition [3, 4], human re-identification [5], and human-computer interaction. In generala human pose estimation model aims to predict the 2D coordinates of different human parts given a 2D human image. Achieving accurate localization, however, is difficult due to the highly articulated human body limbs, occlusion, change of viewpoint, and foreshortening. Nowadays there exists two main topics in human pose estimation: single person pose estimation and multi-person pose estimation. Obviously, multi-person is more challenging than single person pose estimaion. But single person is the fundamentation for multi-person pose estimation.

(a) Small Box with a Long Caption

(b) Big Box

Figure 2.1: Two Subfigures

CHAPTER 3

Conclusion et perspectives

Cuius acerbitati uxor grave accesserat incentivum, germanitate Augusti turgida supra modum, quam Hannibaliano regi fratris filio antehac Constantinus iunxerat pater, Megaera quaedam mortalis, inflammatrix saevientis adsidua, humani cruoris avida nihil mitius quam maritus; qui paulatim eruditiores facti processu temporis ad nocendum per clandestinos versutosque rumigerulos conpertis leviter addere quaedam male suetos falsa et placentia sibi discentes, adfectati regni vel artium nefandarum calumnias insontibus adfligebant.

Acknowledgement

At last, i would like to point out that i couldnt nish this internship successfully without someones, and here i give my most sincere thanks to them. Firstly, i will express thanks to Zhicheng WANG, one of my team leader. Its him that taught me the basic knowledge of DL, like CNN, Resnet, frameworks, etc. He also explained papers to me clearly and carefully. In fact, he leads me into the domain CVDL. Next, iwould like to thank for Gang YU, my another team leader. During August, i did some research work about SOT(single object tracking), which is supervised by him. He gave me lots of help, like correcting my wrong opinions and reviewing my codes. Finally but with same importance, my colleagues helped me a lot when i encountered some problems about software or hardware, i will always appreciate that.

Bibliography

[1] Yilun Chen, Zhicheng Wang, Yuxiang Peng, Zhiqiang Zhang, Gang Yu, and Jian Sun. Cascaded pyramid network for multi-person pose estimation. *arXiv preprint arXiv:1711.07319*, 2017.

[2] Chao Peng, Tete Xiao, Zeming Li, Yuning Jiang, Xiangyu Zhang, Kai Jia, Gang Yu, and Jian Sun. Megdet: A large mini-batch object detector. pages 6181–6189, 2018.

[3] Chunyu Wang, Yizhou Wang, and Alan L Yuille. An approach to pose-based action recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 915–922, 2013.

[4] Weilong Yang, Yang Wang, and Greg Mori. Recognizing human actions from still images with latent poses. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 2030–2037. IEEE, 2010.

[5] Liang Zheng, Yujia Huang, Huchuan Lu, and Yi Yang. Pose invariant embedding for deep person re-identification. *arXiv preprint arXiv:1701.07732*, 2017.