

# Person Search by Multi-Scale Matching

Xu Lan<sup>1</sup>, Xiatian Zhu<sup>2</sup>, and Shaogang Gong<sup>1</sup>

<sup>1</sup> Queen Mary University of London,  
x.lan@qmul.ac.uk, s.gong@qmul.ac.uk

<sup>2</sup> Vision Semantics Ltd  
eddy@visionsemantics.com

**Abstract.** We consider the problem of person search in unconstrained scene images. Existing methods usually focus on improving the person detection accuracy to mitigate negative effects imposed by misalignment, mis-detections, and false alarms resulted from noisy people auto-detection. In contrast to previous studies, we show that sufficiently reliable person instance cropping is achievable by slightly improved state-of-the-art deep learning object detectors (e.g. Faster-RCNN), and the under-studied multi-scale matching problem in person search is a more severe barrier. In this work, we address this multi-scale person search challenge by proposing a Cross-Level Semantic Alignment (CLSA) deep learning approach capable of learning more discriminative identity feature representations in a unified end-to-end model. This is realised by exploiting the in-network feature pyramid structure of a deep neural network enhanced by a novel cross pyramid-level semantic alignment loss function. This favourably eliminates the need for constructing a computationally expensive image pyramid and a complex multi-branch network architecture. Extensive experiments show the modelling advantages and performance superiority of CLSA over the state-of-the-art person search and multi-scale matching methods on two large person search benchmarking datasets: CUHK-SYSU and PRW.

**Keywords:** Person Search; Person Detection and Re-Identification; Multi-Scale Matching; Feature Pyramid; Image Pyramid; Semantic Alignment.

## 1 Introduction

Person search aims to find a probe person in a gallery of whole unconstrained scene images [41]. It is an extended form of person re-identification (re-id) [12] by additionally considering the requirement of automatically detecting people in the scene images besides matching the identity classes. Unlike the conventional person re-id problem assuming the gallery images as either manually cropped or carefully filtered auto-detected bounding boxes [40, 24, 3, 15, 20, 44, 25, 39, 37, 2, 44], person search deals with raw unrefined detections with many false cropping and unknown degrees of misalignment. This yields a more challenging matching problem especially in the process of person re-id. Moreover, auto-detected person boxes often vary more significantly in scale (resolution) than the conventional



































