

Macro-Micro Adversarial Network for Human Parsing

Yawei Luo^{1,2}, Zhedong Zheng², Liang Zheng^{2,3}, Tao Guan¹, Junqing Yu¹, and
Yi Yang²

¹ School of Computer Science and Technology,
Huazhong University of Science and Technology

Froyalvane, qd.gt, yjqingG@hust.edu.cn

² CAI, University of Technology Sydney

³ Singapore University of Technology and Design

Fzdzheng12, lliangzheng06, yee.ii.yang@gmail.com

Abstract. In human parsing, the pixel-wise classification loss has drawbacks in its low-level local inconsistency and high-level semantic inconsistency. The introduction of the adversarial network tackles the two problems using a single discriminator. However, the two types of parsing inconsistency are generated by distinct mechanisms, so it is difficult for a single discriminator to solve them both. To address the two kinds of inconsistencies, this paper proposes the Macro-Micro Adversarial Net (MMAN). It has two discriminators. One discriminator, Macro D, acts on the low-resolution label map and penalizes semantic inconsistency, e.g., misplaced body parts. The other discriminator, Micro D, focuses on multiple patches of the high-resolution label map to address the local inconsistency, e.g., blur and hole. Compared with traditional adversarial networks, MMAN not only enforces local and semantic consistency explicitly, but also avoids the poor convergence problem of adversarial networks when handling high resolution images. In our experiment, we validate that the two discriminators are complementary to each other in improving the human parsing accuracy. The proposed framework is capable of producing competitive parsing performance compared with the state-of-the-art methods, i.e., mIoU=46.81% and 59.91% on LIP and PASCAL-Person-Part, respectively. On a relatively small dataset PPSS, our pre-trained model demonstrates impressive generalization ability. The code is publicly available at <https://github.com/RoyalVane/MMAN>.

Keywords: Human parsing, Adversarial network, Inconsistency, Macro-Micro

1 Introduction

Human parsing aims to segment a human image into multiple semantic parts. It is a pixel-level prediction task which requires to understand human images in both the global level and the local level. Human parsing can be widely applied to human behavior analysis [9], pose estimation [34] and fashion synthesis [40].

